European Commission

Study on Mapping Data Flows

Final Report

Authors:

VVA: Luena Collini, Lison Rabuel, Malin Carlberg

Tech4i2: Paul Foley, Alexandra Gemmell

Internal identification

Contract number: LC-01291841
SMART number 2019/0010

EUROPEAN COMMISSION

Directorate-General for Communications Networks, Content and Technology

Directorate E — Future Networks

Unit E.2 — Cloud & Software Unit

Contact: Alexandra Paul

E-mail: cnect-e2@ec.europa.eu

European Commission

Study on Mapping Data Flows

Directorate-General for Communications Networks, Content and Technology 2021

LEGAL NOTICE

This document has been prepared for the European Commission however it reflects the views only of the authors, and the European Commission is not liable for any consequence stemming from the reuse of this publication. More information on the European Union is available on the Internet (http://www.europa.eu).

PDF	ISBN 978-92-76-41937-2	doi:10.2759/7355	KK-03-21-388-EN-N

Manuscript completed in October 2021

Luxembourg: Publications Office of the European Union, 2021

© European Union, 2021

The reuse policy of European Commission documents is implemented by the Commission Decision 2011/833/EU of 12 December 2011 on the reuse of Commission documents (OJ L 330, 14.12.2011, p. 39). Except otherwise noted, the reuse of this document is authorised under a Creative Commons Attribution 4.0 International (CC-BY 4.0) licence (https://creativecommons.org/licenses/by/4.0/). This means that reuse is allowed provided appropriate credit is given and any changes are indicated.

For any use or reproduction of elements that are not owned by the European Union, permission may need to be sought directly from the respective rightholders.

Table of contents

List of Tables

List of Figures

Abstract

Measuring volume and trends of cloud based data flows[1] remains a complex and yet sporadic exercise. To date, only a very limited number of studies, targeted to either a specific sector, a geography or a 'data flow' type, have tried to map data flows. None of these studies have succeeded in developing a robust, systematic and replicable data collection methodology to identify, map, and monitor data flows in a holistic and aggregated manner.

This study, which is part of the European Commission initiative 'The European Data Flow Monitoring',[2] provides an innovative and replicable methodology to estimate and monitor the volume and types of data flows within the EU, EFTA countries and the UK. For the purpose of the study, enterprise data flows to cloud and edge data centres are defined as the movement of data from the end user to the data centre in question.

One of the core objectives of the study is to support the evaluation of the Free Flow of Non-Personal Data Regulation at macro-economic level with data related to volume and trends in data flows to assess if barriers to data movement have effectively been removed.

The second core objective is to deliver on one of the key actions of the European Strategy for Data[3] by making a first attempt to develop the analytical framework for estimating data flows. The framework provides the tools for a continuous analysis of data flows and the economic development of the EU's data processing sector. It can be used in future to monitor data flow trends across and within the European Union to provide evidence in support of EU policy, trade and investment decisions.

1 The term cloud based data flow is used for enterprise data flowing to cloud and edge data centres. It describes data flowing from enterprises in private and public sectors to cloud data or edge centres;

2 'The European Data Flow Monitoring' (https://digital-strategy.ec.europa.eu/en/policies/european-data-flow-monitoring)

3 European Commission Communication, 'A European strategy for data', COM(2020) 66 final.

Résumé

Mesurer le volume et les tendances des flux de données basés sur le cloud[4] reste un exercice complexe mais sporadique. À ce jour, seul un nombre très limité d'études, ciblées soit sur un secteur spécifique, soit sur une géographie ou un type de « flux de données », ont essayé de cartographier les flux de données. Cependant, aucune de ces études ne réussit à développer une méthodologie de collecte de données robuste, systématique et reproductible pour identifier, cartographier et surveiller les flux de données de manière holistique et agrégée.

Cette étude, qui fait partie de l'initiative de la Commission européenne « The European Data Flow Monitoring »[5], fournit une méthodologie innovante et reproductible pour estimer et surveiller les flux de données au sein de l'UE, des pays de l'AELE et du Royaume-Uni. Dans cette étude, le flux de données d'entreprise vers les centres de données cloud principaux et périphériques est défini comme le mouvement des données de l'utilisateur final vers le centre de données en question.

L'un des principaux objectifs de cette étude est de soutenir l'évaluation du règlement sur la libre circulation des données non personnelles au niveau macro-économique avec des données relatives au volume et aux tendances des flux de données afin d'évaluer si les obstacles à la circulation des données ont effectivement été supprimés.

Le deuxième objectif principal de cette étude est de faire une première tentative pour développer le cadre analytique pour estimer les flux de données afin de mettre en œuvre l'une des actions clés de la Stratégie Européenne pour les Données[6]. Ce cadre fournit les outils nécessaires pour mener une analyse continue des flux de données et du développement économique du secteur de l'informatique de l'UE. Il peut être utilisé dans les années à venir pour surveiller les tendances des flux de données, à la fois au sein de l'UE et entre l'UE et le reste du monde, pour étayer les décisions de l'UE en matière de politique, de commerce et d'investissement.

4 Le terme flux de données basé sur le cloud est utilisé pour les données d'entreprise circulant vers les centres de données cloud et périphériques. Il décrit les flux de données des entreprises des secteurs privé et public vers les centres de données cloud ou de périphérie.

5 The European Data Flow Monitoring' (https://digital-strategy.ec.europa.eu/en/policies/european-data-flow-monitoring).

6 Communication de la Commission européenne, « Une stratégie européenne pour les données », COM(2020) 66 final.

Executive Summary

a. Introduction

This report presents the work carried out for the study 'Mapping Data Flows, SMART 2019/0010', commissioned by the Directorate-General for Communications Networks, Content and Technology (DG CNECT), Cloud and Software Unit, of the European Commission and carried out by VVA Economics & Policy and Tech4i2.

Acknowledging the increasing role played by data in our economy,[7] the European Commission released in 2020 a Communication on a **European Strategy for Data**[8] developing a common European data space, interconnecting cloud infrastructure, setting up a cloud service marketplace for EU users, as well as setting strategic EU investments in new technologies, such as edge computing, high-performance computing/quantum computing.

One of the key actions of the European Data Strategy is to create a **'European analytical framework for measuring data flows'**,[9] given the strategic importance of monitoring data flows to inform EU decision-making and trade and investment choices in the area of cloud computing.

b. Objectives of the study

This study, which is part of the European Commission initiative 'The European Data Flow Monitoring',[10] provides an innovative method for estimating the volume and types of enterprise data flowing to cloud infrastructures and for investigating where data flows across the EU. Enterprise data flows to cloud and edge data centres is the movement of data from the end user to the data centre in question. For example, if a user was to use the OVH services, the data would flow from user in the home country to the closest OVH cloud data centre. In this study, the assumption has been made that 80 per cent of cloud data flows to main cloud data centres and 20 per cent flows to edge data centres[11].

The report provides a holistic approach and an integrated view of enterprise data flowing to cloud data centres and edge centres within the EU; between the EU and the UK; and between EU and EFTA countries. It captures several dimensions of the data flows, namely the sectors[12] (NACE sectors C-N) triggering the flows, enterprise size (small,

7 In 2018, the value of the data economy in Europe reached 2.4 per cent of European GDP, and it is estimated to reach almost 6 per cent of EU GDP in 2025. Moreover, the global data volume is expected to grow five times the 2018 figures in 2025. Source: Lisbon Council, International Data Corporation. 'Final Study Report: The European Data Market Monitoring Tool' (https://datalandscape.eu/study-reports/final-study-report-european-data-market-monitoring-tool-key-facts-figures-first-policy).

8 European Commission Communication, 'A European strategy for data', COM(2020) 66 final.

9 Idem.

10 'The European Data Flow Monitoring' (https://digital-strategy.ec.europa.eu/en/policies/european-data-flow-monitoring).

11 The assumption is based from the European Commission Communication, 'A European strategy for data', COM(2020) 66 final.

12 C Manufacturing, D Electricity Gas, Steam and air conditioning supply, E Water Supply, F Construction, G Wholesale and retail trade, H Transportation and storage, I Accommodation and food service activities, J Information and Communication, K Financial and insurance activities, L Real estate activities, M Professional, scientific and technical activities, N Administrative and support services.

medium, large enterprises),[13] services,[14] and data-intensive enterprise activities[15] and data types (personal/non-personal).[16] The study was not able to investigate data flows between the cloud and edge data centres of the businesses providing cloud services. These exchanges might arise so that businesses providing cloud service can back up data or operate more efficiently. Furthermore, the study did not investigate data flowing to private cloud infrastructure on company premises. This provides data for 2016 to 2020, with forecsts up to 2030.

The methodology presented here can be used in future to monitor data flow trends across the European Union (the EU), between the EU and EFTA countries and the UK in support of EU policy, trade and investment decisions. Furthermore, the findings provide useful evidence to support the evaluation of the Regulation of the free flow of data and the tools for a continuous analysis of data flows and the economic development of the EU's data processing sector.

c. Overall data flows

Overall, in the EU27 member states, the flow of **data stemming from enterprises[17] buying cloud services used over the internet** was 504,763 TB/month (5.8 EX/year) in 2020. By 2030 the amount is forecast to be 15.2 times greater at 7,669,835 TB/month (87.8 EX/year) and by 2025 the amount is forecast to be 4.4 times greater (than the 2020 value) at 2,201,058 TB/month (25.2 EX/year).

Across the 31 countries examined[18], **data flows to cloud and edge data centres by enterprises buying cloud services used over the internet** were estimated to be 708,790 TB/month (8.1 EX/year) in 2020. By 2030 the amount is forecast to be 14.9 times greater at 10,577,600 TB/month (121.1 EX/year). By 2025 the amount forecast is to be 4.3 times greater (the 2020 value) at 3,025,388 TB/month (34.6 EX/year).

d. Country level analysis

One element of the analysis was to estimate enterprise data flows (TB/month) to the cloud and edge data centres of the ten main cloud service providers in countries that possess cloud infrastructure between 2016 and 2021[19]. **The largest volume of enterprise data flows is served by cloud and edge data centres in Germany**. In 2020, Germany received 151,968 TB/month (1.74 EX/year) of cloud data flows from other countries. This represents 30.7 per cent of cloud data flows to EU27 Member States and 25.9 per cent of cloud data flows across the 31 countries studied. The EU27 Member State receiving the **second highest inflow of data served by its cloud and edge data centres in 2020 was the Netherlands** (86,963 TB/month, 1 EX/year).

13 Eurostat definition: small enterprises: 10-49 persons employed; medium-sized enterprises: 50-249 persons employed; small and medium sized enterprises (SMEs): 1-249 persons employed; large enterprises: 250 or more persons employed.

14 See 'Definitions' in Section 1.2.

15 'Data intensive activities' refers to data stored in cloud infrastructures and generated by activities that are highly dependent on the use, production or/and provision of the cloud technology, such as industrial data, which is one of the typical data to feed into customer relationship management (CRM) activities. For more information, please refer to Annex 1.

16 'Data type' refers to personal and non-personal data. For more information, please refer to Annex 1.

17 In this study, enterprises are defined as private NACE sectors (C to M) and public NACE sectors (N to S). Enterprises consist of mostly businesses (see footnote 8) but also some public sectors (N Administrative and Support Services, P Education, Q Human Health and Social Work, R Arts and Entertainment and Recreation, S Other Service Activities).

18 Liechtenstein is not included in the analysis due to a lack of data on Eurostat

19 Analysis focuses on 2016 to 2021 because information about cloud infrastructure in these years is known. Many commentators predict considerable growth in cloud edge data centres in the next few years. Extrapolation to create forecasts in the future, when the distribution of infrastructure will change considerably, would create spurious results.

Both these countries (Germany and the Netherlands) received more enterprise data flowing to their cloud and edge data centres in 2020 than the UK (66,064 TB/month, 0.76 EX/year). In 2020, the UK received 12.8 per cent of total data flowing to cloud and edge facilities from the 31 countries studied. EFTA countries received 3 per cent of total data flowing to cloud and edge facilities (15,506 TB/month, 0.18 EX/year).

Analysis discovered that in 2020, the highest outflow of data (data flowing from a country to a cloud and edge data centre) was generated by the United Kingdom (152,815 TB/month, 1.75 EX/ year), followed by Germany (93,474 TB/month, 1.07 EX/year), and France (75,039 TB/month, 0.86 EX/year).

e. Net data flows by country

A further stage of analysis examined *net* enterprise data flowing to cloud and edge data centres for the 31 countries examined in the study. This was found by subtracting data inflows to a country from data outflows generated by the country. Three countries have net enterprise data inflows to their cloud and edge data centres. All other countries have net data outflows.

The country with the highest *net* data inflow to its cloud and edge data centres in 2020 was Germany. German net enterprise data inflow to its cloud and edge data centres in 2020 was 58,494 TB/month (0.67 EX/year). **Ireland** has the second highest net enterprise data inflow to its cloud and edge data centres (45,981 TB/month, 0.53 EX/year). The **Netherlands** has the third highest net enterprise data inflow to its cloud and edge data centres in 2020 (45,504 TB/month (0.52 EX/year).

The EU27 Member State with the largest net outflow of enterprise data to cloud and edge data centres in other countries is Italy; 42,923 TB/month (0.49 EX/year). Of the 31 countries examined, the largest net enterprise data out flows were recorded by the **UK**; 86,751 TB/month (1 EX/year).

f. Cloud service adoption and forecasts to 2030

In terms of sectors, in 2020, the largest data flows to cloud come from the Health sector (NACE Q, 12.9 per cent of all flows, 91,600 TB/month; 1.048 EX/year), **Retail and Wholesale** (NACE G, 12.9 per cent, 91,400 TB/month; 1.046 EX/year) and **Education** (12.5 per cent, 88,600 TB/month; 1.01 EX/year). In 2025, it is expected that Retail and Wholesale will have 13.1 per cent of all flows (390,300 TB/month; 4.47 EX/year), the Health sector will have 12.3 per cent of all flows (377,900 TB/month; 4.32 EX/year), and Education will have 12.1 per cent of all flows (367,800 TB/month; 4.21 EX/year). In 2030, it is expected that Retail and Wholesale will have 13.2 per cent of all flows (1,370,670 TB/month; 15.69 EX/year), Health will have 12.1 per cent of all flows (1,293,560 TB/month; 14.80 EX/year), and Education will have 11.9 per cent of all flows (1,262,730 TB/month; 14.45 EX/year).

Moreover, in 2020 the largest volume of data flowing to main cloud and edge data centres comes from enterprises with 250 or more employees. In the EU27 Member States, enterprises with 250 or more employees triggered 61 per cent of all data flows to cloud (309,500 TB/month [3.54 EX/year] of total EU27 flows of 504,700 TB/month [5.78 EX/year] in 2020).[20] EU27 Member States enterprises with less than 10 employees account

20 In EFTA countries, data flows from enterprises with 250 or more employees comprise 48 per cent of flows (24,700 TB/month of 51,200 TB/month). In the UK, data flows from enterprises with 250 or more employees comprise 66 per cent of flows (100,500 TB/month of 152,800 TB/month).

for 8.5 per cent of all data flows to cloud (42,870 TB/month; 0.49 EX/year). EU27 Member States enterprises with 10 to 49 employees achieved 13.2 per cent of all data flows to cloud (66,560 TB/month; 0.76 EX/year). EU27 Member States enterprises with 50 to 249 employees obtained 17 per cent of all data flows to cloud (85,800 TB/month; 0.98 EX/year).

By 2030 it is expected that 65 per cent of EU27 enterprises with 10 to 49 employees will buy cloud services used over the internet. This is almost twice the number in 2020 (34 per cent) and higher than the 52 per cent of EU27 enterprises with 10 to 49 employees who are expected to buy cloud services used over the internet by 2025[21]. **46 per cent of EU27 enterprises with 50 to 249 employees bought cloud services used over the internet in 2020**. In 2025 and 2030, this is expected to increase to 68 per cent and 83 per cent, respectively[22]. **For EU27 enterprises with over 250 employees, in 2020, 66 per cent bought cloud services used over the internet.** In 2025 this is expected to increase to 90 per cent, and then to 97 per cent by 2030[23].

In terms of services[24], the proportion of enterprises buying low and medium levels of cloud services used over the internet is declining. An increasing number of enterprises buy high level cloud services used over the internet. In 2020 19 per cent of EU27 enterprises used low levels of cloud services, 30 per cent used medium levels of cloud services and 51 per cent used high levels of cloud services. In 2025, it is forecasted that 15 per cent of cent of enterprises will be using low levels of cloud services, 22 per cent using medium levels of cloud services, and 63 per cent will be using high levels of cloud services. By 2030, the situation is expected to change significantly - 72 per cent of enterprises are expected to be using high levels of cloud services, 16 per cent are forecast to use medium services and only 12 per cent will be purchasing low levels of cloud services.

Concerning data types, 41 per cent of the total data stored in cloud infrastructure is personal data and 59 per cent is non-personal data. Of these 41 per cent of personal data, 11 per cent is generated by a user/individual. Concerning non-personal data, out of 59 per cent, around 22 per cent is generated by a machine. Within machine-generated data, 19 per cent is industrial (productive) data, namely used as a direct input into the production process from a supply chain perspective and the remaining 3% are indicated as “other data” by survey respondents.

21 EFTA: 2020 – 63 per cent, 2025 – 90 per cent, 2030 – 98 per cent. UK: 2020 – 41 per cent, 2025 – 72 per cent, 2030 – 95 per cent.

22 EFTA: 2020 – 81 per cent, 2025 – 97 per cent, 2030 – 99 per cent. UK: 2020 – 63 per cent, 2025 – 92 per cent, 2030 – 98 per cent.

23 EFTA: 2020 – 90 per cent, 2025 – 98 per cent, 2030 – 100 per cent. UK: 2020 – 77 per cent, 2025 – 87 per cent, 2030 – 97 per cent.

24 Eurostat definitions of low, medium and high cloud computing (CC) services are as follows; low CC services are email, office software, storage of files, medium CC services include low CC services plus hosting of the enterprises’ database, high CC services are accounting software applications, CRM software, computing power.

Résumé

a. Introduction

Ce rapport présente les travaux réalisés pour l'étude « Mapping Data Flows, SMART 2019/0010 », commandée par la Direction générale des réseaux de communication, des contenus et des technologies (DG CNECT), Unité Cloud et logiciels, de la Commission européenne et réalisée par VVA Economics & Policy et Tech4i2.

Reconnaissant le rôle croissant joué par les données dans notre économie[25], la Commission européenne a publié en 2020 une communication sur une **stratégie européenne pour les données**[26] développant un espace de données européen commun, interconnectant l'infrastructure cloud, mettant en place un marché de services cloud pour les utilisateurs de l'UE, ainsi que la définition d’investissements stratégiques de l'UE dans les nouvelles technologies, telles que l'informatique de pointe, l'informatique haute performance/l'informatique quantique.

L'une des actions clés de la stratégie européenne en matière de données consiste à créer un «cadre analytique européen pour mesurer les flux de données»[27], étant donné l'importance stratégique du suivi des flux de données pour éclairer la prise de décision et les choix commerciaux et d'investissement de l'UE dans le domaine du cloud computing.

b. Objectifs de l'étude

Cette étude, qui fait partie de l'initiative de la Commission européenne « The European Data Flow Monitoring »,[28] fournit une méthode innovante pour estimer le volume et les types de données d'entreprise circulant vers les infrastructures cloud et pour étudier où les données circulent dans l'UE. Les flux de données d'entreprise vers les centres de données cloud principaux et périphériques correspondent au mouvement des données de l'utilisateur final vers le centre de données en question. Par exemple, si un utilisateur devait utiliser le services OVH, les données seraient transmises de l'utilisateur du pays d'origine au centre de données cloud OVH le plus proche. Dans cette étude, l'hypothèse a été faite que 80 % des données cloud circulent vers les principaux centres de données et 20 % vers les centres de données périphériques[29].

Le rapport propose une approche holistique et une vue intégrée des flux de données d'entreprise vers les centres de données cloud et les centres périphériques au sein de l'UE ; entre l'UE et le Royaume-Uni ; et entre les pays de l'UE et de l'AELE. Il capture plusieurs dimensions des flux de données, à savoir les secteurs[30] (secteurs NACE CN)

25 En 2018, la valeur de l'économie des données en Europe a atteint 2,4% du PIB européen, et on estime qu'elle atteindra près de 6% du PIB de l'UE en 2025. De plus, le volume mondial de données devrait croître cinq fois par rapport aux chiffres de 2018. en 2025. Source : Conseil de Lisbonne, International Data Corporation. « Rapport d'étude final : l'outil européen de surveillance du marché des données » (https://datalandscape.eu/study-reports/final-study-report-european-data-market-monitoring-tool-key-facts-figures-first-policy)

26 European Commission Communication, ‘A European strategy for data’, COM(2020) 66 final.

27 Ibid.

28 The European Data Flow Monitoring’ (https://digital-strategy.ec.europa.eu/en/policies/european-data-flow-monitoring)

29 L'hypothèse est basée sur la communication de la Commission européenne, « Une stratégie européenne pour les données», COM(2020) 66 final.

30 C Fabrication, D Electricité Gaz, Vapeur et climatisation, E Eau, F Construction, G Commerce de gros et de détail, H Transport et stockage, I Hébergement et restauration, J Information et communication, K Finance et assurances, L Réel activités immobilières, M Activités professionnelles, scientifiques et techniques, N Services administratifs et de soutien

déclenchant les flux, la taille de l'entreprise (petites, moyennes, grandes entreprises)[31], les services[32] et les activités d'entreprise à forte intensité de données[33] et les types de données (personnelles/non personnelles)[34]. L'étude n'a pas été en mesure d'examiner les flux de données entre le cloud et les centres de données de périphérie des entreprises fournissant des services cloud. Ces échanges peuvent survenir afin que les entreprises fournissant un service cloud puissent sauvegarder des données ou fonctionner plus efficacement. De plus, l'étude n'a pas examiné les flux de données vers l'infrastructure de cloud privé dans les locaux de l'entreprise. Cela fournit des données pour 2016 à 2020, avec des prévisions jusqu'en 2030.

La méthodologie présentée ici peut être utilisée à l'avenir pour surveiller les tendances des flux de données dans l'Union européenne (UE), entre l'UE et les pays de l'AELE et le Royaume-Uni à l'appui des décisions politiques, commerciales et d'investissement de l'UE. En outre, les résultats fournissent des preuves utiles pour étayer l'évaluation du règlement sur la libre circulation des données et les outils pour une analyse continue des flux de données et le développement économique du secteur de l'informatique de l'UE.

c. Flux de données globaux

Dans l'ensemble, dans les États membres de l'UE27, le flux de données provenant des entreprises[35] achetant des services cloud utilisés sur Internet était de 504 763 To/mois (5,8 EX/an) en 2020. D'ici 2030, le volume devrait être 15,2 fois plus important, à 7 669 835 To. /mois (87,8 EX/an) et d'ici 2025, le montant devrait être 4,4 fois supérieur (à la valeur de 2020) à 2 201 058 To/mois (25,2 EX/an).

Dans les 31 pays examinés[36], les flux de données vers les centres de données cloud et de périphérie par les entreprises achetant des services cloud utilisés sur Internet ont été estimés à 708 790 To/mois (8,1 EX/an) en 2020. D'ici 2030, le montant devrait être de 14,9 fois. supérieur à 10 577 600 To/mois (121,1 EX/an). D'ici 2025, le montant prévu devrait être 4,3 fois supérieur (valeur de 2020) à 3 025 388 To/mois (34,6 EX/an).

d. Analyse au niveau des pays

L'un des éléments de l'analyse consistait à estimer les flux de données d'entreprise (To/mois) vers les centres de données cloud principaux et périphériques des dix principaux fournisseurs de services cloud dans les pays possédant une infrastructure cloud entre 2016 et 2021[37]. Le plus grand volume de flux de données d'entreprise est servi par les centres de données cloud et de périphérie en Allemagne. En 2020, l'Allemagne a reçu

31 Définition Eurostat : petites entreprises : 10-49 personnes occupées ; entreprises moyennes : 50-249 personnes occupées ; petites et moyennes entreprises (PME) : 1 à 249 personnes occupées ; grandes entreprises : 250 personnes ou plus occupées.

32 Voir « Définitions » dans la section 1.2.

33 Les activités à forte intensité de données font référence aux données stockées dans les infrastructures cloud et générées par des activités fortement dépendantes de l'utilisation, de la production ou/et de la fourniture de la technologie cloud, telles que les données industrielles, qui sont l'une des données typiques à alimenter dans la relation client. activités de gestion (CRM). Pour plus d'informations, veuillez vous référer à l'annexe 1.

34 Le type de données fait référence aux données personnelles et non personnelles. Pour plus d'informations, veuillez vous référer à l'annexe

35 Dans cette étude, les entreprises sont définies comme les secteurs privés de la NACE (C à M) et les secteurs publics de la NACE (N à S). Les entreprises sont constituées majoritairement d'entreprises (voir référence 8) mais aussi de certains secteurs publics (N Services administratifs et de soutien, P Éducation, Q Santé humaine et travail social, R Arts et spectacles et loisirs, S Autres activités de services).

36 Lichtenstein n'est pas inclus

37 L'analyse se concentre sur 2016 à 2021 car les informations sur l'infrastructure cloud de ces années sont connues. De nombreux commentateurs prédisent une croissance considérable des centres de données en périphérie du cloud dans les prochaines années. L'extrapolation pour créer des prévisions dans l'avenir, lorsque la répartition de l'infrastructure changera considérablement, créerait des résultats erronés.

151 968 To/mois (1,74 EX/an) de flux de données cloud en provenance d'autres pays. Cela représente 30,7 % des flux de données cloud vers les États membres de l'UE27 et 25,9 % des flux de données cloud dans les 31 pays étudiés. L'État membre de l'UE27 recevant le deuxième plus grand afflux de données desservi par ses centres de données cloud principaux et périphériques en 2020 était les Pays-Bas (86 963 To/mois, 1 EX/an).

Ces deux pays (Allemagne et Pays-Bas) ont reçu plus de données d'entreprise circulant vers leurs centres de données cloud principaux et périphériques en 2020 que le Royaume-Uni (88 798 To/mois, 1,02 EX/an). En 2020, le Royaume-Uni a reçu 12,7% du total des données circulant vers les installations cloud et périphériques des 31 pays étudiés. Les pays de l'AELE ont reçu 3 % du total des données circulant vers les installations cloud et périphériques (15 506 To/mois, 0,18 EX/an).

L'analyse a découvert qu'en 2020, le flux de données le plus élevé (données circulant d'un pays vers un centre de données cloud et de périphérie) a été généré par le Royaume-Uni (152 815 To/mois, 1,75 EX/an), suivi de l'Allemagne (93 474 To/ mois, 1,07 EX/an) et France (75 039 TB/mois, 0,86 EX/an).

e. Flux nets de données par pays

Une autre étape d'analyse a examiné les données nettes d'entreprise circulant vers les centres de données cloud et périphériques pour les 31 pays examinés dans l'étude. Cela a été trouvé en soustrayant les flux de données vers un pays des flux de données générés par le pays. Trois pays enregistrent des flux nets de données d'entreprise dans leurs centres de données cloud et de périphérie. Tous les autres pays enregistrent des sorties nettes de données.

Le pays avec le plus grand afflux net de données dans ses centres de données cloud et de périphérie en 2020 était l'Allemagne. L'afflux net de données d'entreprise allemandes dans ses centres de données cloud et de périphérie en 2020 était de 58 494 To/mois (0,67 EX/an). L'Irlande a le deuxième plus grand flux net de données d'entreprise vers ses centres de données cloud et de périphérie (45 981 To/mois, 0,53 EX/an). Les Pays-Bas ont le troisième flux net de données d'entreprise le plus élevé dans leurs centres de données cloud et de périphérie en 2020 (45 504 To/mois (0,52 EX/an).

L'État membre de l'UE27 avec le plus grand flux net de données d'entreprise vers les centres de données cloud et de périphérie dans d'autres pays est l'Italie ; 42 923 To/mois (0,49 EX/an). Sur les 31 pays examinés, les sorties nettes de données d'entreprise les plus importantes ont été enregistrées par le Royaume-Uni ; 86 751 To/mois (1 EX/an).

f. Adoption des services cloud et prévisions jusqu'en 2030

En termes de secteurs, en 2020, les flux de données les plus importants vers le cloud proviennent du secteur de la santé (NACE Q, 12,9 % de tous les flux, 91 600 To/mois ; 1,048 EX/an), du commerce de détail et de gros (NACE G, 12,9 % pour cent, 91 400 TB/mois ; 1,046 EX/an) et Éducation (12,5 %, 88 600 TB/mois ; 1,01 EX/an). En 2025, on s'attend à ce que la vente au détail et en gros représente 13,1 % de tous les flux (390 300 TB/mois ; 4,47 EX/an), le secteur de la santé représentera 12,3 % de tous les flux (377 900 TB/mois ; 4,32 EX/ par an), et l'éducation représentera 12,1 % de tous les flux (367 800 TB/mois ; 4,21 EX/an). En 2030, on s'attend à ce que la vente au détail et en gros représente 13,2 % de tous les flux (1 370 670 To/mois ; 15,69 EX/an), la Santé représentera 12,1 % de tous les flux (1 293 560 To/mois ; 14,80 EX/an) , et l'éducation représentera 11,9% de tous les flux (1 262 730 TB/mois ; 14,45 EX/an).

De plus, en 2020, le plus grand volume de données circulant vers les principaux centres de données cloud et de périphérie provient d'entreprises de 250 employés ou plus. Dans les États membres de l'UE27, les entreprises de 250 employés ou plus ont déclenché 61 % de tous les flux de données vers le cloud (309 500 To/mois [3,54 EX/an] du total des flux de l'UE27 de 504 700 To/mois [5,78 EX/an] en 2020)[38]. Les entreprises des États membres de l'UE27 comptant moins de 10 employés représentent 8,5 % de tous les flux de données vers le cloud (42 870 To/mois ; 0,49 EX/an). Les entreprises des États membres de l'UE27 comptant de 10 à 49 employés ont réalisé 13,2 % de tous les flux de données vers le cloud (66 560 To/mois ; 0,76 EX/an). Les entreprises des États membres de l'UE27 comptant de 50 à 249 employés ont obtenu 17 % de tous les flux de données vers le cloud (85 800 To/mois ; 0,98 EX/an).

D'ici 2030, on s'attend à ce que 65 % des entreprises de l'UE27 comptant de 10 à 49 employés achètent des services cloud utilisés sur Internet. C'est presque le double du nombre en 2020 (34%) et supérieur aux 52% des entreprises de l'UE27 comptant 10 à 49 employés qui devraient acheter des services cloud utilisés sur Internet d'ici 2025.[39] 46 % des entreprises de l'UE27 comptant de 50 à 249 employés ont acheté des services cloud utilisés sur Internet en 2020. En 2025 et 2030, ce chiffre devrait atteindre 68 % et 83 %, respectivement[40]. Pour les entreprises de l'UE27 comptant plus de 250 employés, en 2020, 66% ont acheté des services cloud utilisés sur Internet. En 2025, ce chiffre devrait passer à 90 %, puis à 97 % d'ici 2030[41].

En termes de services[42], la proportion d'entreprises achetant des niveaux bas et moyens de services cloud utilisés sur Internet est en baisse. Un nombre croissant d'entreprises achètent des services cloud de haut niveau utilisés sur Internet. En 2020, 19 % des entreprises de l'UE27 utilisaient de faibles niveaux de services cloud, 30 % utilisaient des niveaux moyens de services cloud et 51 % utilisaient des niveaux élevés de services cloud. En 2025, il est prévu que 15 % des entreprises utiliseront des niveaux bas de services cloud, 22 % utiliseront des niveaux moyens de services cloud et 63 % utiliseront des niveaux élevés de services cloud. D'ici 2030, la situation devrait considérablement changer : 72 % des entreprises devraient utiliser des niveaux élevés de services cloud, 16 % devraient utiliser des services moyens et seulement 12 % achèteront de faibles niveaux de services cloud.

Concernant les types de données, 41 % du total des données stockées dans l'infrastructure cloud sont des données personnelles et 59 % sont des données non personnelles. Sur ces 41 pour cent de données personnelles, 11 pour cent sont générés par un utilisateur/individu. Concernant les données non personnelles, sur 59 pour cent, environ 22 pour cent sont générées par une machine. Parmi les données générées par machine, 19 % sont des données industrielles (productives), à savoir utilisées comme intrant direct

38 Dans les pays de l'AELE, les flux de données des entreprises de 250 salariés ou plus représentent 48 pour cent des flux (24 700 To/mois de 51 200 To/mois). Au Royaume-Uni, les flux de données des entreprises de 250 employés ou plus représentent 66 % des flux (100 500 To/mois de 152 800 To/mois).

39 AELE : 2020 – 63 %, 2025 – 90 %, 2030 – 98 %. Royaume-Uni : 2020 – 41 %, 2025 – 72 %, 2030 – 95 %.

40 AELE : 2020 – 81 %, 2025 – 97 %, 2030 – 99 %. Royaume-Uni : 2020 – 63 %, 2025 – 92 %, 2030 – 98 %.

41 EFTA: 2020 – 90 per cent, 2025 – 98 per cent, 2030 – 100 per cent. UK: 2020 – 77 per cent, 2025 – 87 per cent, 2030 – 97 per cent

42 Les définitions d'Eurostat des services de cloud computing (CC) à faible, moyen et haut niveau sont les suivantes ; les services CC faibles sont la messagerie électronique, les logiciels de bureau, le stockage de fichiers, les services CC moyens incluent les services CC faibles plus l'hébergement de la base de données des entreprises, les services CC élevés sont les applications logicielles de comptabilité, le logiciel CRM, la puissance de calcul.

dans le processus de production du point de vue de la chaîne d'approvisionnement et les 3 % restants sont indiqués comme « autres données » par les répondants à l'enquête.

1. Introduction

This report provides a summary of the work carried out to develop a forward-looking and agile methodology for mapping data flows within the EU, between the EU and the UK, and between the EU and EFTA countries. This study 'Mapping Data Flows, SMART 2019/0010' was commissioned by the Directorate-General for Communications Networks, Content and Technology (DG CNECT), Cloud & Software Unit, of the European Commission and carried out by VVA Economics & Policy and Tech4i2.

The overall objectives of the study are:

- To develop and test a new, self-sutained and replicable methodology to estimate data flows and data stocks.
- To provide a holistic approach and integrated views of enterprise data flowing to cloud and edge infrastructures:
  - within the EU;
  - between the EU and the UK;
  - between EU and EFTA countries.
- To capture several dimensions of the data flows, namely the sectors triggering the flows (NACE sectors C-N),[43] enterprise size (small, medium, large enterprises),[44] cloud service type[45] and cloud-intensive enterprise activities,[46] and data types (personal/non-personal).[47]
- To map and assess cloud data centre storage capacity and utilisation 'data stock' (i.e. data volume stored in cloud infrastructure) across the EU.

The following sections set the scene for the study by presenting the overall policy and economic context in which the study takes place, as well as the definitions of the key terms and concepts used throughout the report.

1.1 Why it matters: the bigger picture

In the last decade, data has become a critical factor of production. As mentioned by the European Commission Communication in a 'European Strategy for Data',[48] data is an essential resource for enterprises to develop new products and services, and its availability

43 Manufacturing, Electricity Gas, Steam and air conditioning supply, Water Supply, Construction, Wholesale and retail trade, Transportation and storage, accommodation and food service activities, Information and Communication, Financial and insurance activities, Real estate activities, Professional, scientific and technical activities, Administrative and support services

44 Eurostat definition: small enterprises: 10-49 persons employed; medium-sized enterprises: 50-249 persons employed; small and medium sized enterprises (SMEs): 1-249 persons employed; large enterprises: 250 or more persons employed.

45 See 'Definitions' in Section 1.2.

46 'Data intensive activities' refers to data stored in cloud infrastructures and generated by activities that are highly dependent to the use, production or/and provision of the cloud technology, such as industrial data, which is one of the typical data to feed into customer relationship management (CRM) activities. For more information, please refer to Annex 1.

47 'Data type' refers to personal and non-personal data. Personal data, refers to 'any information relating to an identified or identifiable natural person (…), directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person' (European Regulation 2016/679 of the European parliament and the Council, art. 4 defines personal data);Non-Personal data, refers to data as other than personal, (European Regulation 2018/1807 (COD) of the European Parliament and the Council).

48 European Commission Communication, 'A European strategy for data', COM(2020) 66 final.

has enabled more sophisticated forecasting techniques to allow better economic decisions. In this context, the digital transformation of the EU economy relies on the availability and uptake of secure, energy-efficient, affordable and high-quality data processing capacities. Cloud infrastructures and services, both in data centres and at the edge, could hence provide the ground for economic growth and innovation.

The Covid-19 pandemic has shown how pivotal the data economy is and the importance of data-driven businesses and decision-making. Governments, enterprises and international organisations have been relying on data more than ever to tackle the challenges brought on by the pandemic (the need for digital channels to treat the demand of vaccines, testing and treatments, increased digital interactions due to teleworking). Moreover, during the pandemic, data storage and data processing capacity have become crucial for enabling teleworking across the EU via the set-up of cloud platforms.

On the policy side, efforts at European level have been channelled into creating the right conditions to pave the way for a data-driven economy. Since 2014, the European Commission has been working on creating a **Digital Single Market**[49] where individuals and businesses can access and engage in online activities under conditions of fair competition and aligned with EU rules and values of privacy and data protection.

To remove the remaining key barriers to the movement of non-personal data and to address legal uncertainties spurred by new data technologies, the **Regulation on a Framework for the free flow of non-personal data**[50] entered into force in 2019 with the objective of improving the free flow of non-personal data across borders and making it easier to switch providers of data storage and processing services. Such Regulation complements the framework set by the GDPR covering personal data and aims at removing obstacles to the free movement of non-personal data between different EU countries.

Acknowledging the increasing role played by data in our economy,[51] the European Commission released in 2020 a Communication on a **European Strategy for data**[52] aiming at developing a common European data space, interconnecting cloud infrastructure, setting up a cloud service marketplace for EU users, as well as setting strategic EU investments in new technologies, such as edge computing, high-performance computing/quantum computing. One of the key actions of the European Data Strategy is to create a ‘European analytical framework for measuring data flows’,[53] given the strategic importance of monitoring data flows to inform EU decision-making and investment choices in the area of cloud computing.

Due to the interconnected way European enterprises (both businesses and public sector) operate, which goes beyond the EU’s borders, the importance of data flows for the competitiveness of the European economy is further acknowledged in the recent European

49 European Commission Communication, ‘A Digital Single Market Strategy for Europe’, COM(2015) 192 final.

50 European Commission, Regulation (EU) 2018/1807 on a framework for the free flow of non-personal data in the European Union.

51 In 2018, the value of the data economy in Europe reached 2.4 per cent of European GDP, and it is estimated to reach almost 6 per cent of EU GDP in 2025. Moreover, the global data volume is expected to grow five times the 2018 figures in 2025. Source: Lisbon Council, International Data Corporation. ‘Final Study Report: The European Data Market Monitoring Tool’ (https://datalandscape.eu/study-reports/final-study-report-european-data-market-monitoring-tool-key-facts-figures-first-policy)

52 European Commission Communication, ‘A European strategy for data’, COM(2020) 66 final.

53 Idem.

Commission Communication '**Digital Compass'**.[54] Building a vision for Europe's digital transformation, the Communication presents targets to radically transform the European economy and society by 2030. The target is to materialise the vision set out in the data strategy that is to have 80% of data processing done at the edge by 2025. In particular, by 2030, the ambition is that 75% of European enterprises have taken up cloud computing services, big data analytics, artificial intelligence and deployed 10,000 climate neutral highly secure hedge nodes.

Given the strategic importance of monitoring data flows to assess the competitiveness of the European digital economy, while fostering the principle of free movement of data across the EU, the European Commission requested the present study to deliver on one of the key actions of the European Strategy for Data.

The study which is part of the European Commission initiative 'The European Data Flow Monitoring',[55] provides an innovative method for estimating the volume and types of enterprise data flowing to cloud infrastructures and for investigating where data flows across the EU. Enterprise data flows to cloud and edge infrastructures is the movement of data from the end user to the data centre in question. For example, if a user was to use OVH services, the data would flow from user in the home country to the closest OVH cloud data centre. In this study, the assumption has been made that 80 per cent of cloud data flows to main cloud data centres and 20 per cent flows to edge data centres[56].

The report provides a holistic approach and an integrated view of enterprise data flowing to cloud data centres and edge centres within the EU; between the EU and the UK; and between EU and EFTA countries. It captures several dimensions of the data flows, namely the sectors[57] (NACE sectors C-N) triggering the flows, enterprise size (small, medium, large enterprises),[58] services,[59] and data-intensive enterprise activities[60] and data types (personal/non-personal).[61] The study was not able to investigate data flows between the cloud and edge infrastructures of the businesses providing cloud services. These exchanges might arise so that businesses providing cloud service can back up data or operate more efficiently. Furthermore, the study did not investigate data flowing to private cloud infrastructure on company premises.

The methodology presented here can be used in future to monitor data flow trends across the European Union (the EU), between the EU and EFTA countries and the UK in support of EU policy, trade and investment decisions. Furthermore, the findings provide useful evidence to support the evaluation of the Regulation of the free flow of data and the tools for a continuous analysis of data flows and the economic development of the EU's data processing sector.

54 European Commission Communication, '2030 Digital Compass: the European way for the Digital Decade', COM(2021)118 final.

55 'The European Data Flow Monitoring' (https://digital-strategy.ec.europa.eu/en/policies/european-data-flow-monitoring)

56 The assumption is based from the European Commission Communication, 'A European strategy for data', COM(2020) 66 final.

57 C Manufacturing, D Electricity Gas, Steam and air conditioning supply, E Water Supply, F Construction, G Wholesale and retail trade, H Transportation and storage, I Accommodation and food service activities, J Information and Communication, K Financial and insurance activities, L Real estate activities, M Professional, scientific and technical activities, N Administrative and support services

58 Eurostat definition: small enterprises: 10-49 persons employed; medium-sized enterprises: 50-249 persons employed; small and medium sized enterprises (SMEs): 1-249 persons employed; large enterprises: 250 or more persons employed.

59 See 'Definitions' in Section 1.2.

60 'Data intensive activities' refers to data stored in cloud infrastructures and generated by activities that are highly dependent on the use, production or/and provision of the cloud technology, such as industrial data, which is one of the typical data to feed into customer relationship management (CRM) activities. For more information, please refer to Annex 1.

61 'Data type' refers to personal and non-personal data. For more information, please refer to Annex 1.

1.2 Definitions

This report adopts a number of terms that are used throughout this report. For ease of reference, they are defined here:

Internet traffic and data flows

- The Eurostat term ***enterprise*** is used to describe ***private sector*** organisations (in NACE sectors C to M) ***and public sector*** organisations (NACE sectors N to S);[62]
- The term ***internet traffic*** is adapted from an International Telecommunication Union (ITU) definition and used to describe internet traffic generated by household and enterprise fixed-broadband subscribers, measured at the end user access point by adding upload and download traffic;[63].

- The term ***cloud based data flow*** is used to describe ***enterprise data flowing to cloud and edge data centres. It*** describes data flowing from enterprises in private and public sectors to cloud data or edge centres;[64]

Data

- **Personal data**: refers to 'any information relating to an identified or identifiable natural person (…), directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person' (European Regulation 2016/679 of the European parliament and the Council, art. 4 defines personal data);
- **Non-Personal data:** refers to data as other than personal, (European Regulation 2018/1807 (COD) of the European Parliament and the Council);
- **Mixed data**: consists of both personal and non-personal data;
- **Machine generated data**: Data recorded, collected or produced by connected devices, assets or networks independent of any human intervention; User generated data: data generated by human intervention

A series of additional terms need more extensive definitions in order to understand the approach developed in the study. In particular, it is important to introduce the concepts of cloud computing and edge computing, the types of cloud and edge data centres that are

62 Eurostat describe an 'enterprise' as 'an organisational unit producing goods or services which has a certain degree of autonomy in decision-making. An enterprise can carry out more than one economic activity and it can be situated at more than one location' (https://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Enterprise). Eurostat use the term enterprise for organisational units in NACE sectors C to S. The term 'enterprise' thus includes organisational units usually considered to be the private sector (NACE C. Manufacturing, D. Electricity, gas, steam and air conditioning supply, E. Water supply; sewerage, waste management and remediation activities, F. Construction, G. Wholesale and retail trade; repair of motor vehicles and motorcycles, H. Transportation and storage, I. Accommodation and food service activities, J. Information and communication, K. Financial and insurance activities, L. Real estate activities, M. Professional, scientific and technical activities. We adhere to the Eurostat approach of also describing organisational units in NACE sectors N to S (N. Administrative and support service activities, O. Public administration and defence; compulsory social security, P. Education, Q. Human health and social work activities, R. Arts, entertainment and recreation, S. Other service activities) as enterprises. This later group of enterprises are commonly regarded as 'public sector' organisations. The term 'enterprise' does not differentiate between the intermediate or ultimate country ownership of an enterprise. On some occasions the term 'business' is used to avoid the awkward repetition or juxtaposition of the term 'enterprise'.

63 ITU World Telecommunications/ICT Indicators Database Indicator i135tfb

64 The terms 'cloud based data flow' and 'enterprise data flowing to cloud and edge data centres' are used interchangeably in this report to avoid awkward repetition or juxtaposition of terms.

considered within the scope of the study, and the distinction between public and private cloud.

Cloud computing

Cloud computing is defined as the on-demand availability of computer system resources, especially data storage, processing (cloud storage and processing) and computing power, without direct active management by the user.[65]

All devices that need to access this data or use applications must first connect to the cloud, usually over the internet.[66] Since everything becomes centralised, the cloud is generally easy to secure and control while still allowing for reliable remote access from anywhere at any time.

The three most common types of cloud services are:

- **Infrastructure as a Service** (IaaS): IaaS provides users access to raw computing resources such as processing power, data storage capacity, and networking, in a secure data centre;[67]
- **Platform as a Service** (PaaS): Cloud provides infrastructure and a software framework (software development kits for various programming languages) where businesses can develop and run their own applications, such as web applications that can be created quickly and easily;[68]
  Software as a Service (SaaS): Cloud offers application-level services tailored to a wide variety of business needs, such as customer relationship management (CRM) and business analytics. SaaS is usually used for applications that need a lot of web or mobile access, for example mobile sales management software.[69]
- Eurostat defines low cloud computing services as email, office software and storage files. Medium cloud computing services is defined as email, office software, storage files, and hosting of the enterprise's database. High cloud computing services has been defined as accounting software applications, CRM software, and computing power, by Eurostat.
- The Eurostat definition '***buying cloud computing services used over the internet***' is used to describe an enterprise that has purchased and uses/utilises cloud services.[70]

Cloud computing can have some limitations, in particular increased latency of access to information when the cloud data centre is located a long way from the user.[71] When it comes to time-critical applications requiring low latency, such as communicating with

65 Montazerolghaem, A. et al., 2020. 'Green Cloud Multimedia Networking: NFV/SDN Based Energy-Efficient Resource Allocation'. IEEE Transactions on Green Communications and Networking. 4,3. p873–889.

66 Gyarmathy K., 2019. 'Edge Computing vs. Cloud Computing: What You Need to Know' (https://www.vxchnge.com/blog/edge-computing-vs-cloud-computing). Enterprises can develop their own 'private' cloud infrastructure located on their own premises accessed via a local network. In this case connectivity is not 'over the internet'. The Eurostat definition to describe an enterprise that utilises cloud services is an enterprise 'buying cloud computing services used over the internet'. It is therefore possible that enterprises with cloud usage independent of the internet might be under-represented in Eurostat data.

67 Akamai. 2021. 'Cloud computing services' (https://www.akamai.com/uk/en/resources/cloud-computing-services.jsp)

68 LeadingEdge. 2021. 'What are the types of cloud computing?' (https://www.leadingedgetech.co.uk/it-services/it-consultancy-services/cloud-computing/what-are-the-types-of-cloud-computing/)

69 Dynamix. 2021. 'What are the different types of services offered by cloud computing?' (https://www.dynamixsolutions.com/what-are-the-different-types-of-services-offered-by-cloud-computing/)

70 Eurostat. 2021. 'Cloud computing - statistics on the use by enterprises' (https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Cloud_computing_-_statistics_on_the_use_by_enterprises)

71 Apriorit. 2021. 'What Is the Difference Between Cloud Computing and Edge Computing?' (https://www.apriorit.com/dev-blog/531-cloud-edge-computing)

autonomous vehicles, drones and augmented reality, there is a need for a faster, real-time and more flexible solution.[72] Edge computing can be one of the responses.

Edge computing

Edge computing is the technology moving data storage, processing and analysis of time-critical applications away from a central cloud network closer to the edge, where the user (which can be a person, a machine, or a device) is located. This reduces the travel time of data, allowing processing to happen in real-time, effectively reducing latency closer to zero.[73] Edge computing also reduces some of the amount of enterprise data flowing to cloud data centres that might previously have flowed across internet networks to main cloud centres in more distant locations.[74]

Edge computing devices are designed for quick data processing on-site and place less emphasis on data storage. Cloud computing, in contrast, is based on considerably larger cloud data centres designed for huge amounts of data storage and data processing.

The 'edge' is typically defined as 'where the physical meets the digital' - the locations where the users are, where data and information is consumed, generated and transferred between users and cloud applications.[75] For cloud service providers, this means the edge of their cloud network where data is exchanged with users, partners and other systems.

Cloud service providers are driving the edge closer to customers through boosting edge computing capabilities. This is possible due to the increasing sophistication of data centre management software (DCIM)[76] and growth in highly interconnected networks[77]. These developments help to ensure quality of service and place computer power closer to users and end-point devices.

Undertaking data processing and storage at the edge also has the advantage of moving the requirement for data processing and storage tasks away from some mobile devices and 'internet of things' (IoT) main processors. Computing and data processing are undertaken at the edge, not within the IoT device. This may help to save power and provides opportunities to increase performance.[78]

A report addressed to the European Commissioner for the Internal Market in 2021 provided further dimensions and locations to define the **'near' and 'far' edge**, see Figure 1. The report defines the'near edge' as typically some hundreds of kilometres away from the customer.[79] The 'far edge' is described as being at a high proximity to customers and premises, within a range of 100km.

72 The time lag is generally only a few hundred milliseconds, but it makes a lot of difference in time-sensitive applications like autonomous cars, where real-time data is required for seamless functionality. Furthermore, a large quantity of data travelling back and forth over the network puts significant strain on bandwidth, further compounding the problem.

73 Srivastava A. 2021. 'What's the difference between edge computing and cloud computing?' (https://www.quora.com/Whats-the-difference-between-edge-computing-and-cloud-computing)

74 Priyanka V. 2019. 'Demystifying edge vs. Cloud computing' (https://dzone.com/articles/demystifying-the-edge-vs-cloud-computing)

75 Barras. D. 2020. 'How to speak like a data centre geek: Digital Edge vs Mobile Edge vs Micro Edge' (https://blog.equinix.com/blog/2020/08/07/how-to-speak-like-a-data-center-geek-digital-edge-vs-mobile-edge-vs-micro-edge/)

76 Gartner Glossary. 2021. https://www.gartner.com/en/information-technology/glossary/data-center-infrastructure-management-dcim

77 Wu C and Buyya R. 2015. Data Centre Networks. https://www.sciencedirect.com/topics/computer-science/interconnected-network

78 Chandra et al. 2016. 'Offloading to improve the battery life of mobile devices'. IEEE Pervasive Computing. 15, 4. (https://www.researchgate.net/publication/309452510_Offloading_to_Improve_the_Battery_Life_of_Mobile_Devices)

79 Airbus et al. 2021. 'European industrial technology roadmap for the next generation cloud-edge offering' (https://ec.europa.eu/newsroom/repository/document/2021-18/European_CloudEdge_Technology_Investment_Roadmap_for_publication_pMdz85DSw6nqPppq8hE9S9RbB8_76223.pdf)

Main cloud data centres

The main difference between a data center and a cloud data centre is where data is stored. In a data center, data is most often stored on the premises of an organization. The cloud is completely off premises and data is accessible from anywhere via the internet[80].

The **main cloud data centres** are able to achieve considerable financial economies of scale in providing large data storage capacity and data processing.[81] They represent the core of the cloud computing network and provide important computing power and storage capacity.[82]

For this study, we estimate there are 20 main cloud data centres in EU27 Member States and six in EFTA countries and the UK.[83] Main cloud data centres can expand large data storage capacities and computing processing for users quickly[84]. They offer users functionality and scalability in a quicker and more cost-effective way than purchasing hardware and undertaking complex data management *in situ*.[85]

These main cloud data centres, sometimes called **hyperscale cloud centres[86]**, require large financial investment around many hundreds of millions of euros to construct.[87] They usually have a minimum of 5,000 servers linked with an ultra-high-speed fibre network connection. Power is the largest operational expense.[88]

Edge data centres

Edge data centres are yet difficult to define given edge computing technologies are currently still in their infancy. The general consensus viewpoint is that **edge data centres** are smaller facilities (than main cloud centres) that extend the edge of the network to deliver cloud computing resources and cached streaming content to local end users.[89] Since they are positioned closer to end users, they can deliver services faster with reduced latency thanks to edge caching.

Edge servers, which cache content for localities, can help to ease the burden for main cloud data centres. If data, particularly higher bandwidth image files and video files, can be located closer to the user, the amount of time it takes to load a resource is reduced and it

---

80 Brown S. 2020. Cloud vs. Data Center: Which Is Right for Your Business? https://www.rutter-net.com/blog/cloud-computing-vs.-data-center-which-is-right-for-your-business

81 Waldhauser B. 2019. 'The edge and edge data centres' (https://www.dotmagazine.online/issues/on-the-edge-building-the-foundations-for-the-future/edge-data-centers)

82 Airbus et al (2021. Ibid.) estimate that large cloud provide around 30 MW. Section 2 examining cloud data centre capacity estimates average hyperscale cloud data centre capacity to be approximately 20 EX.

83 This estimate was derived from several websites, including TeleGeography (https://global-internet-map-2021.telegeography.com), CloudScene (https://cloudscene.com/browse/on-ramps), and DataCentreKnowledge (https://www.datacenterknowledge.com/cloud/telegeography-maps-world-s-cloud-data-centers). Cloud services providers were then contacted to validate estimations. Hyperscale cloud centres can have storage capacities of up to 10 EX or higher.

84 Users can start using cloud data services almost immediately after signing a contract.

85 Barras. D. 2020. Ibid.

86 In this study, hyperscale cloud centres will be referred to as 'main cloud data centres'.

87 Judge P. 2020. 'Amazon aims to fast track three data centres in North Virginia' (https://www.datacenterdynamics.com/en/news/amazon-aims-fast-track-three-data-centers-northern-virginia)

88 AFL Hyperscale. 2020. 'Understanding Different Types of Data Centre'. This study estimates hyperscale centres consumed 1.5 per cent of total global energy consumption in 2020 (https://www.aflhyperscale.com/understanding-different-types-of-data-center). Hyperscale cloud centres can have storage capacities of up to 10 EX or higher

89 Ciena. 2021. 'What is edge cloud?' (https://www.ciena.com/insights/what-is/What-is-Edge-Cloud.html). Data Foundry. 2020. 'Edge Data Centres Explained: What is Edge Computing?' (https://www.datafoundry.com/blog/edge-data-centers-explained-what-is-edge-computing). Gyarmathy K. 2019b. 'Checklist: How to Define an Edge Data Centre' (https://www.vxchnge.com/blog/what-is-an-edge-data-center). Sverdlik Y. 2015. 'How Edge Data Center Providers are Changing the Internet's Geography' (https://www.datacenterknowledge.com/archives/2015/08/26/how-edge-data-center-providers-are-changing-the-internets-geography).

reduces the volume of traffic flowing to main cloud data centres in more distant locations that might previously have flowed across internet networks.[90]

For IoT devices and internet networks, edge data centres also serve as clearinghouses to store or distribute data being generated. Edge data centres can also provide remote digital processing capabilities. By undertaking processing away from the IoT device power can be saved and batteries last longer in the device.[91]

Edge centres and locations are evolving. Some edge centres are being located inside or adjacent to internet exchange points through which most local traffic flows. **Internet exchange points** (IXPs) route internet traffic - they are supercomputers in terms of their processing power. They process and route millions of data packets every second and work to configure network traffic efficiently at major traffic points on the internet.

Telecommunication companies often provide internet services to their business and household telecommunication subscribers. They are located on the 'edge' and route their customers' internet traffic to internet exchange points and edge centres.

Micro data centres are smaller versions of edge data centres, usually about the size of a school locker,[92] which are ideal in remote locations. Micro data centres can be used to bring edge computing even closer to end users and undertake data processing to assist edge centres and minimise latency.

An edge node (or gateway node, or edge communication node) is a computer that acts as an end user portal for communication with other nodes in cluster computing[93]. This means users are able to complete tasks on the edge nodes rather than on the master nodes, which prevents capacity losses[94].

Telecommunications data centres

Telecommunications data centres are facilities owned and operated by Telecommunications or Service Provider companies. These data centres provide high connectivity and are mainly responsible for providing mobile services and driving content delivery[95] but not usually cloud computing services.[96] Having said that, some telecommunication companies have however established strategic agreements with cloud service providers to start providing edge service capabilities and offer their partners cloud services. Telecommunications businesses have also recently started using their telecommunications data centres to offer cloud services. Some telecommunications companies are also using space within their data centres to provide additional services such as colocation. It is worth noting that the provision of cloud computing services is not the primary function of telecommunications data centres. For that reason, they have not been included in the cloud and edge data centres in this study.

Colocation data centres[97] house privately-owned servers and networking equipment.[98] The equipment it provides (i.e. physical space and possibly servers, racks) to host cloud

90 Data Foundry. 2020. 'Edge Data Centres Explained: What is Edge Computing?' (https://www.datafoundry.com/blog/edge-data-centers-explained-what-is-edge-computing)

91 Gyarmathy K. 2019b. Ibid

92 Isberto M. 2020. 'What Is a Micro Data Centre?' (https://www.colocationamerica.com/blog/micro-data-center)

93 What Is. 2020. 'Edge Node.' (https://whatis.techtarget.com/definition/edge-node).

94 OVH Cloud. 2021. 'About edge nodes.' (https://docs.ovh.com/gb/en/analytics/edge-nodes/)

95 Content delivery is the process of delivering media over a medium, for example, the internet.

96 AFL Hyperscale. 2020. Ibid. The primary function of these data centres was not the provision of cloud computing services. They are not therefore included among the cloud and edge data centres examined in this study.

97 Colocation data centres are not taken into account in this study.

98 365Data Centres. 2021. 'What is colocation and how does it compare to the public cloud?' (https://www.365datacenters.com/portfolio-items/colocation-compare-public-cloud/)

computing capability is then leased in whole or part to customers. Colocation data centres can be leased to cloud service providers, telecommunication companies and/or hundreds or thousands of individual business customers.

Enterprise data centres

Instead of leasing equipment or servers in colocation data centres, some businesses create their own **enterprise data centre**[99]. Enterprise data centres are owned and operated by the company or a subsidiary of the main enterprise.[100] Some use external companies to undertake initial fit-outs and network installation before being maintained internally. They can be located on the company's premises ('on-premise') or off-site.

Other types of data centres

In addition to the types of data centres mentioned above, there are other data centres including Cloud On-Ramp and Points of Presents.

Some data centres that provide on-ramp access to cloud services that enable customers to have direct access to cloud service providers.[101] **Cloud On-Ramp**, sometimes known as a 'cloud exchange', is a direct connection service inside a cloud data centre that provides a business's IT network with direct connectivity to adjacent cloud service providers' edge computing infrastructure.[102] This is possible because cloud service providers rent space in cloud on-ramp data centres, rather than building their own data centre. This space is thus providing edge services.

Many cloud service providers are setting up **Points of Presence** (PoPs), also known as local hubs, in data centres and micro data centres in high internet usage areas to speed up their response to customer queries.[103] Points of presence are generally located near internet exchange points (IXPs)[104] with which they have peering agreements.[105] Proximity to IXPs (that route internet traffic) enables much faster response to customer queries.[106] As noted above, points of presence can be as simple as a server in a data centre or a micro data centre. In contrast, edge data centres provide points of presence, but with full deployment of data storage and data processing technology.

The roadmap developed by 27 European CEOs presented to the European Commissioner for Internal Market in May 2021[107] provides a useful overview of the parameters and geographies for key terms such cloud, near edge and far edge discussed earlier, see Figure 1. IoT devices and sensors are located to the left of the graphic, the main hyperscale data centres are located on the right.

99 Enterprise data centres are not taken account of in this study.

100 AFL Hyperscale. 2020. Ibid

101 Cloudscene. 2020. 'Cloud On-ramp' (https://cloudscene.com/glossary#cloudonramp)

102 Upstack. 2021. 'Cloud On-ramps and their benefits' (https://upstack.com/data-center/cloud-on-ramps-direct-connection-advantages)

103 Stackpath. 2020. 'Points of presence vs. edge locations' (https://blog.stackpath.com/point-of-presence/)

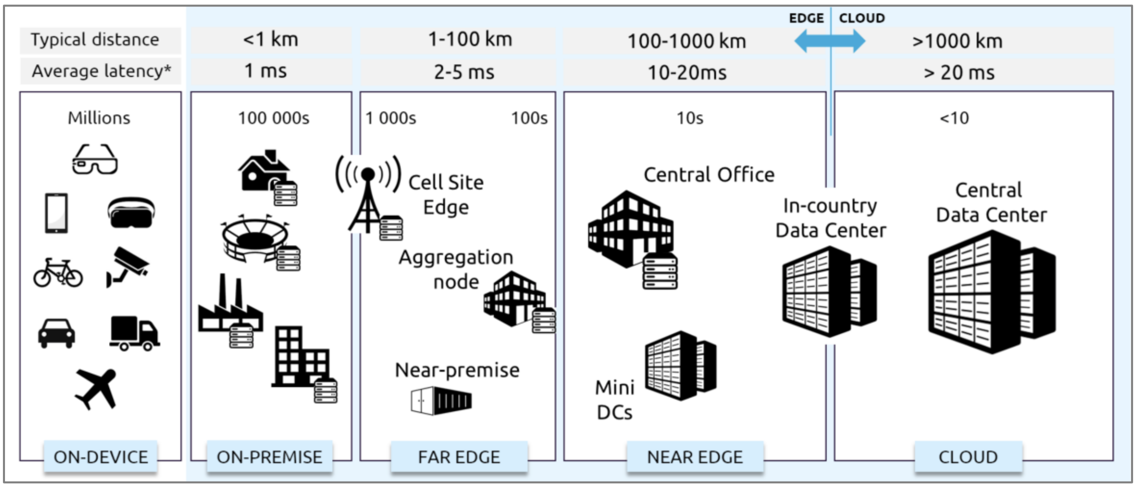
104 Cloudfare. 2020. 'What is an Internet exchange point?' (https://www.cloudflare.com/en-gb/learning/cdn/glossary/internet-exchange-point-ixp/)

105 Peering involves the interconnection of administratively separate internet and/or cloud networks for the purpose of exchanging traffic between the users of each network. NetNod. 2020. 'What is an IXP and what is peering?' (https://www.netnod.se/ix/what-is-an-ixp-and-what-is-peering)

106 Rahi Systems. 2021. 'Point of presence micro data centres' (https://www.rahisystems.com/blog/pop-point-of-presence-micro-data-centers/)

107 Airbus et al. 2021. Ibid.

Figure 1: Characteristics and geographies for IoT, edge and cloud data centres



Source. Roadmap. 2021. page 12[108]

Public, private and hybrid cloud

A **public cloud** is a cloud service offered to multiple customers by a cloud provider, such as OVH Cloud.[109] Public cloud service providers may offer cloud-based services such as infrastructure as a service (IaaS), platform as a service (PaaS), or software as a service (SaaS) to users for either a monthly or pay-per-use fee, eliminating the need for customers to host these services on site and/or in their own data centre.[110]

Private cloud is cloud resource solely dedicated to a single end use.[111] Most private clouds are on-premise.[112] Organisations are now building private clouds on rented, vendor-owned data centres located off-premise.[113]

Hybrid cloud enables users to combine private and public cloud. Hybrid cloud integrates at least one on-premise private cloud with at least one public cloud service. This combination provides enterprises with flexibility to switch workloads to the most efficient platform as needed[114].

Overview of the report

This section has provided an introduction to the main policy initiatives and key definitions in the fields of data, data flows, cloud and edge computing, the main types of cloud data centres, and public and private cloud used in the study and in this report.

108 Latency does not depend only on distance. Other factors influencing latency are: i) Access technology (5G and FTTH are much lower than 4G); ii) transport topology and technology; iii) core network configuration (user plane location, breakout point); iv) network optimisation (traffic prioritisation, bandwidth allocation and edge node selection).

109 Cloudflare. 2020. 'What is public cloud? Public vs. private cloud' (https://www.cloudflare.com/learning/cloud/what-is-a-public-cloud)

110 Vmware.2020. 'What is public cloud?' (https://www.vmware.com/topics/glossary/content/public-cloud)

111 Cloudflare. 2020. 'Private cloud vs. public cloud' (https://www.cloudflare.com/learning/cloud/what-is-a-private-cloud). TechTarget. 2021. 'What is private cloud?' (https://searchcloudcomputing.techtarget.com/definition/private-cloud)

112 The Eurostat definition to describe an enterprise that uses cloud services is an enterprise 'buying cloud computing services used over the internet'. Strict adherence to this definition would exclude an enterprise with private cloud infrastructure located on its own premises accessed via a local network. It is possible private cloud usage might be under-represented in Eurostat data for enterprises with on-premise private cloud.

113 Citrix. 2021. 'What is private cloud?' (https://www.citrix.com/en-gb/glossary/what-is-private-cloud.html)

114 Vennam, S. 2019. IBM. 'Hybrid Cloud'. (https://www.ibm.com/uk-en/cloud/learn/hybrid-cloud).

The study investigates enterprise data flowing to and from cloud data and edge centres from businesses ‘buying cloud computing services used over the internet’. While there was a desire to understand the volumes of data flowing between the cloud data centres owned by different cloud service providers[115], this information was not forthcoming in interviews cloud service provider stakeholders[116]. The study therefore does not look at data flowing between cloud data centres.

The remainder of the report is structured as follows:

- Chapter 2 provides an overview of the **key findings** of the study for enterprise data flows to cloud and edge data centres in the EU.
- Chapter 3 examines **enterprise data flowing to cloud data centres**. Information was not available to enable mapping of data flowing between two or more cloud data centres.
- Chapter 4 investigates where some of the different types of cloud infrastructure are located. Analysis of sources, including cloud service providers, provides details of the locations of 26 **cloud centres** and 73 **edge data centres** in the 32 countries examined in this study to understand where cloud data flows to.
- Chapter 5 summarises the lessons learned over the course of the study especially with regard to methodological aspects.
- Chapter 6 summarises the main conclusions of the study.

Among the edge data centres considered in the mapping of cloud infrastructure, some might be on colocation or on-ramp. Most telecommunications companies in Europe (with the exception of Cloud Provider 8) have a market share of less than 1 per cent across Europe and have therefore been excluded from analysis[117]. The other types of data centres such as enterprise data centres and telecomunication data centres are not considered in the mapping due to lack of information or smaller market shares[118].

The report gives an overview of enterprise data flows to cloud and edge data centres in the EU27 Member States, EFTA countries and the UK. Analysis is disaggregated by NACE sector, enterprise size, level of cloud service and type of data (personal and non-personal).

Enterprise data flow to cloud and edge data centres is the movement of data from the end user to the data centre in question. For example, if a user was to use OVH services, the data would flow from a user in the country where they were located to the closest OVH cloud data centre. In this study, the assumption has been made that 80 per cent of cloud data flows to main cloud data centres and 20 per cent flows to edge data centres[119].

In addition, the report gives some details on the location of the main cloud and edge data centres, their capacity and their utilisation. This allows one to understand where data is flowing to from the EU27 Member States, EFTA countries and the UK. Further details about cloud and edge data flows between countries can be found in section 4.1. This provides analysis and estimates of the cloud data centre capacity and utilisation.

115 For backup purposes and to redistribute the load of different data centres.

116 Flow between data centres for exchange and back up purposes is known to take place. This would be additional to cloud data flows examined in this study. It would thus be interesting to know the magnitude of these flows.

117 They are generally concern particular localities, information is sparse and by agreement with the project officer this study does not have the resources to undertake this fine grained level of analysis.

118 Smaller market share means that the proportion of cloud data flows by these means only account for a very small amount and thus by excluding them would not make a significant difference to the overall results of the study.

119 The assumption is based from the European Commission Communication, ‘A European strategy for data’, COM(2020) 66 final.

2. Overview of data flows

2.1 Data flows in Europe

This section provides an overview of more fulsome results presented in section 1.

As specified in Section 1.2 above, enterprise is used throughout this report to describe private sector organisations (in NACE sectors C to M[120]) and public sector organisations (NACE sectors N to S[121]).

Analysis estimates that flow of data stemming from enterprises buying cloud services used over the internet was 708,790 TB/month (8.1 EX/year) in 2020 across EU27 Member States, EFTA countries and the UK. This is expected to increase to 3,025,388 TB/month (34.6 EX/year) in 2025 and 10,577,600 TB/month (121.1 EX/year) in 2030. Across the 31 countries examined, **by 2030 the amount is forecast to be 14.9 times** greater than the estimated 2020 value.

In 2020, for EU27 Member States, it was calculated that there was 504,763 TB/month (5.8 EX/year) of data flowing to cloud and edge data centres. **By 2030, the amount is forecast to be 15.2 times greater at 7,669,835 TB/month (87.8 EX/year)**. This estimate is based on data from 26 main cloud data centres and 73 edge data centres across the EU27 Member States, EFTA countries and the UK[122]. In between these two years (in 2025), EU27 Member States data flows is forecasted to be 2,201,058 TB/month (25.2 EX/year).

Predications state that the EFTA countries will be 12 times greater from 2020 (51,211 TB/month [0.59 EX/year]) to 2030 (616,687 TB/month [7.1 EX/year]), in terms of data flowing to cloud and edge data centres. In 2025, it is expected that EFTA countries will have 193,390 TB/month (2.2 EX/year) of data flowing to cloud and edge data centres.

The UK is forecasted to experience a significant increase in data flowing to cloud and edge data centres from 2020 to 2030 (15 times greater). In 2020 data flows was equal to 152,815 TB/month (1.7 EX/year), in 2025 630,940 TB/month (7.2 EX/year), and in 2030 2,291,077 TB/month (26.2 EX/year).

See Table 1 below for further details of total flows of data to cloud and edge data centres in 2020.

120 C Manufacturing, D Electricity and Gas, E Water Supply, F Construction, G Wholesale and Retail, H Transportation, I Accommodation and Food, J ICT, K Finance and Insurance Activities, L Real Estate, M Professional, Scientific and Technical.

121 N Administrative and Support Activities, P Education, Q Human Health and Social Work, R Arts, Entertainment, Recreation, S Other Service Activities.

122 Full list is displayed in Figure 9 (page 34) and Figure 23 (page 82).

Table 1: Overview of flow of data to cloud and edge data centres in 2020, 2025 and 2030

	2020 Flow to cloud (TB/month)	2025 Flow to cloud (TB/month)	2030 Flow to cloud (TB/month)
EU27 Member States	504,763	2,201,058	7,669,835
EFTA	51,211	193,390	616,687
UK	152,815	630,940	2,291,077
Total EU27 EFTA UK	708,790	3,025,388	10,577,600
Belgium	24,567	94,965	307,673
Bulgaria	2,704	12,447	53,605
Czech Republic	9,290	44,533	164,557
Denmark	23,023	89,898	281,185
Germany	93,474	424,732	1,583,855
Estonia	3,061	14,383	46,130
Ireland	10,595	43,096	148,570
Greece	3,296	15,916	65,304
Spain	43,096	186,759	682,122
France	75,039	335,685	1,180,543
Croatia	2,667	13,121	48,180
Italy	58,974	286,944	931,104
Cyprus	683	3,343	12,864
Latvia	1,451	7,250	30,524
Lithuania	3,620	16,513	59,092
Luxembourg	1,403	5,410	20,133
Hungary	6,350	29,050	100,920
Malta	945	4,639	15,851
Netherlands	41,459	165,765	550,383
Austria	10,388	47,586	170,350
Poland	18,470	86,978	298,755
Portugal	8,964	36,112	123,077
Romania	5,073	28,410	127,238
Slovenia	2,611	10,994	39,412
Slovakia	2,852	13,010	48,445
Finland	17,469	62,462	196,263
Sweden	33,238	121,056	383,699
Iceland	1,294	4,869	15,566
Norway	17,067	63,655	198,783
Switzerland	32,851	124,866	402,338
United Kingdom	152,815	630,940	2,291,077

In 2020, EU27 Member States data flows to cloud and edge data centres are 10 times greater than EFTA countries and 3 times greater than the UK. By 2030, the EU27 Member States data flows will be 12 times higher than EFTA countries and 3 times higher than the UK. This implies that the EU27 Member States and the UK will be growing at similar rates, however the EFTA countries will be growing at a slightly slower pace.

2.2 Data flow to cloud and edge data centres and the adoption of cloud computing services by sector

Eurostat provides information about enterprises buying cloud computing services used over the internet disaggregated in number of ways. This section presents the results of analysis examining differences in data flow to cloud and edge data centres and the adoption of cloud computing services by sectors.

Figure 2 provides an overview of the proportion of data flowing to cloud and edge data centres for NACE sectors C to S. **Largest data flows to cloud come from the Health sector**[123] (NACE Q, 12.9 per cent of all flows, 91,600 TB/month; 1.048 EX/year**), Retail and Wholesale** (NACE G, 12.9 per cent, 91,400 TB/month; 1.046 EX/year) and **Education** (12.5 per cent, 88,600 TB/month; 1.01 EX/year). Followed by Public administration, Professional services and ICT sectors.

Conversely, the smallest data flows arise from Water Supply (NACE E 0.5 per cent, 3,500 TB/month; 0.04 EX/year), Electricity and Gas (NACE D 0.5 per cent, 3,800 TB/month; 0.04 EX/year) and Real Estate (NACE L, 0.6 per cent, 4,200 TB/month; 0.05 EX/year).

Figure 2: Data flow to cloud and edge data centres by sectors in 2020

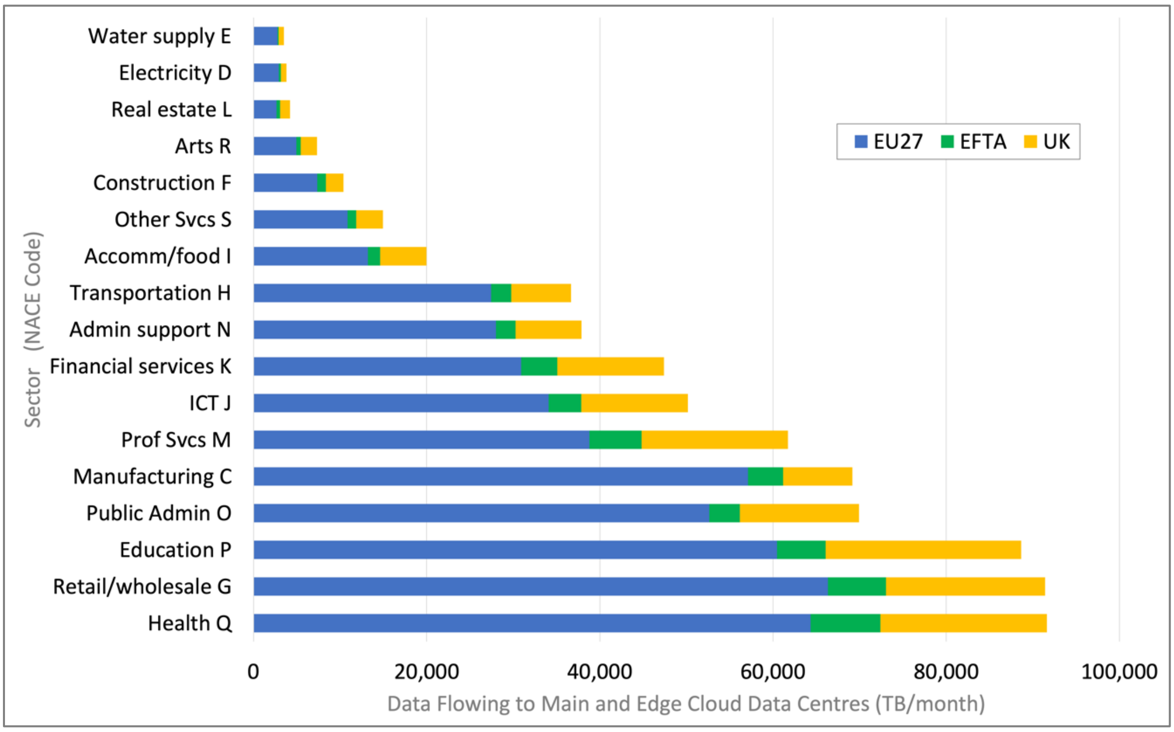
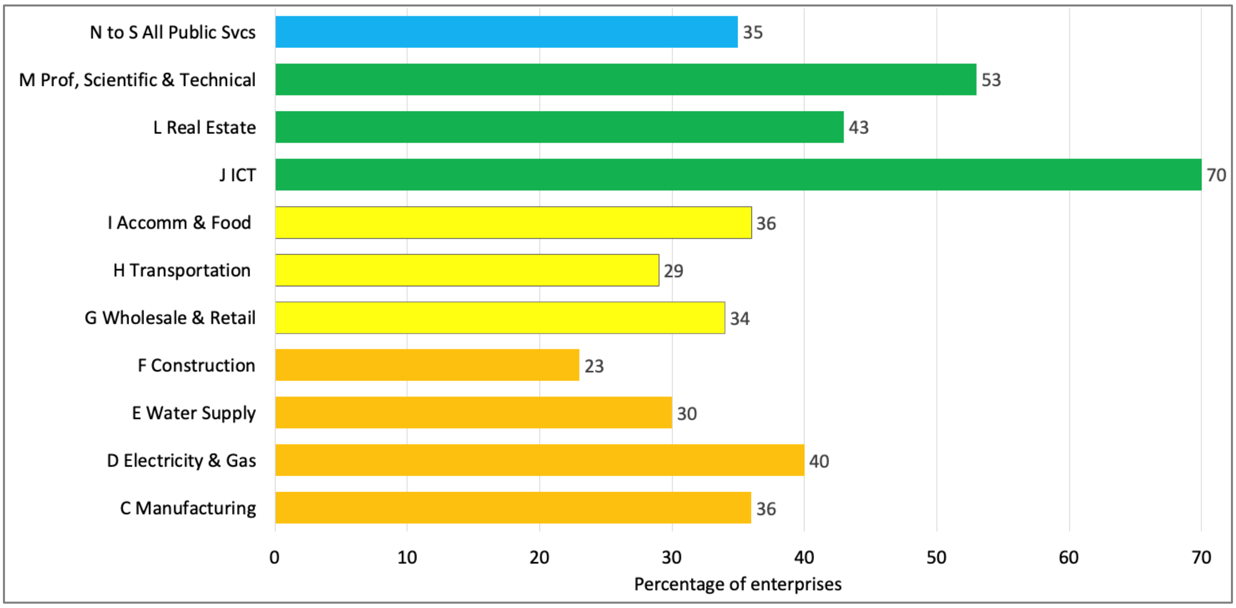


Figure 3 provides an overview of the proportion of enterprises operating in the EU27 Member States utilising cloud services in business and public sectors (by NACE codes C to S). **The ICT sector (70 per cent) has far higher levels of cloud utilisation than the averages for the EU** (36 per cent in 2020), EFTA (64 per cent) and UK (53 per cent). Construction has the lowest levels of enterprises buying cloud services used over the internet. Lowest levels of cloud utilisation are found in the Construction (23 per cent of enterprises) and Transportation sectors (29 per cent).

123 This includes hospitals, medical and dental practices, residential nursing and care, social work without accommodation, and child day-care.

The ICT sector (NACE sector J) in Figure 3 has the highest percentage levels of enterprises buying cloud computing services used over the internet (70 per cent). But due to its relatively small size in terms of number of employees (6.48 million in 2020) it does not have the highest volume of data flowing to main cloud and edge centres (34,080 TB/month in 2020). For example, Figure 3 shows that the health sector (NACE Q, included in the public services group NACE N to S) has 35 per cent of enterprises buying cloud computing services used over the internet. Due to the large size of the sector (21.5m employees in 2020), in comparison with ICT, it has far higher volumes of data flowing to main cloud and edge centres (64,300 TB/month in 2020)[124].

Figure 3: Cloud service adoption by enterprises in different sectors in EU27, EFTA and UK 2020



Source: Eurostat data (i soc_cicce_use)

2.3 Data flow to cloud and edge data centres and the adoption of cloud computing services by enterprise size

The previous section noted that Eurostat provides information about enterprises buying cloud computing services used over the internet disaggregated in number of ways. This section presents the results of analysis examining differences in data flow to cloud and edge data centres and the adoption of cloud computing services by enterprise size (employees).

The findings on the proportion of enterprises in different employee size bands highlight that in the EU27 Member States, the vast majority of enterprises are small – 96.9 per cent of enterprises have less than 20 employees.[125]

Large enterprises with 250 or more employees comprise 0.2 per cent of enterprises but they employ 35.1 per cent of the workforce. As Figure 5 shows, it is also known that larger

124 Employment statistics for 2020 are derived from SBS: Annual enterprise statistics by size class for special aggregates of activities (NACE Rev. 2) [SBS_SC_SCA_R2__custom_613220].

125 Similar circumstances prevail in EFTA and the UK.

enterprises with over 250 employees have higher levels of cloud service adoption. Thus **the largest amounts of data flowing to main cloud and edge data centres come from enterprises with 250 or more employees** (see Figure 4). In the **EU27 Member States these enterprises created 61 per cent of all data flows to cloud and edge** (309,500 TB/month [3.54 EX/year] of total EU27 flows of 504,700 TB/month [5.78 EX/year] in 2020).[126] In 2025, enterprises with over 250 employees accounted for 59 per cent of data flowing to main cloud and edge centres (1,298,100 TB/month [14.86 EX/year] of 2,201,000 TB/month [25.19 EX/year]). In 2030, enterprises with over 250 employees accounted for 56 per cent of data flowing to main cloud and edge centres (4,266,600 TB/month [48.82 EX/year] of 7,669,800 [7.31 EX/year]).

Figure 4: Data flow to cloud and edge data centres by different sizes of enterprise in 2020

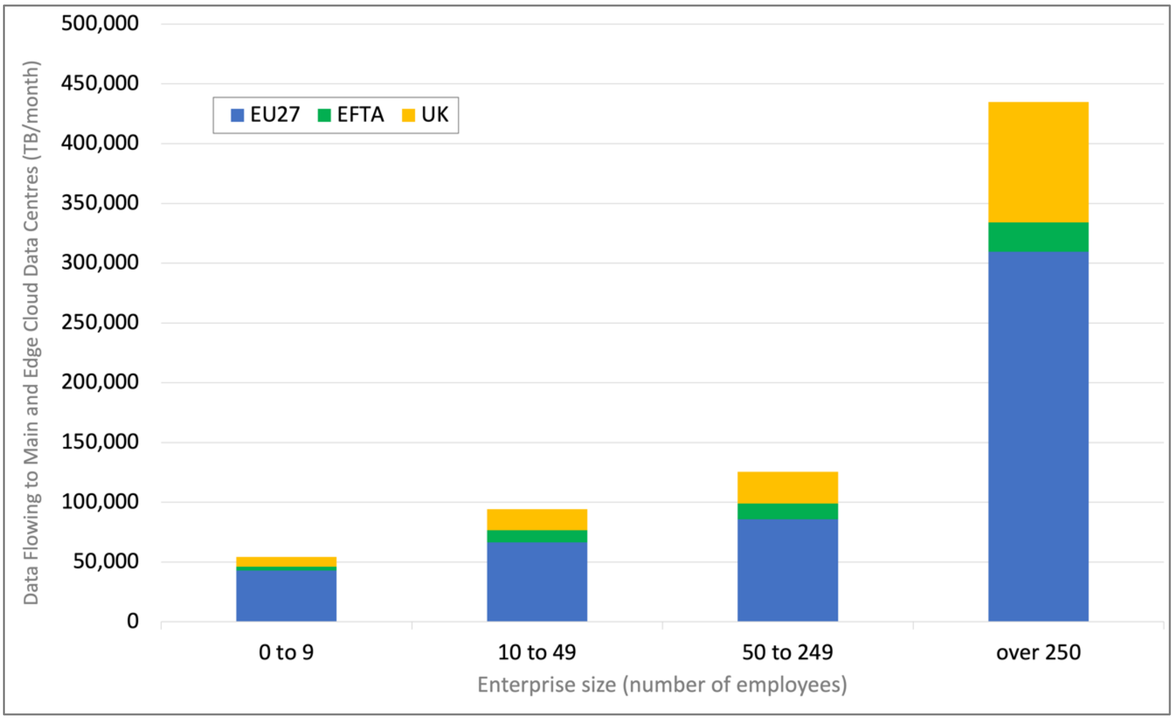


Figure 4 shows how **data flows are highest among businesses with over 250 employees.**

Figure 5 shows historic data from Eurostat between 2016 and 2020. Analysis extrapolated values from 2021 to 2030[127].

In Figure 5, data relating to EU27 Member States is presented in blue. The light coloured line is for enterprises with 20 to 49 employees - in 2020 34 per cent of these EU27 enterprises were buying cloud services used over the internet. The blue shading darkens as enterprise sizes become larger, for example in 2020, 46 per cent of enterprises with 50 to 249 employees were buying cloud services. Finally, the dark blue line shows that in 2020 66 per cent of the largest enterprises with over 250 employees were buying cloud services used over the internet.

126 In EFTA countries data flows from enterprises with 250 or more employees comprise 48 per cent of flows (24,700 TB/month of 51,200 TB/month). In the UK data flows from enterprises with 250 or more employees comprise 66 per cent of flows (100,500 TB/month of 152,800 TB/month).

127 The methodological annex provides an overview of extrapolation methods.

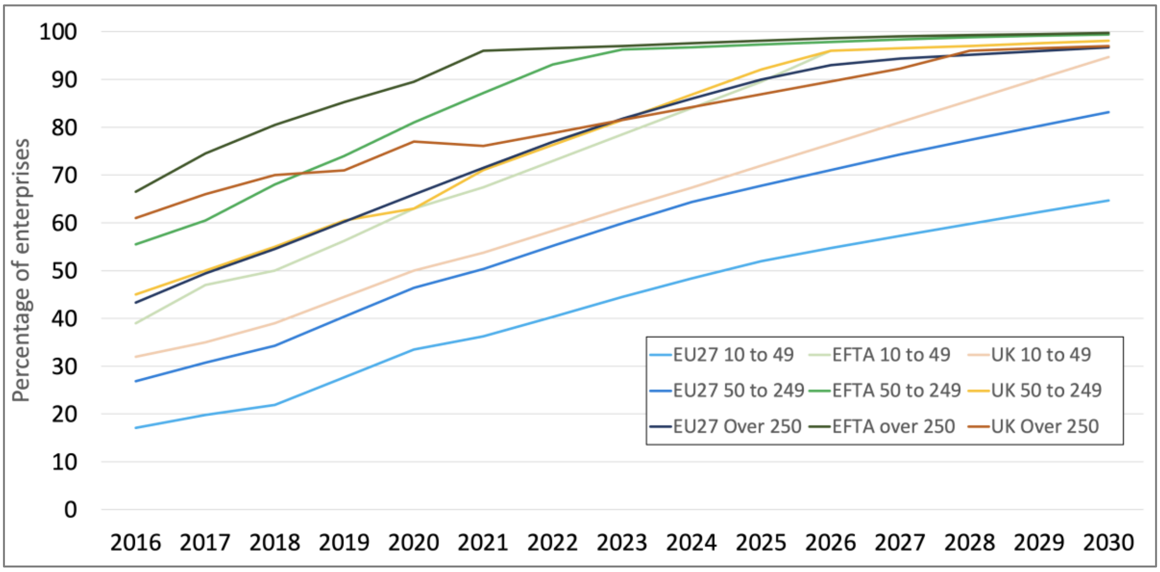
In 2030 it is expected that 65 per cent of EU27 enterprises with 10 to 49 employees will buy cloud services used over the internet. This is almost twice the number in 2020.

Cloud service use by medium sized enterprises with 50 to 249 employees is expected to reach 83 per cent in 2030, 1.8 times greater than the 2020 figure. Among the largest EU27 enterprises (with 250 or more employees) cloud use is expected to almost reach saturation point at 97 per cent in 2030, this is almost 1.5 times the number in 2020.

Initially higher levels of cloud services use can be seen in EFTA countries against EU27, all lines for these countries are shaded green. In enterprises with 10 to 49 employees in the three EFTA countries cloud adoption was 63 per cent in 2020. 81 per cent of enterprises with 50 to 249 employees have adopted cloud services and 90 per cent of enterprises with over 250 employees have adopted cloud services. In 2025, for these three different enterprise sizes adoption levels are forecasted to be at 90 per cent, 97 per cent, and 98 per cent, respectively. For all sizes of EFTA enterprises cloud adoption is expected to exceed 98 per cent in 2030.

Similar growth levels can be seen for the UK. UK lines are shaded brown. In UK enterprises with 10 to 49 employees, cloud adoption was 50 per cent in 2020. Growth to 72 per cent and 95 per cent was predicted for 2025 and 2030, respectively. 63 per cent of enterprises with 50 to 249 employees was estimated to have adopted cloud services in 2020. It is expected that this will increase to 92 per cent in 2025 and 98 per cent in 2030. Enterprises with over 250 employees have higher levels of cloud service adoption for 2020 (77 per cent), but more similar levels in 2025 and 2030 (87 per cent and 97 per cent, respectively).

Figure 5: Cloud service adoption by enterprises of different sizes in EU27, EFTA and UK



Source: Eurostat data (i soc_cicce_use) and study team extrapolation[128]

128 Linear extrapolation was undertaken to forecast values of different sizes of enterprise to 2030. Some of these extrapolations led to the creation of values greater than 100 per cent. Clearly this is unrealistic as percentage values can never be more than 100. To adjust these errors, data from s-shaped adoption curves was utilised. The data represents the flatter parts of the s-shaped curve at the start and end technology adoption. The utilisation of s-shaped methods ensured values were never greater than 100 per cent.

2.4 Data flow to cloud and edge data centres and the adoption of cloud computing services by level of service use

Previous sections examined data flow to cloud and edge data centres and the adoption of cloud computing services by NACE sectors and enterprise size. This section presents the results of analysis examining differences Eurostat data about the level of cloud services used by enterprises.

Figure 6 shows a dichotomy in the adoption of high level cloud services, medium level services and low level services.[129] **The graphic shows that the proportion of enterprises using low and medium level cloud services is declining for the EU27 Member States, EFTA countries and the UK since 2018 as more and more enterprises buy high level cloud services used over the internet.**

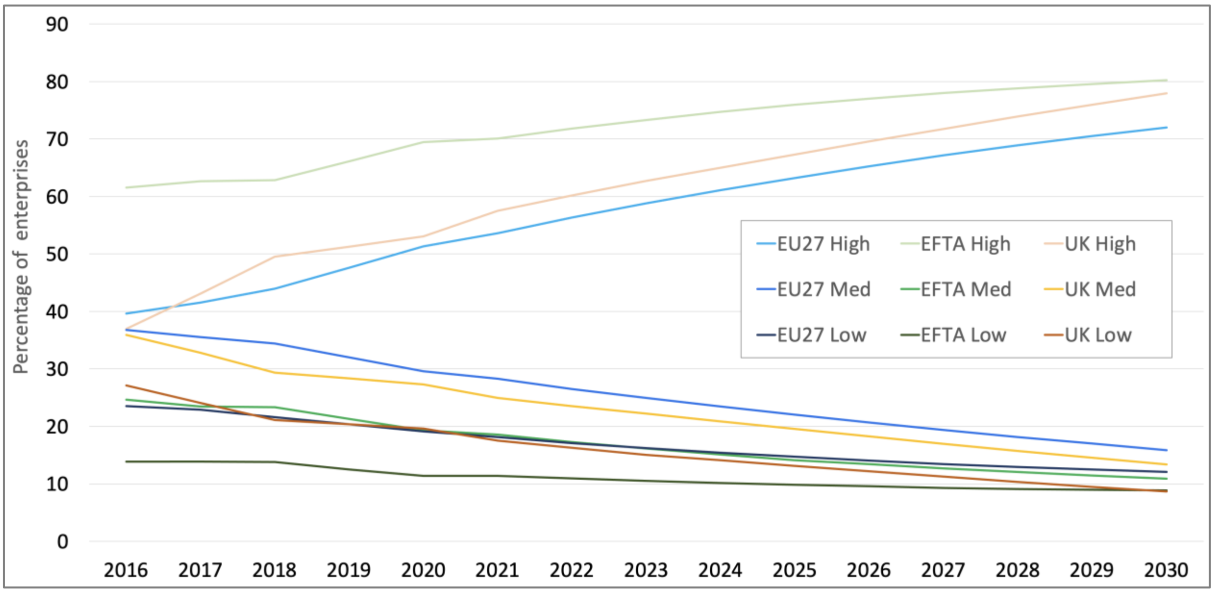
In EU27 Member States, in 2016 a similar number of enterprises were buying medium levels of cloud services (37 per cent) as were buying high levels of services (40 per cent). **By 2030 the situation is expected to change significantly - 72 per cent of enterprises are expected to be using high levels of cloud services, 16 per cent are forecast to use medium services and only 12 per cent will be purchasing low levels of cloud services**. The 2030 values are only slightly different to the 2025 forecasted value – 63 per cent, 22 per cent, and 15 per cent, respectively.

Similar changes can be observed in EFTA countries (see the green lines in figure 6). In 2016, 62 per cent were using high level cloud services. By 2025, this figure is expected to have grown to 76 per cent and further growth to 80 per cent in 2030. Over the same time period for the three EFTA countries, medium level cloud services are expected to decline from 25 per cent to 14 per cent, then to11 per cent. Uptake of low level cloud services have always been small – 14 per cent in 2016. They are expected to decrease to 10 per cent in 2025 and to 9 per cent in 2030.

The UK (see brown lines) follows a similar trajectory to EFTA countries. With an identical number of enterprises using low level services in 2030 (9 per cent) and almost identical per centage (78 per cent) with high levels of cloud use. Prior to this, in 2025, it is forecasted that 67 per cent of enterprises will be using high levels of cloud services, 20 per cent, medium level cloud services, and 13 per cent low level cloud services.

129 Low services include email and office software. Medium include file storage and database hosting (alongside those included as low). High includes accounting applications, CRM, computing power (alongside those included as low and medium).

Figure 6: Cloud service utilisation by enterprises using different levels of cloud services (low, medium and high) in EU27, EFTA and UK from 2016 to 2030



Source: Eurostat data (i soc_cicce_use) and study team extrapolation

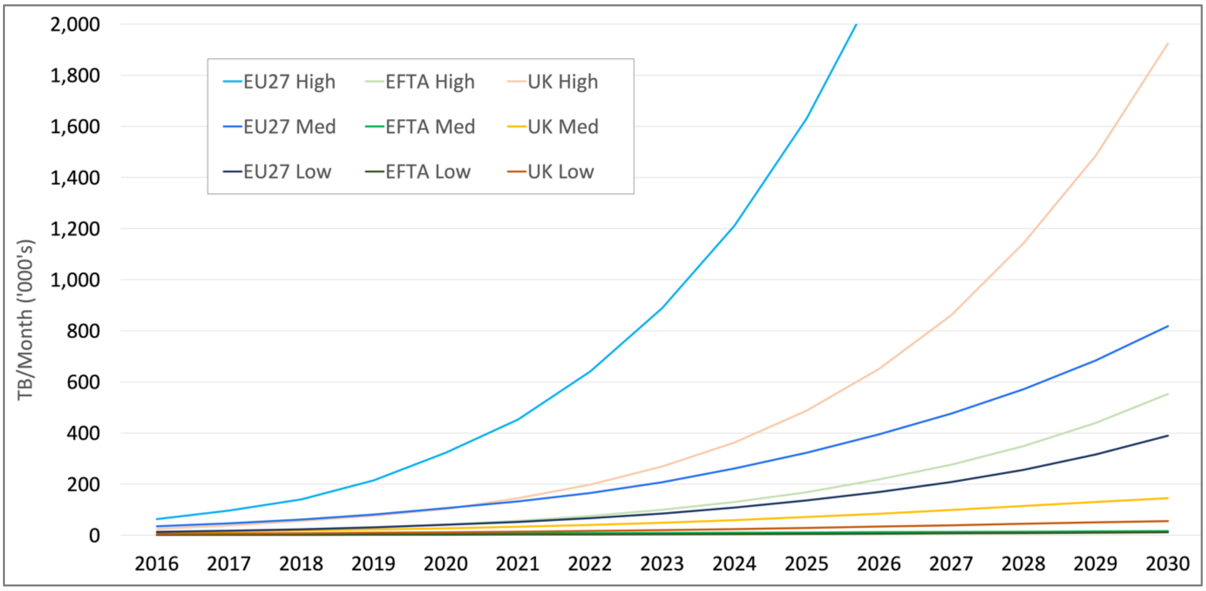
Table 1 showed how increasing number of enterprises buying cloud computing services used over the internet in EU27 Member States are forecast to lead to more than 15 fold increase in the flow of data to cloud and edge data centres between 2020 (504,760 TB/month) and 2030 (7,669,800 TB/month).

Enterprises are also moving to higher levels of services that will lead to increases in the flow of data to cloud and edge data centres.

Figure 6 shows that there is forecast to be a steady decrease in the proportion of enterprises buying low and medium level cloud services between 2016 and 2030. There is forecast to be an increase in proportion of enterprises using higher levels of cloud services that, due to their greater intensity of cloud usage, will exchange more data between enterprises and cloud and edge data centres.

Figure 7 forecasts the likely flow of data to cloud and edge data centres between 2020 and 2030 for the three different levels (low medium and high) of cloud computing services.

Figure 7: Enterprise data flowing to cloud and edge data centres by enterprises using different levels of cloud services (low, medium and high) in EU27, EFTA and UK



Source: Eurostat data (i soc_cicce_use) and study team extrapolation

Growth in the number of enterprises utilising high level cloud services (almost doubling in the EU27 Member States), combined with 18 fold growth in data utilisation per employee between 2016 (EU27 enterprise average 14.7 GB/month) and 2030 (264 GB/month), is forecast to lead to large increases in enterprise data flowing to cloud and edge data centres.

Enterprise data flowing to cloud and edge data centres using high level service is forecast to grow from 63,340 TB/month (0.72 EX/year) in EU27 Member States in 2016 to 6,140,000 TB/month (66.67 EX/year) in 2030[130]. In the three EFTA countries, data flowing to cloud and edge data centres using high level service is forecast to grow from 10,620 TB/month (0.12 EX/year) in EU27 Member States in 2016 to 553,000 TB/month (6.33 EX/year) in 2030.

In the UK data flowing to cloud and edge data centres using high level service is forecast to grow from 25,190 TB/month (0.31 EX/year) in EU27 Member States in 2016 to 1,924,000 TB/month (22.02 EX/year) in 2030.

As would be expected from declining adoption rates for medium and low level cloud services, enterprise data flowing to cloud and edge data centres has lower growth rates than high level service use, see Figure 7.

2.5 Geography of data flows

The final stage of analysis examined the geography of enterprise data flowing to cloud and edge data centres by enterprises buying cloud computing services used over the internet.

Analysis made the assumption that the amount of enterprise data flowing to cloud and edge data centres will be proportional to the market share of the main cloud service providers. In each country estimates were made of data flowing to cloud and edge data centres for

130 To enable differences in smaller values in Figure 7 to be identified the y-axis is restricted to 2 million TB/month.

the ten main cloud services providers examined in the study. This analysis estimated enterprise data generated by a country.

The final element of analysis was to estimate to where the data flowing to cloud and edge data centres of the ten main cloud services providers might flow. The locations of main cloud and edge data centres for the ten main cloud services providers identified by the study. For each of the 31 countries examined in the study the nearest main cloud and edge data centre for the ten main cloud services providers was identified. This enabled the end point destination (cloud or edge data centre) to be identified. The cloud or edge data centre destination of data generated by the 31 countries that was flowing to cloud and edge data centres could thus be modelled.

This analysis estimated enterprise data flowing to cloud and edge data centres in countries that possessed cloud infrastructure. The analysis provided interesting insights to the amount (TB/month) of enterprise data flowing to cloud service providers in different countries 2016 to 2021[131].

By far the largest amount of enterprise data flows are served by cloud and edge data centres in Germany. In 2020, Germany received 151,968 TB/month (1.74 EX/year) of cloud data flows from other countries. This represents 30.7 per cent of cloud data flows to EU27 Member States and 25.9 per cent of cloud data flows across the 31 countries studied.

The EU27 Member State receiving the second highest inflow of data served by its cloud and edge data centres in 2020 was the Netherlands (86,963 TB/month, 1 EX/year).

Both these countries (Germany and the Netherlands) received more enterprise data flowing to their cloud and edge data centres in 2020 than the UK (88,798 TB/month, 1.02 EX/year). In 2020, the UK received 12.7 per cent of total data flowing to cloud and edge facilities from the 31 countries studied. EFTA countries received 3 per cent of total data flowing to cloud and edge facilities (15,506 TB/month, 0.18 EX/year).

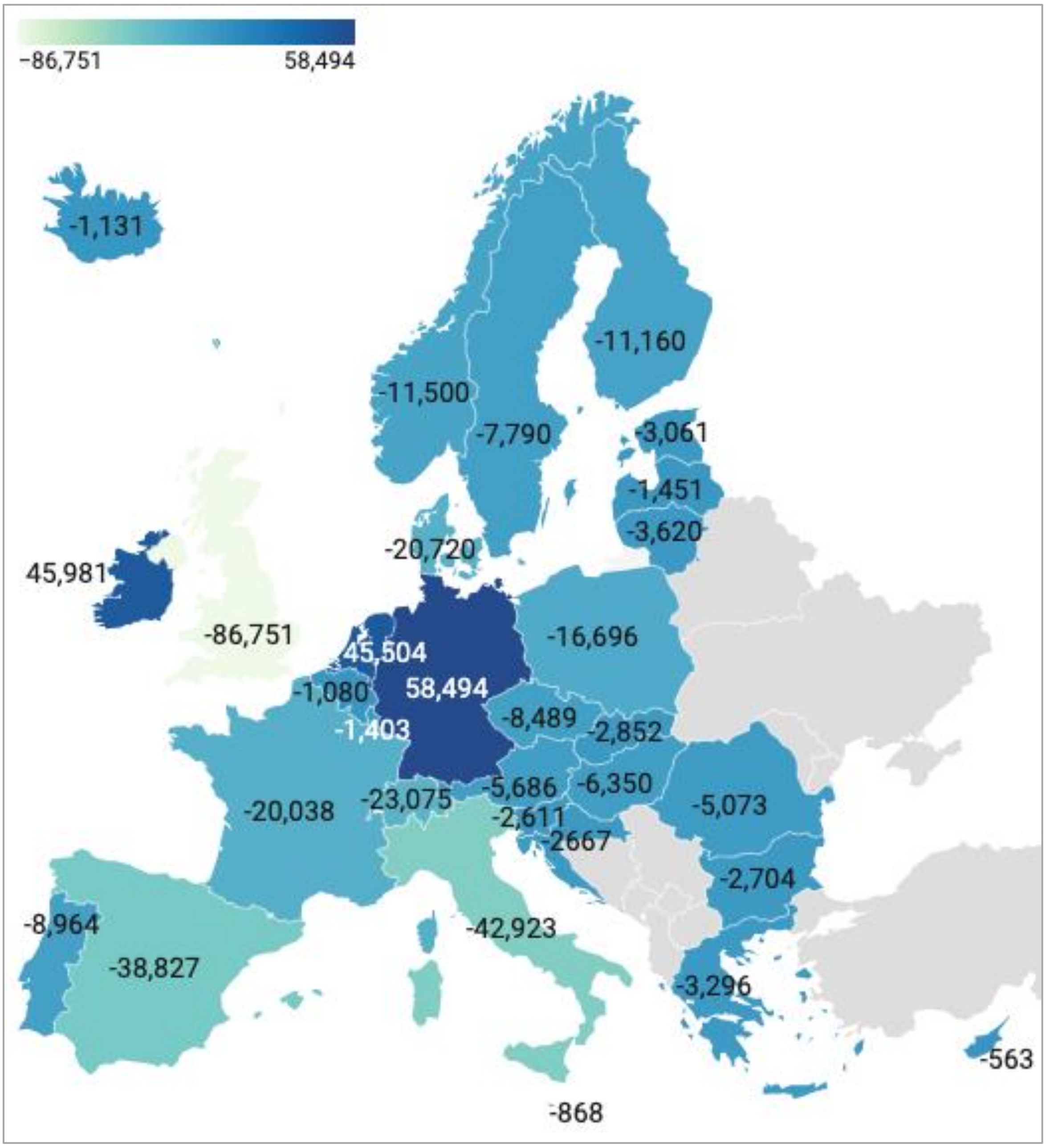
The final stage of analysis examined net enterprise data flowing to cloud and edge data centres for the 31 countries examined in the study. This was found by subtracting data inflows to a country from data outflows generated by the country. Net inflow/outflow results are presented in Figure 8.

Figure 8 reveals that three countries have net enterprise data inflows to their cloud and edge data centres. All other countries have net data outflows. Minus figures depicted in Figure 8 indicate there is more enterprise data flowing out of the country to cloud and edge data centres in other countries than data flowing into the country (if the country has cloud or edge data centres; 12 countries do not have this infrastructure[132]).

131 Analysis focuses on 2016 to 2021 because information about cloud infrastructure in these years is known. Many commentators predict considerable growth in cloud edge data centres in the next few years. Extrapolation to create forecasts in the future, when the distribution of infrastructure will change considerably, would create spurious results.

132 The 12 countries that do not have any cloud infrastructure are Bulgaria, Estonia, Greece, Croatia, Latvia, Lithuania, Luxembourg, Hungary, Portugal, Romania, Slovenia, and Slovakia.

Figure 8 Data flows geography: Net inflow/outflow map for data flowing to cloud and edge data centres for the 31 countries examined in the study (TB/month) in 2020.



The country with the highest net data inflow to its cloud and edge data centres in 2020 was Germany. German net enterprise data inflow to its cloud and edge data centres in 2020 was 58,494 TB/month (0.67 EX/year). Germany has a main cloud data centre for one of the top three cloud service providers.

Ireland has the second highest net enterprise data inflow to its cloud and edge data centres (45,981 TB/month, 0.53 EX/year). Ireland has a relatively small population (5 million in 2020) and only generates 10,595 TB/month (0.12 EX/year) of enterprise data flowing to cloud and edge data centres. However, it is unique in having main cloud data centres for the top three cloud service providers with the largest market share.

The Netherlands has the third highest net enterprise data inflow to its cloud and edge data centres in 2020 (45,504 TB/month (0.52 EX/year). The Netherlands has main cloud data centres for two of the top three cloud service providers.

The EU27 Member State with the largest net outflow of enterprise data to cloud and edge data centres in other countries is Italy. In 2020, Italy was thought to have one main cloud data centre and three edge centres. As a result, net enterprise data outflow from Italy is 42,923 TB/month (0.49 EX/year).

Of the 31 countries examined largest net enterprise data outflows were recorded by the UK. The UK has quite a lot of data infrastructure, including a main cloud data centre for one of the top three cloud service providers, but it does not have main cloud data centres for the second and third largest cloud service providers. As a result, the net enterprise data out flow in the UK in 2020 was 86,751 TB/month (1 EX/year).

As previously explained, Figure 8 shows the net flows of data flowing to cloud and edge data centres. Table 2 below, shows the total inflows of data flowing to cloud and edge data centres in the EU27 Members States, EFTA countries and the UK. 12 countries (as mentioned previously) do not have any cloud infrastructure, thus total inflows of data is equal to zero for these countries.

Table 2: Inflow of data flowing to cloud and edge data centres for the 31 countries examined in the study (TB/month) from 2016 to 2021.

	2016	2017	2018	2019	2020	2021
EU27	109,214	156,579	217,237	307,682	435,847	588,108
EFTA	4,059	5,789	7,860	11,082	15,506	20,883
UK	19,023	26,516	36,860	49,073	66,064	88,798
Total	132,296	188,885	261,957	367,837	517,417	697,789
Belgium	5,524	7,986	11,136	16,231	23,486	31,790
Bulgaria	0	0	0	0	0	0
Czechia	183	280	404	571	801	1,111
Denmark	623	888	1,214	1,674	2,302	3,126
Germany	35,943	51,997	72,720	105,202	151,968	180,747
Estonia	0	0	0	0	0	0
Ireland	16,234	22,629	31,153	41,779	56,576	75,730
Greece	0	0	0	0	0	0
Spain	1,167	1,671	2,207	3,072	4,269	5,707
France	14,248	20,553	28,535	39,586	55,001	75,152
Croatia	0	0	0	0	0	0
Italy	3,369	4,857	6,836	10,689	16,051	45,827
Cyprus	25	42	59	83	119	167
Latvia	0	0	0	0	0	0
Lithuania	0	0	0	0	0	0
Luxembourg	0	0	0	0	0	0
Hungary	0	0	0	0	0	0
Malta	17	25	38	54	77	110
Netherlands	21,654	31,242	43,643	61,609	86,963	117,891
Austria	1,014	1,547	2,170	3,229	4,701	6,495
Poland	362	550	756	1,176	1,774	2,386
Portugal	0	0	0	0	0	0
Romania	0	0	0	0	0	0
Slovenia	0	0	0	0	0	0
Slovakia	0	0	0	0	0	0
Finland	1,775	2,472	3,271	4,533	6,309	8,316
Sweden	7,077	9,840	13,096	18,193	25,448	33,552
Iceland	45	63	87	119	163	218
Norway	1,551	2,187	2,893	4,011	5,567	7,388
Switzerland	2,464	3,539	4,880	6,952	9,776	13,278
United Kingdom	19,023	26,516	36,860	49,073	66,064	88,798
Unknown	48,932	69,861	96,888	136,049	191,373	258,086

2.6 Data flows by data type

Data flows by type (personal and non-personal) were estimated using the way data was generated (machine-generated vs user-generated) and stored as a proxy. It is important to mention that mixed dataset can consist of both personal and non-personal data. Mixed datasets represent the majority of datasets used in the data economy and are common because of technological developments such as the Internet of Things (i.e.digitally connecting objects), artificial intelligence and technologies enabling big data analytics.

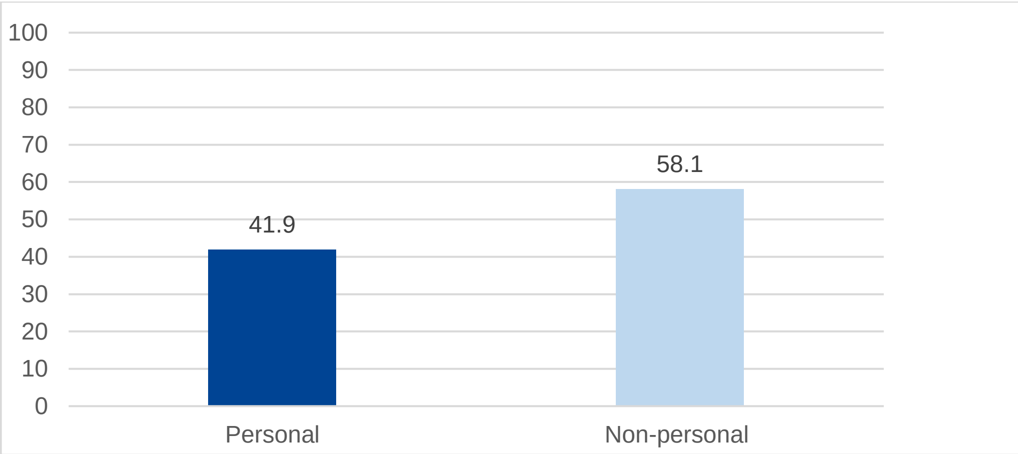
This was done using findings from the literature review and online survey conducted for the study (see Chapter 3 for more details).

The online survey was designed in order to complement the economic data flows findings presented above, which predominantly use secondary sources as opposed to primarily sources used for the survey. For example, in this study, the analysis found that 12,620

TB/day of enterprise data was flowing to cloud servers. The survey found that on average 59 per cent of total data stored in cloud infrastructure is non-personal data (see Figure 9). If survey respondents were representative of all EU27 Member State enterprises, this could suggest that 7,446 TB/day of enterprise data flowing to cloud servers is associated with non-personal data in 2020.

As there were issues regarding the representativeness of the survey given that, findings cannot be extrapolated in this report and only raw findings can be presented. The survey, conducted over a sample of 174 respondents, show **that 41 per cent of the total data stored in cloud infrastructure is personal data and 59 per cent is non-personal data**.

Figure 9: Share between personal and non-personal data stored in cloud infrastructure by enterprises



Moreover, the survey findings shed light on the estimation of data generated by users. According to survey respondents, of the 41 per cent of personal data (namely data relating to identified or identifiable person), 11 per cent is generated by a user/individual.

Concerning non-personal data, namely any data other than personal, out of 59 per cent, around 22 per cent is generated by a machine. Within machine-generated data, 19 per cent is industrial data. Industrial data was defined in the survey as data used as a direct input into the production process from a supply chain perspective. A full overview of survey respondent is presented in annex 9.

2.7 Cloud and edge data centre capacity and utilisation

The majority of analysis undertaken in this study focuses on the amount of data flowing to cloud and edge data centres from countries. Insights are also further disaggregated by NACE sector and company size (employees). Chapter 4 takes a closer look at the characteristics of cloud and edge data centres. It examines the location of main cloud and edge data centres and provides estimates of cloud data capacity and the amount of capacity utilised by enterprises buying cloud services used over the internet.

Cloud computing services are relatively new. The first commercial cloud services were introduced around 2006.[133] The industry has grown and changed considerably in the last 15 years. Significant early growth in huge hyperscale cloud centres is now being succeeded by considerable growth in edge computing, which reduces the travel time of data to cloud service users, reducing latency closer to zero. Edge computing also reduces the amount of cloud traffic that would previously have flowed across networks to main cloud centres.

Section 1.2 provided an overview of definitions of cloud computing and the different types of main cloud data centres and near and far edge data centres. The section highlighted that growth in small edge centres, the development of small micro data centres and the utilisation of telecommunications companies' data centres for edge cloud computing activities, makes defining edge centres and identifying locations for edge data centres difficult.

Analysis in this study estimates the volume of data flowing to cloud data centres from enterprises buying cloud computing services used over the internet. However, some of this data flows to companies providing cloud computing services that are limited in size and have a small market share (less than one per cent of the market for the 31 countries included in analysis). The market analysis suggests these smaller providers have a combined market share of 27 per cent.

The analysis presented in reveals that ten cloud service providers have a market share of approximately 73 per cent in the countries analysed. The location of main cloud and edge data centres for these ten providers has been investigated, see Figure 10. **Analysis therefore focuses on data flows to the 26 main cloud data centres and 73 edge data centres of the ten leading cloud service providers that have 73 per cent market share in the 31 countries examined in this study**[134].

The volume of data flowing to cloud and edge data centres from enterprises buying cloud computing services used over the internet for the smaller cloud service providers[135] that have 27 per cent of the market is calculated in analysis. While this analysis has been able to estimate the volume of data flowing to these cloud service providers from enterprises, the location of their main cloud and edge data centres is unknown. It has not therefore been possible to estimate to where this cloud data flows. The methodological annex provides further details of these calculations.

Figure 10 provides a list of cloud and edge centres for the ten main cloud service providers used when investigating flows to cloud data centres in this study. Section 4.1 describes how several sources for this information were cross-checked by contact with cloud service providers.

133 Gartner. 2020. 'Your data centre may not be dead but it's morphing' (https://www.gartner.com/en/documents/3990241/your-data-center-may-not-be-dead-but-it-s-morphing). Amazon introduced its 'simple storage service' in 2006 and Google released the Google App engine in 2008.

134 Section 4.1 provides an overview of the location of these centres. The ten cloud service providers examined have market shares of more than 1 per cent. Most telecommunications companies (with the exception of Deutsche Telekom) in Europe have a market share of less than 1 per cent across Europe and have therefore been excluded from analysis.

135 These are not part of the 73 per cent of the market for which the location of main and edge data centres is known.

Figure 10: Locations of the main cloud and edge data centres in EU27 Member States, EFTA and the UK

	Cloud Provider 1	Cloud Provider 2	Cloud Provider 3	Cloud Provider 4	Cloud Provider 5	Cloud Provider 6	Cloud Provider 7	Cloud Provider 8	Cloud Provider 9	Cloud Provider 10
AT		E	E		E	E			E	E
BE		E	M							
CZ		E								
CY		E	E		E	E				
DE	M	M + 4E	4E	M	3E	M		M	M	E
DK		E			E					
ES		E	E							
FI		E	M							
FR		M + 2E	2E	M	2E		M + 2E			M
IE		M + E	M + E		M + E					
IT		E	E	M	E					
MT		E	E							
NL		E	M + E	M	M + E	E				
PL		E					E			
SE		M + E	E		E					
CH		E	E		2E	M				
IS	E	E	E		E					
NO		E		M	2E					
UK	E	M + 4E	E	M	E	M			M	E

'M' indicates a main data centre. 'E' indicates and edge data centre. The number indicates the number of main and edge centres.

Due to commercial confidentiality, many cloud service providers were reluctant to provide information about the location and storage capacities of their cloud and edge data centres.

Since this research is an exploratory investigation of cloud and edge data centres in the EU, EFTA and the UK,

In the absence of insights about cloud and edge data storage capacity from cloud service providers this study utilised several sources of information[136] about global cloud data centre capacity. The methodology is presented in section 4. Analysis estimated cloud and edge data centre storage capacities in Europe. The analysis primarily utilised a Cisco study that provided global cloud storage estimates for 2016 and 2021[137].

Analysis showed that in EU27 Member States, cloud data centre storage capacity was 109 EX in 2016. Business cloud data centre storage capacity was estimated to be 76 EX.

Table 3: Estimates of cloud and edge data centre storage capacity in EU27 Member States, EFTA and the UK 2016 and 2021

	Cloud and edge data centre storage capacity 2016 (EX)	Cloud and edge data centre storage capacity 2021 (EX)	Business cloud and edge data centre storage capacity 2016 (EX)	Business cloud and edge data centre storage capacity 2021 (EX)
EU27	109	453	76	317

136 Cloud service provider web sites and TeleGeography (https://global-internet-map-2021.telegeography.com). CloudScene (https://cloudscene.com/browse/on-ramps). DataCentreKnowledge(https://www.datacenterknowledge.com/cloud/telegeography-maps-world-s-cloud-data-centers). In addition, direct contact was made with cloud service providers to seek information.

137 Cisco. 2018. 'Cisco Global Cloud Index: Forecast and methodology 2016 to 2021'. p20-21 (https://virtualization.network/Resources/Whitepapers/0b75cf2e-0c53-4891-918e-b542a5d364c5_white-paper-c11-738085.pdf).

EFTA	3.5	14.5	2.4	10.2
UK	16	68	11	47
Total	**129**	**535**	**90**	**374**

In 2016, it was estimated that 43 per cent of global cloud storage capacity was utilised. In **2021 it was estimated four-fold growth would have occurred and in EU27 Member States cloud data centre storage capacity was 453 EX in 2016**. **Business cloud data centre storage capacity was estimated to be 317 EX.** It was estimated that half of this cloud capacity was utilised in 2021. The methodology, further insights and caveats about forecasts are provided in section 4.

It was possible to extrapolate Cisco estimates and provide forecasts for 2025 and 2030, see Table 4.

Table 4: Estimates of cloud and edge data centre storage capacity in EU27 Member States, EFTA and the UK in 2025 and 2030

	Cloud and edge data centre storage capacity 2025 (EX)	Cloud and edge data centre storage capacity 2030 (EX)	Business cloud and edge data centre storage capacity 2025 (EX)	Business cloud and edge data centre storage capacity 2030 (EX)
EU27	1,390	5,356	973	3,749
EFTA	45	172	31	120
UK	207	798	145	559
Total	**1,642**	**6,326**	**1,149**	**4,428**

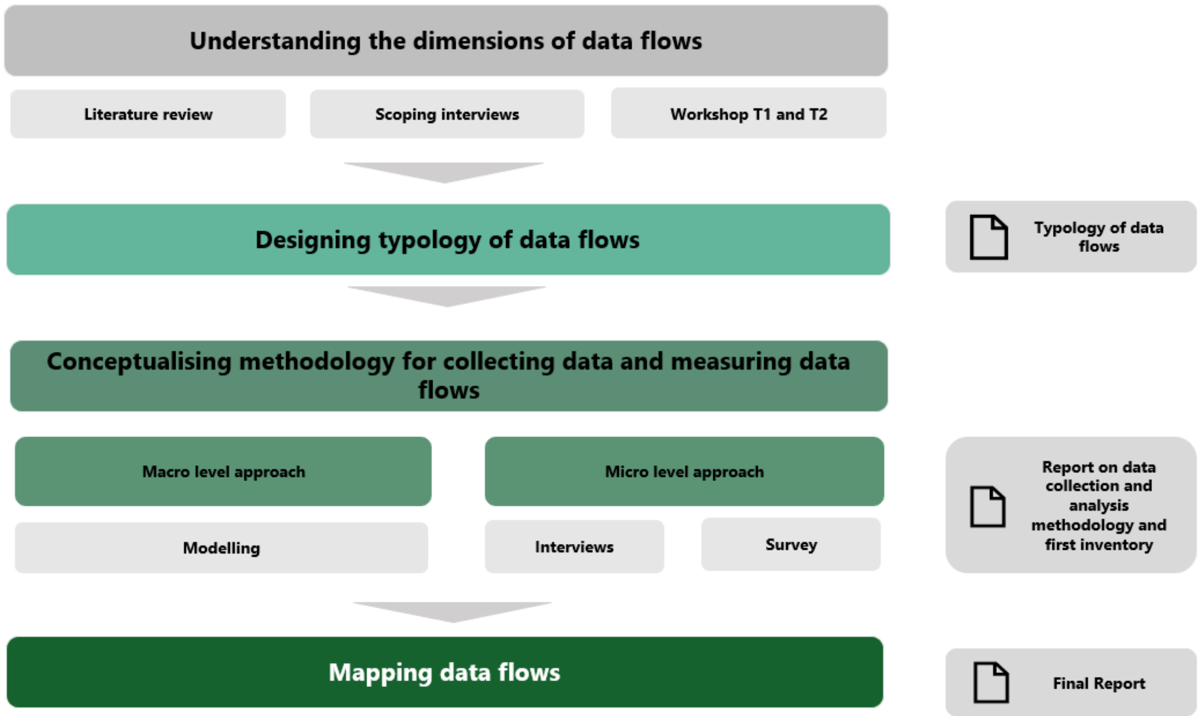
Forecasts suggest that in 2025 EU27 cloud data centre storage capacity is forecast[138] to be 1,390 EX and business cloud and edge data centre storage capacity would be 973 EX in 2025, this is 2.1 fold increase on business storage capacity in 2021. In 2030, EU27 cloud data centre storage capacity increases to 5,356 EX and business cloud data centre storage capacity is estimated to be 3,749 EX. This is 8.3 fold increase on business storage capacity in 2021. Table 3 extrapolate forecasts for EFTA and the UK.

138 Using linear extrapolation methods. These are described in the methodological annex.

3. A methodology for mapping data flows

This chapter presents the methodology designed for mapping data flows. The methodology, alongside the overview of data flows presented in the previous chapter, is a key deliverable of the study as it provides the modelling framework developed to estimate enterprise data flows to cloud and edge infrastructure in this study. The study has been designed to be replicable in the future for updating the mapping **of data flows when it comes to enterprise data flowing to public cloud and edge infrastructures.**

Figure 11: Overview of the methodology for estimating cloud based data flows



First, the study team conducted a literature review on the existing types of data flows and associated measurement methodologies, undertook scoping interviews, and organised an expert workshop to discuss the existing different types of data flows.

Different dimensions of data flows were explored to then design a typology of data flows (for more details, see Annex 1:Designing a typology of data and data flows). Seven dimensions were initially identified: *1) data types, 2) technology enabling data flows, 3) actors using cloud services, 4) functional areas related to activities within an entity, 5) services, 6) sectors, 7) geography*). As a third step, the study team designed a methodology to measure data flows and collect primary and secondary data on the seven data flow dimensions identified. **The data collection process allowed the team to reduce these seven to five measurable dimensions:** *country (namely narrowing down the geography dimension in origin and destination flows by country), sector (selecting only data intensive nace sectors), enterprise size (namely narrowing down the actors dimension into size of enterprise), services (namely focusing on cloud service utilisation and technology)* and *data* *type.* The overview of data flows disentangled by those five dimensions is presented in Chapter 2.

A macro-level analysis, (looking at the large-scale, aggregation of factors), was designed to estimate data flows with available data from secondary sources, based on the first four of those characteristics: by country, by sector, by enterprise size and cloud service utilisation. This has allowed the team to make reliable estimates of data flows in the first four dimensions of the typology, as it is explained in Section 3.1.

A micro-level analysis, namely the analysis of survey and interviews, aimed to complement the macro-level one by providing more detailed information based on primary data from the interviews and the online survey. The objective of the micro-level approach was to gather data on the additional dimension that could not be captured by the macro-level analysis, namely the **type of data**. This is explained in Section 3.2.

Finally, findings from both the macro-level and micro-level analyses were integrated to provide a comprehensive mapping of data flows covering the five dimensions of the typology, as Section 3.3 details.

Table 5: The study dimensions and insights provided by micro-level and macro-level analysis

Data flow dimension	Data collection source	Approach chosen	Objective
Geography	Secondary data provides information about the size of enterprises buying low, medium and high cloud services used over the internet. Enterprise size bands 0 to 9, 10 to 49, 50 to 249 and 250 or more employees	Macro-level	To map enterprise data flowing to cloud data centres from and to EU countries, EFTA and UK
Sector	Secondary data provides information about enterprises buying cloud services used over the internet in NACE sectors C to S	Macro-level	To map enterprise data flowing to cloud data centres by NACE sectors (C to S)
Enterprise size	Secondary data provides information about enterprises buying cloud services used over the internet in 31 countries	Macro-level	To map enterprise data flowing to cloud data centres by enterprise size
Nature of cloud services utilised	Secondary data provides information on the nature of cloud service utilised (low-medium-high) by enterprises Primary data provides information on the percentage of cloud applications used	Macro-level and micro-level	To map enterprise data flowing to cloud data centres by low/medium/high To provide an estimation on the main services used
Data type	Primary data (survey) provides information on the percentage of data stored per type (personal or non-personal)	Micro-level	To provide estimation on data stored by data type (personal/non-personal)

3.1 A methodology to examine enterprise cloud data flows

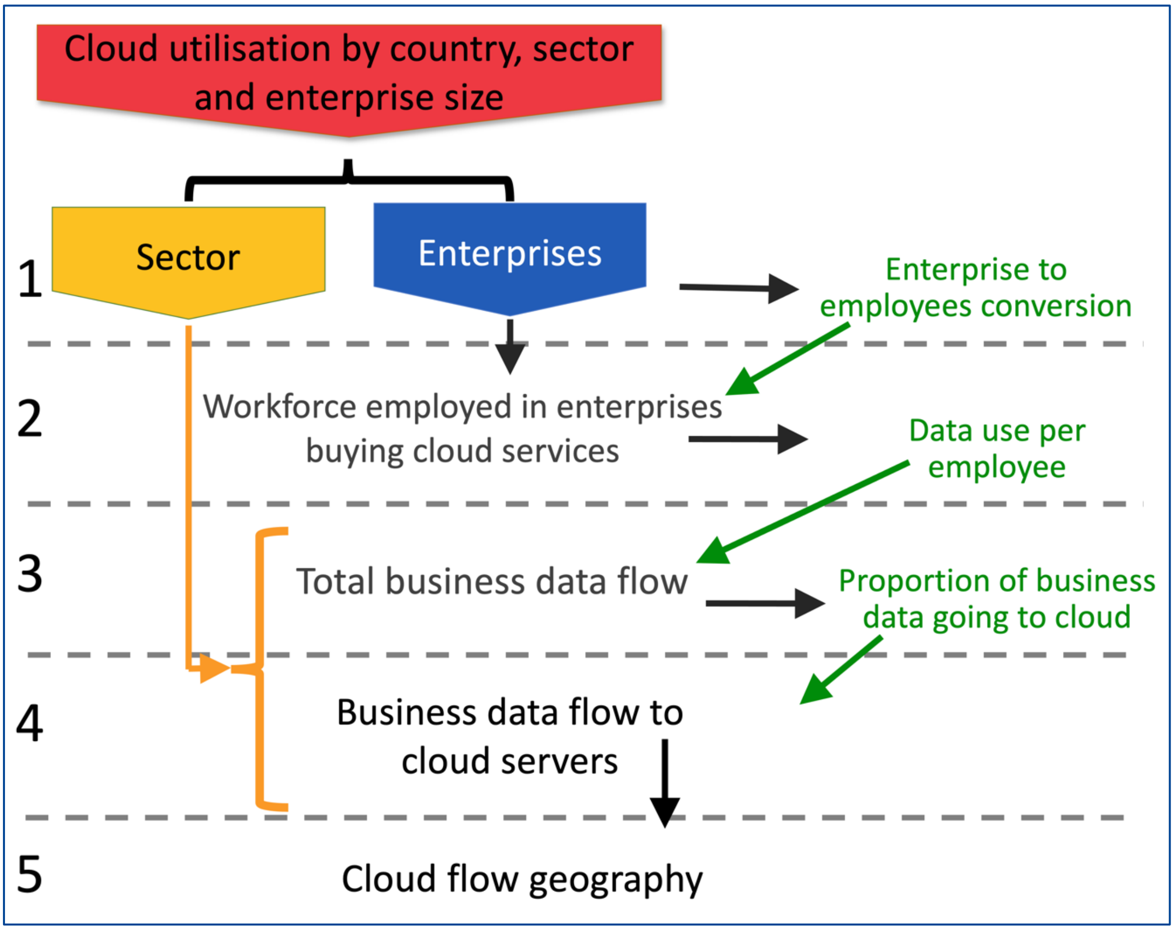
This section provides an overview of the macro-level analysis. The focus will be on both the methodological approach and the results achieved from the investigation of data flows associated with enterprises buying cloud computing services used over the internet.

3.1.1 Introduction to the five-stage macro-level approach

For clarity, the precise areas of focus for macro-level analysis are presented in section 2. This section provides a brief overview of the methodological approach for each stage of macro analysis. More fulsome methodological details are described in the separate sections that follow for each of the five stages of the macro-level approach.

The methodological annex provides focuses solely on the methodology and provides additional details of the rationale for the approach and details of methods adopted.

Figure 12: Macro level methodology to estimate enterprise data flows to cloud and edge data centres



At the core of the methodology is the analysis of robust Eurostat statistics about cloud service use by enterprises.[139] Figure 12 represents this first stage in the red box at the top left corner of the graphic. Insights to these statistics for the research are presented in the next section. These statistics provide a robust foundation due to the rigorous methods by which the data is collected by national statistical agencies.

Figure 12 depicts the five stages of analysis. Each stage of analysis is described below. To assist with understanding each stage, Eurostat statistics for Belgium are used to demonstrate the methodological approach in this introduction to the methodology. By focusing on the application of the methodology for one country the approach should be easier to understand than by observing the analysis for 31 countries

- In 2018 Belgium had 635,500 enterprises, employing 4.565 million people. The average size of an enterprise is 7.2 employees.[140]
- **40 per cent of enterprises (254,200) in Belgium bought cloud computing services that are used over the internet**.

The **first stage of the analysis** (row 1) explains why the analysis of cloud use by enterprise could lead to spurious results if data is not standardised. First stage analysis is described in section 3.1.2.

The preceding statistics about Belgian enterprises help to define the problem. There are 254,000 enterprises that use cloud services. If all cloud-using enterprises only had one employee there would be 254,000 employees. If each cloud-using enterprise had 15 employees, the total number of employees in these enterprises would be 3.81 million (15 x 254,200). All other things being equal, 3.81 million people in cloud-using enterprises will create more data flowing to cloud and edge data centres than 254,000 people[141].

To address this deficiency of analysis, data about enterprises buying cloud computing services used over the internet is standardised; this forms the **second stage** (see row 2). Second stage analysis is described in section 3.1.3.

Standardisation provides more meaningful and accurate insights on the number of employees, rather than enterprises, employed in enterprises of different (workforce) sizes and in different sectors using cloud services. This will then serve to estimate more accurately the volume of data flowing from enterprise to cloud and edge infrastructure across the EU27, UK and EFTA.

Eurostat provide enterprise data by size bands for employees[142]. For example, in Belgium in 2018:

- The 602,300 enterprises with ‘0 to 9’ persons employed 993,100 employees (average size 1.65 persons); approximately 14 per cent of enterprises in this band size use cloud services[143].
- 26,600 enterprises with ‘10 to 49’ persons employed 515,100 employees (average size 19.4 persons); 36 per cent of enterprises in this band size use cloud services.

139 Eurostat indicator isoc_cicce_use. Data was extracted on 2 March 2021. The data provides sectoral and enterprise size (employees) insights to enterprises buying cloud computing services used over the internet. Eurostat data is used for the time period 2016 to 2020.

140 The average size of all enterprises in Belgium (7.2 employees) is provided only for illustrative purposes to enable comparison in the bullet points below with the average size of enterprises in different size bands. The figure of 7.2 employees is not used in analysis.

141 This is also dependent on the intensity of the use of data.

142 Eurostat has data for EU 27MS, EFTA (excluding Liechtenstein) and UK.

143 Eurostat data for this indicator (0-9 employees) is sparse, thus this per cent has been calculated through thorough estimations.

- 4,100 enterprises with ‘50 to 249’ persons employed 415,900 employees (average size 101.4 persons); 55 per cent of enterprises in this band size use cloud services.
- 986 enterprises with ‘more than 250’ persons employed 1,036,900 employees (average size 1,051.6 persons); 79 per cent of enterprises in this band size use cloud services.

Analysis enables the standardisation process to be completed by calculating the number of employees who are employed in enterprises of different sizes that buy cloud computing services used over the internet. For example, in Belgium analysis estimates that in 2018 there were 2.48 million employees working in enterprises buying cloud computing services used over the internet.

The **third stage macro analysis** combines Stage 2 insights (about the numbers of employees in enterprises that buys cloud services over the internet) with ITU statistics on fixed broadband traffic (exabytes (EX) per annum) to estimate the total enterprise internet traffic.

The investigation of internet traffic by enterprises includes both those enterprises utilising cloud services that they have bought and are using over the internet and those undertaking other activities online. This provides insights to total traffic flow per employee (cloud and non-cloud traffic) which can then be used when estimating the amount of enterprise data flowing to cloud and edge data centres in Stage 4.

To continue our example, in Belgium (2018), the average internet traffic per employee was 0.98 GB/day (29.8 GB/month, and 357.6 GB/year). Across the Belgian workforce of 4.795 million people this equates to 4,584 TB/day (139,581 TB/month, and 1,675,201 TB/year).[144]

The **fourth stage of analysis** (row 4) uses results from analysis in stage two and stage three to estimate data flows to cloud and edge data centres.

Analysis in stage four focuses on the proportion of total business traffic flows for enterprises that is exchanged with cloud and edge data centres (among enterprises utilising cloud). In Belgium, this is estimated that the flow of data to cloud and edge data centres by enterprises buying cloud computing services used over the internet is 425 TB/day (12,949 TB/month, and 155,409 TB/year) in 2018.

Macro-level analysis undertaken in Stage 4 estimates that the volume of data flowing from enterprises[145] to cloud and edge data centres in EU27 Member States was 8,061 TB/day in 2018 (245,471 TB/month, and 2.8 EX/year). The aggregated figure for EU27, EFTA countries and the UK is 11,785 TB/day (358,845 TB/month, and 4.1 EX/year) in 2018. Enterprise data flowing to cloud and edge data centres is forecast to increase over thirty-fold: to 251,883 TB/day (7,669,835 TB/month, and 87.8 EX/year) in EU27 Member States in 2030 and 347,376 TB/day (10,577,600 TB/month, and 121.1 EX/year) for EU27, EFTA countries and the UK. At the intermediate point in 2024, enterprise data flowing to cloud data and edge centres is forecast to be 54,857 TB/day (1,670,394 TB/month, and 19.1 EX/year) in EU27 Member States and 75,527 TB/day (2,299,805 TB/month, and 26.3 EX/year) for EU27, EFTA countries and the UK.

144 The description of the methodology below describes how an estimate of total internet traffic in Belgium can be used with information about internet traffic flowing to households (70 per cent of all internet traffic) and enterprises (30 per cent) to calculate average traffic flow per employee.

145 The modelling in this paper is largely based on Eurostat ‘enterprise’ statistics. Eurostat describe an ‘enterprise’ as ‘an organisational unit producing goods or services which has a certain degree of autonomy in decision-making. An enterprise can carry out more than one economic activity and it can be situated at more than one location’ (https://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Enterprise). The Eurostat definition of an ‘enterprise’ does not differentiate between the intermediate or ultimate country ownership of an enterprise.

The **final stage of analysis** (row 5) utilises information about the locations of the 26 main cloud data centres and 73 edge data servers to examine **the origin and destination of** the 75,527 TB/day of data flowing to cloud and edge data centres will flow across the EU27 Member States, the UK and EFTA. For eight of the main cloud service providers, the nearest cloud and edge data centres to each of the 31 countries investigated were identified. The cloud and edge data flows from each country were allocated to these. This enabled insight about the destination and origin of where the enterprise cloud and edge data flows that were identified in Stage 4 were flowing to.

For example, analysis estimates that in 2018 the flow of data to cloud and edge data centres by Belgian enterprises buying cloud computing services used over the internet is 12,950 TB/month (data flows generated by Belgian enterprises, a proportion of which will flow out of Belguim into other countries with cloud and edge data centres). Analysis found that the flow of data to cloud and edge data centres in Belgium from enterprises in Belguim and other countries examined in the study buying cloud computing services used over the internet was 11,130 TB/month (inflows of data).

More detailed descriptions of each stage of macro analysis are provided in the remaining sections of this chapter.

3.1.2 Stage One: Cloud use by enterprise

The first stage of the study examined Eurostat data from 2016 to 2020 about enterprises buying cloud computing services used over the internet. The goal of the study was to map the geographical flows of enterprise data flowing to cloud and edge data centres by enterprises buying cloud computing services used over the internet. It was therefore important at the start of the study to understand how many enterprises were using cloud computing services in the 31 countries investigated in the study. Enterprises not using cloud computing services could then be omitted from analysis.

The characteristics of enterprises buying cloud computing services was also important in enabling the study to have a more granular approach. Eurostat data enabled the analysis of enterprises buying cloud computing services by a number of enterprise characteristics. These included:

- Country (EU27, EFTA and UK);
- Sector (NACE C to S);
- Enterprise size (‘0 to 9’, ‘10 to 49’, ‘50 to 249’ and ‘over 250’).

In addition, Eurostat provided information about the nature and intensity of cloud services used enterprises buying cloud computing services.

Using the historic Eurostat data from 2016 to 2020 analysis used extrapolation methods to provide forecasts about the characteristics of enterprises buying cloud computing services used over the internet until 2030.

The starting point for the analysis of enterprise data flowing to cloud and edge data centres is the analysis of Eurostat statistics.[146] In particular, the statistics about enterprises buying cloud computing services used over the Internet. These statistics provide robust information, collected by national statistics offices, about the use of cloud services[147] between 2014 and 2018 by:

- Country (EU27, EFTA and UK);
- Sector (public and private sectors – NACE C to N);
- Enterprise size (in bands of '1 to 9', '10 to 19', '20 to 49', '50 to 249' and '250 or more' employees);
- Nature of cloud services utilised (low, medium and high level).[148]

The data is robust and provides immediate insights into cloud use by enterprises for all of the preceding variables. For example:

- The smallest levels of cloud utilisation across the EU27 Member States in 2020 occurred in Bulgaria (11 per cent of enterprises);[149]
- The highest level of cloud utilisation occurred in Finland and Iceland (75 per cent);
- The ICT sector[150] has the highest levels of cloud utilisation (70 per cent);
- Cloud use in EU27 Member States is highest (65 per cent) among businesses with 'more than 250' employees and lowest (20 per cent) for enterprises with '9 or less' employees;[151]
- EU27 growth in cloud service use by enterprises was 17.3 per cent CAGR between 2016 to 2020;
- In 2020 in EU27 Member States 19.1 per cent of enterprises buying cloud services used over the internet were buying low level services. 29.6 were buying medium level services and 51.3 per cent were buying high level[152]. Since 2016 the proportion of enterprises buying low and medium levels has been declining and the percentage buying higher level services has increased.

Analysis of the Eurostat data provides a useful overview of cloud service use by enterprise and also a basis for extrapolating past trends since 2020 to provide insights to the future. Figure 13 shows the forecasts in cloud services adoption by enterprises using Eurostat data[153] for three geographical locations: EU27, EFTA and the UK. Analysis of robust Eurostat data concerning cloud adoption by enterprises and historic data (at present from 2016 to 2020) enables trends to be identified.

---

146 Eurostat indicator isoc_cicce_use. Data was extracted on 3rd March2021. On 21 January 2021 Eurostat updated the indicator; providing statistics for 2020, but not for 2019. Our team extrapolated estimates for 2019.

147 The Eurostat definition is 'Buy cloud computing services used over the internet'.

148 Low services include email and office software. Medium include file storage and database hosting (alongside those included as low). High includes accounting applications, CRM, computing power (alongside those included as low and medium). It will be reasonable to assume that higher level, more intense users, will exchange greater volumes of data.

149 Data in this bullet point list relate to Eurostat indicator isoc_cicce_use for enterprises. Analysis for this paper standardises information within countries using low, medium and high levels of cloud services. The difference between comparing performance for enterprises and the workforce employed in enterprises is explained later in section 4.4.

150 NACE sector J.

151 There is a paucity of Eurostat data for enterprises with fewer than ten employees. Data was only available for five countries.

152 Low services include email and office software. Medium include file storage and database hosting (alongside those included as low). High includes accounting applications, CRM, computing power (alongside those included as low and medium). It will be reasonable to assume that higher level, more intense users, will exchange greater volumes of data.

153 Eurostat. 2020. All enterprises, without financial sector (10 persons employed or more) buying cloud computing services used over the internet. Indicator isoc_cicce_use last updated 24 February 2020.

The latest Eurostat data (2021) shows that in 2020, 36 per cent of enterprises in EU27 Member States were 'buying cloud services used over the internet'.[154] **Between 2016 and 2020, the average rate of growth (CAGR) was 17.3 per cen**t.[155] The 17.3 per cent CAGR figure is used, on a yearly basis from 2021, to extrapolate a forecast of 56 per cent of enterprises using cloud services in 2025 and 78 per cent in 2030.

Similar methods were used to forecast growth in EFTA countries, which already have higher levels of cloud utilisation than EU27 Member States - 64 per cent in 2020. CAGR for EFTA countries between 2016 and 2020 was 12.5 per cent. Using this CAGR figure, it has been estimated that by 2025, 93 per cent of enterprises will be using cloud services and by 2027 (thus 2030) all enterprises will be using cloud services.

The final country, the UK, is in a similar situation to EFTA countries. In 2020, 53 per cent of enterprises were using cloud services. Growth between 2016 and 2020 in the UK was 10.95 per cent, which is less than EU27 and EFTA countries. Using the CAGR figure, it has been estimated that by 2025 75 per cent of enterprises will be using cloud services and by 2030 98 per cent of enterprises will be using cloud services.

Eurostat trends provide a reasonable basis for forecasts. However, care needs to be taken in extrapolating results. Technology adoption often follows an s-shaped curve[156] with low (percentage) levels of adoption from early users. As the technology becomes more popular, percentages then increase. These larger percentage increases is probably the current stage of adoption for cloud in Europe[157]. In the final part of the s-shaped curve, later adopters are generally smaller in numbers and more reticent. Saturation for the adoption of technologies is rarely reached since some people will never adopt the technology. For example, 2 per cent of EU27 households do not have a television.[158]

Numerous forecasts for growth in sales (€/$) from the provision of cloud services and the number of cloud data centres are available. These reports are useful for learning about revenue growth for those offering cloud services,[159] but they do not provide suitable insights to *growth* in the *number* of enterprises using cloud services.

Our team has only been able to find one report that offers forecasts of growth in cloud use by enterprises. This provides a useful cross reference and a relatively good validation of Eurostat CAGR figures for cloud growth described above (EU27 CAGR 17.3 per cent, EFTA 12.45 per cent, UK 10.95 per cent CAGR). This forecast, made by Cisco, predicts 11 per cent CAGR 2016 to 2021 in the number of enterprises purchasing cloud services. This rate is lower than past EU27 Member State and EFTA trends (2016 to 2020), but it is similar to UK growth rates, as forecasted in this study.[160]

154 This is the top level categorisation of cloud use utilised by Eurostat. More granular breakdowns by the characteristics in the first set of four bullet points provided above.

155 When Eurostat data (or other statistics) are updated in the top down model, CAGR can be recalculated by recalibrating the CAGR equation to adjust the value of an investment at the end of the period and the value at the beginning of that period and changing the number of years between the two dates in the exponent of one divided by the number of years.

156 This differs from a linear curve which increases (or decreases) by a constant.

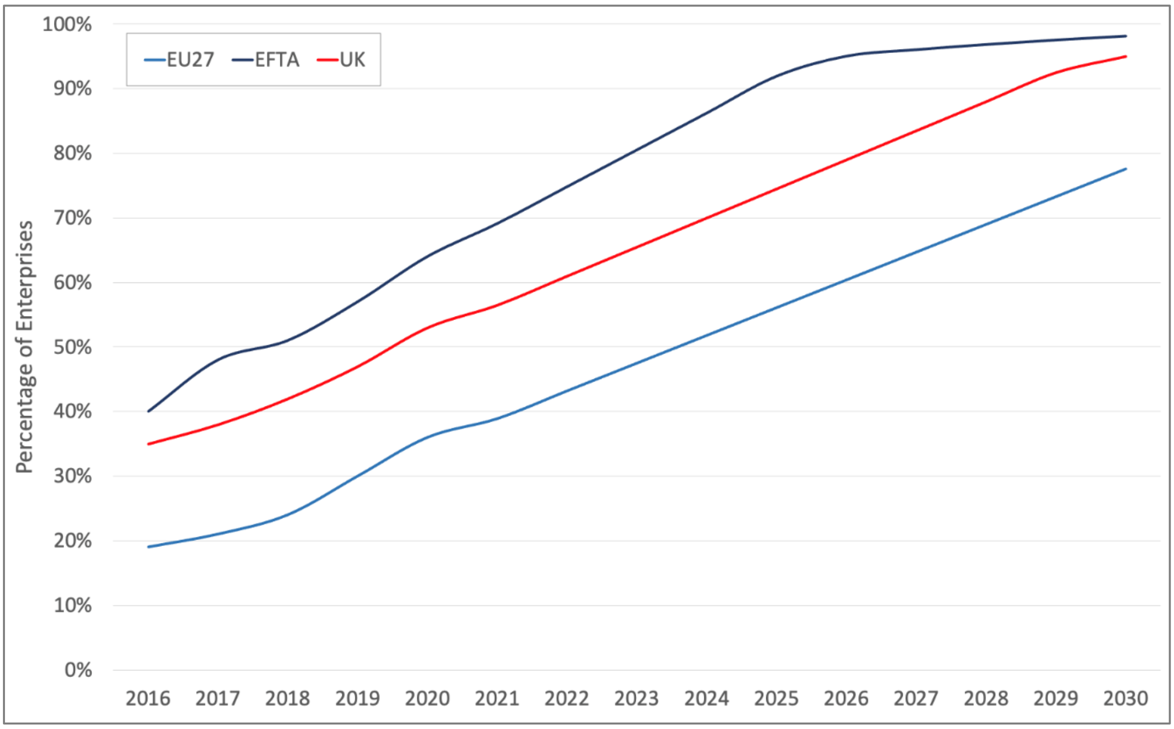
157 EU27, EFTA and UK.

158 European Commission Eurobarometer 335 eCommunications household survey (https://ec.europa.eu/commfrontoffice/publicopinion/archives/ebs/ebs_335_sum_en.pdf)

159 For example IDC. 2019. IDC's IaaSView 2019. https://www.idc.com/getdoc.jsp?containerId =prUS45625619 The study focuses on those already using cloud services.

160 Cisco global cloud index. 2018 (https://virtualization.network/Resources/Whitepapers/0b75cf2e-0c53-4891-918e-b542a5d364c5_white-paper-c11-738085.pdf).

Figure 13: Cloud service utilisation by enterprises 2016 to 2030



Source: Eurostat data (i soc_cicce_use) forecasts use CAGR rates 2016 to 2020, made by the study team, documented in preceding text

In 2020, Covid-19 had an unprecedented and devastating impact on lives and the way society and economies function. Many economic and other forecasters had to make decisions about how to accommodate the massive changes caused by Covid-19 in longer-term forecasts. McKinsey predicts a return to global pre-crisis GDP levels in 2022.[161] In terms of internet traffic, Tech4i2 and others have shown that internet traffic increased by 40 per cent in some locations following lockdowns.[162] But within five weeks of lockdowns being eased, internet traffic volumes returned to pre-lockdown levels.

A small number of articles have been published about the impact of Covid-19 on IT and cloud spending. A Gartner report suggests a decline in IT spending by 7.2 per cent as enterprises evaluate IT spending.[163]

Other studies suggest a small, short-term impact, generally by bringing forward cloud use and spending. For example, from a small global survey of 750 cloud decision makers, Flexera state 31 per cent of respondents suggested cloud use had increased slightly.[164] A survey by MariaDB[165] Corporation discovered that 39 per cent of UK respondents are bringing forward their move to the cloud.

161 McKinsey. 2020. 'Covid19: Global health and crisis response' (https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/Risk/Our%20Insights/COVID%2019%20Implications%20for%20business/COVID%2019%20July%2023/COVID-19-Facts-and-Insights-July-23-vF.pdf)

162 Tech4i2. 2020. 'Covid19 impact on internet use' (https://www.tech4i2.com).

163 Gartner. 2020. 'Forecast: Enterprise IT Spending by Vertical Industry Market, Worldwide, 2018-2024, 1Q20 Update' (https://www.gartner.com/en/documents/3984513/forecast-enterprise-it-spending-by-vertical-industry-mar)

164 Shah, S 2020. 'Why cloud is more vital than ever'. Raconteur (https://www.raconteur.net/technology/cloud-business-2020/benefits-cloud-computing-covid). Markets and Markets. 2020. 'Covid19 impact on cloud computing' (https://www.prnewswire.com/news-releases/covid-19-impact-on-cloud-computing-market--exclusive-report-by-marketsandmarkets-301047436.html) Flexera. 2020. 'Flexera State of Cloud Report 2020' (https://info.flexera.com/SLO-CM-REPORT-State-of-the-Cloud-2020)

165 https://datacentrereview.com/content-library/opinion/1738-covid-19-s-impact-on-cloud-adoption

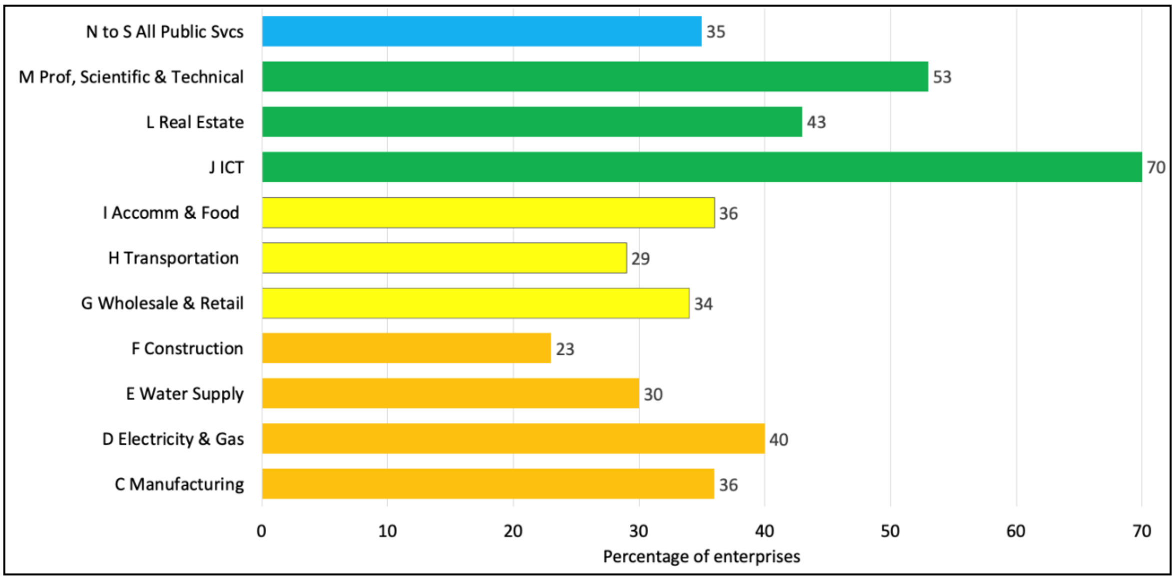
If these increases are indeed taking place, they are not reflected in figures examining cloud spending. Synergy Research Group note that Q2 spent on cloud infrastructure services passed the $30 billion milestone,[166] with a change in the percentage growth rate between Q2 2019 and Q2 2020 of 33 per cent. The change from Q2 2018 to Q2 2019 was 41 per cent. Synergy reports that the change from Q2 2017 to Q2 2018 was 52 per cent.

The cloud market continues to grow, but revenue figures do not suggest an *abnormal* increase in cloud service purchase due to Covid-19.

It is possible that Covid-19 might give a short-term boost to cloud service use, with some enterprises bringing forward expenditure which might otherwise have occurred later. But overall, for modelling purposes in this study, as the preceding studies indicate, it is suggested that the status quo is maintained[167].

The figure below provides an overview of the proportion of enterprises utilising cloud services in business and public sectors (by NACE code). **The ICT sector (70 per cent) for EU27 Member States, EFTA countries and the UK has far higher levels of cloud utilisation than the averages for the EU** (36 per cent in 2020), EFTA (64 per cent) and UK (53 per cent).

Figure 14: Cloud service adoption by enterprises in different sectors in EU27, EFTA and UK in 2020



Source: Eurostat data (i soc_cicce_use)

Manufacturing has a low level of cloud service use (36 per cent), but it comprises a much larger proportion of enterprises than ICT (16 per cent more manufacturing enterprises than ICT enterprises). Annex 3 provides estimates for cloud service adoption by enterprises in NACE sectors C to S from 2016, with forecasts to 2030.

Figure 14 shows historic data from Eurostat between 2016 and 2020, and extrapolated forecasts to 2030 for cloud service adoption by enterprises of different sizes (employees) in EU27 Member States, EFTA and UK.

166 Synergy Research Group. 2020. 'Quarterly Cloud Spending Blows Past $30B; Incremental Growth Continues to Rise' (https://www.srgresearch.com/articles/quarterly-cloud-spending-blows-past-30b-incremental-growth-continues-rise)

167 Excluding any short-term boost that may have arisen due to COVID-19 from this study.

In Figure 15, data relating to EU27 Member States is presented in blue. The light-coloured line is for enterprises with ‘20 to 49’ employees; in 2020, 34 per cent of these EU27 enterprises were buying cloud services used over the internet. The blue shading darkens as enterprise sizes become larger. For example, in 2020, 46 per cent of enterprises with ’50 to 249’ employees were buying cloud services. Finally, the dark blue line shows that in 2020, 66 per cent of the largest enterprises with over 250 employees were buying cloud services used over the internet.

In 2030 it is expected that 65 per cent of EU27 enterprises with ‘10 to 49’ employees will buy cloud services used over the internet. This is almost twice the number in 2020.

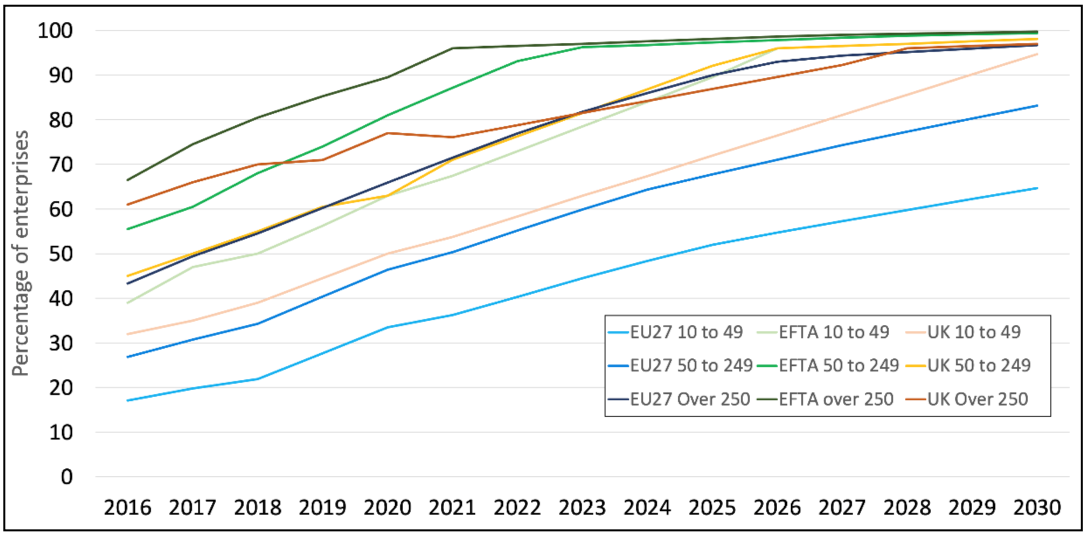
Cloud service use for medium sized enterprises with ‘50 to 249’ employees is expected to reach 83 per cent in 2030, 1.8 times greater than the 2020 figure.

Among the largest EU27 enterprises, those with ‘250 or more’ employees, cloud use is expected to almost reach saturation point at 97 per cent in 2030.

Initially higher levels of cloud services use can be seen in EFTA countries. All lines for these countries are shaded green. In enterprises with ‘10 to 49’ employees, in all the three EFTA countries, cloud adoption was 63 per cent in 2020. In 2025, cloud adoption is above 90 per cent for all enterprise sizes. Accordingly, for all sizes of EFTA enterprises, cloud adoption is expected to exceed 98 per cent in 2030.

Similar growth levels can be seen for the UK; UK lines are shaded brown. In UK enterprises with 10 to 49 employees, cloud adoption was 50 per cent in 2020, 72 per cent in 2025 and 95 per cent was predicted for 2030.

Figure 15: Cloud service adoption by enterprises of different sizes in EU27, EFTA and UK



Source: Eurostat data (i soc_cicce_use) and study team extrapolation[168]

Annex 3 provides estimates for cloud service adoption by enterprises of different (employee) sizes from 2016, with forecasts to 2030.

168 Linear extrapolation was undertaken to forecast values of different sizes of enterprise to 2030. Some of these extrapolations led to the creation of values greater than 100 per cent. Clearly this is unrealistic as percentage values can never be more than 100. To adjust these errors, data from s-shaped adoption curves was utilised. The data represents the flatter parts of the s-shaped curve at the start and end of technology adoption. The s-shaped method ensured that values were never greater than 100 per cent.

This section has described stage one macro-level analysis to investigate cloud data service use by enterprises, sector, country and service type. Eurostat data is used to present information about the characteristics of enterprises in sectors - both private and public - that are buying cloud services used over the internet. Eurostat cloud service use statistics are disaggregated by country, sector and enterprise size (workforce).

The annual increase (CAGR 17.3 per cent) in enterprises buying cloud services is used to forecast cloud use in 2024 and 2030. **In EU27 Member States, the proportion of enterprises using cloud is forecast to increase from 24 per cent in 2018, to 78 per cent in 2030**.

Conclusion

This section has introduced the first stage of analysis to provide insights to enterprise data flowing to cloud and edge centres. At the heart of the analysis are Eurostat statistics about enterprises buying cloud computing services used over the internet. Analysis provides insights by country, sector, and enterprise size. Forecasts use growth rates observed in the countries studied from 2014 to 2020 to extrapolate growth in enterprises buying cloud computing services used over the internet to 2030. **Enterprise insights are important, but as the next section highlights, enterprises can be any size from one to over a million people. Standardisation methods are therefore required to better estimate the *number* of people employed in enterprises buying cloud computing services.**

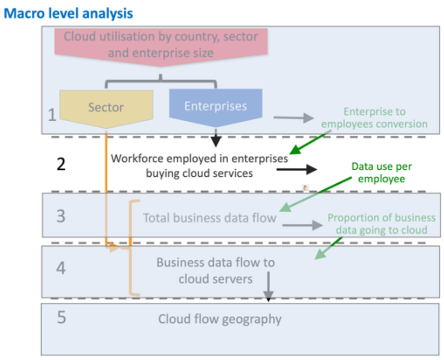
3.1.3 Stage Two: Cloud use by enterprise

Eurostat data and study estimates to 2030 of enterprises buying cloud computing services used over the internet undertaken in stage one provided a foundation for the study. However, a problem with enterprise statistics is that enterprises can range in size from one employee to many thousands of employees.

If there are two enterprises buying cloud services used over the internet and one had one employee and the other had 10,000 employees one would expect the larger enterprise to have much higher levels of data flowing to cloud and edge data servers than the enterprise with one employee.

To address this problem, and better understand the volumes of data used by enterprises of different sizes, it was necessary to use standardisation techniques to estimate the number of employees in enterprises buying cloud services used over the internet. Analysis provided estimates by country and NACE sectors.

The preceding analysis **provided a robust overview of enterprises buying cloud services used over the internet**. Enterprise data is used in the second stage of analysis (see the adjacent diagram) to undertake a standardisation process to estimate the data for this.



Section 2.2 briefly introduced the need for standardisation by using Belgian statistics to highlight that enterprises can range in size from one employee to many thousands of employees. This is

problematical for this study which has the goal of estimating data flows to cloud and edge data servers from enterprises buying cloud services used over the internet.

If there are two enterprises buying cloud services used over the internet and one had one employee and the other had 10,000 employees one would expect the larger enterprise to have much higher levels of data flowing to cloud and edge data servers than the enterprise with one employee. All other things being equal[169] it would be reasonable to assume that the enterprises with 10,000 employees might have approximately 10,000 time more data flowing to cloud and edge data servers than the enterprise with one employee. This difference highlights the need to know more about the number of people working in enterprises buying cloud services used over the internet. This problem can be addressed by using standardisation methods. Standardisation methods are briefly described here. A more detailed overview of the methodology, with additional examples, is provided in the methodological Annex.

3.1.3.1 The need for standardisation

This problem in estimating data flows to cloud and edge data servers from enterprises buying cloud services used over the internet arises because Eurostat provides data about the number of enterprises buying cloud services used over the internet. Eurostat does not provide information about the number of people working in the enterprises buying cloud services used over the internet.

Knowing the number of enterprises buying cloud services used over the internet is not a great deal of help in estimating data flows. There are not figures for average internet traffic per enterprise nor are there figures for average volume of data flowing to cloud and edge data servers per enterprise. As it was shown in the example above, (an enterprise can have one or thousands of employees) enterprises are so diverse in workforce size, and activities (NACE sectors), that average values would be hard to calculate and could provide spurious results if applied to a random selection of companies[170].

The next stage of this study uses an innovative method to calculate average internet traffic per employee. The final stage uses further innovative methods to estimate the proportion of internet traffic per employee that flows to cloud and edge data servers from employees working in enterprises buying cloud services used over the internet.

In the next two stages it is important to know the number of employees working in enterprises buying cloud services used over the internet so that calculations of average internet traffic per employee (stage 3) and average internet traffic per employee flowing to cloud and edge data servers (stage 4) can be applied robustly in the study.

Once again Belgium is used to provide an example of how the standardisation process was applied. Stage one analysis presented Eurostat data that revealed that in 2018 the 635,500 enterprises in Belgium employed 4.795 million people.

Eurostat data also revealed that 40 per cent of enterprises in Belgium bought cloud computing services used over the internet.

169 A comparison with ‘all things equal’ would compare two enterprises (one with one employee the other with 10,000 employees) in the same sector, in the same country or region and at the same time. It is these parameters country, sector and year that are used in the standardisation process.

170 The nearer one got to the population of enterprises in a country, the more reliable estimates using average values would become.

A simple method of standardisation would be to suggest that if 40 per cent of enterprises in Belgium bought cloud computing services used over the internet it would not be unreasonable to assume that 40 per cent of employees would probably be employed in enterprises buying cloud computing services used over the internet. This would equate to 1,918,000 people employed in enterprises buying cloud computing services used over the internet.

The figure of 1,918,000 employees is useful overall estimate, but it can be further refined and made more robust.

Eurostat provide statistics for the percentage of enterprises buying cloud computing services used over internet for four different sizes (employees) of enterprises. Greater accuracy can be achieved if the preceding method is applied to the four different sizes of enterprise. Once again Belgium is used as an example.

Eurostat provide enterprise data by size bands for employees[171]. For example, in 2018 in Belgium:

Enterprises with '0 to 9' persons employed: There are 602,300 enterprises in Belgium with '0 to 9' persons employed. In total they employ 1.60 million employees. Eurostat data for 2018 reveals that approximately 14 per cent of enterprises in this size band use cloud services[172]. Analysis uses the previous assumption that 14 per cent of the workforce are likely to be employed in enterprises buying cloud computing services used over the internet. This would equate to **225,150** people employed in enterprises buying cloud computing services used over the internet.

Enterprises with '10 to 49' persons employed: There are 26,600 enterprises with '10 to 49' persons employing 834,100 employees in Belgium in 2018. Eurostat estimate that 36 per cent of enterprises in this size band use cloud services. Using the previous assumption that 36 per cent of the workforce are likely to be employed in enterprises buying cloud computing services used over the internet. This would equate to **300,270** people employed in enterprises buying cloud computing services used over the internet.

Enterprises with '50 to 249' persons employed: There are 4,100 enterprises in Belgium with '50 to 249' persons employed. In total they employ 673,600 employees. Eurostat data for 2018 reveals that 55 per cent of enterprises in this size band use cloud services. Analysis uses the previous assumption that 55 per cent of the workforce are likely to be employed in enterprises buying cloud computing services used over the internet. This would equate to **370,480** people employed in enterprises buying cloud computing services used over the internet.

Enterprises with 'more than 250' persons employed: There are 4,100 enterprises with 'over 250' persons employing 1.67 million employees in Belgium in 2018. Eurostat estimate that 79 per cent of enterprises in this size band use cloud services. Using the previous assumption that 79 per cent of the workforce are likely to be employed in enterprises buying cloud computing services used over the internet. This would equate to **1,326,560** people employed in enterprises buying cloud computing services used over the internet.

171 Eurostat has data for EU 27MS, EFTA (excluding Liechtenstein) and UK

172 Eurostat data for this indicator (0-9 employees) is sparse, thus this per cent has been calculated through thorough estimations.

The above approach is more refined than the previous broad based national analysis which estimated 1,918,000 people employed in enterprises buying cloud computing services used over the internet.

When one totals the four estimates (for company size bands) above the figures estimate 2,222,460 people employed in enterprises buying cloud computing services used over the internet. This estimate is 304,460 (13.6 per cent) higher than the broad based national analysis.

The methodological annex describes why a more detailed approach, examining enterprise size bands, provides better and usually higher estimates of people employed in enterprises buying cloud computing services used over the internet than broad based national analysis. The main underlying reason is because in most economies a relatively small proportion of enterprises (4,100; 0.6 per cent of all enterprises in Belgium in 2018) with more than 250 employees, which are high adopters of buying cloud computing services used over the internet (79 per cent in Belgium), employ a relatively large proportion of the workforce (35 per cent).

The annex also highlights how the approach was further enhanced by undertaking the above analysis for each of the NACE sectors studied (C to S) for each enterprise size band in each country studies for every year for which data was provided by Eurostat (2016 to 2020) or for forecasts of future trends from 2021 to 2030). The analysis required almost 30,000 computations.

This more detailed analysis derived an estimate of 2,480,000 people in Belgium in 2018 employed in enterprises buying cloud computing services used over the internet.

3.1.3.2 Standardisation of the Workforce in enterprises that utilises cloud services in EU27 Member States, EFTA countries and the UK

Research has replicated the preceding Belgian manufacturing example, using the Eurostat data for 2020 for all 31 countries examined in the study.[173] Data about the percentage of enterprises utilising low, medium, and high cloud services in 19 NACE sectors for four size groups of enterprises was calculated per year for the EU27 Member States, EFTA countries and the UK using the 2018 Eurostat data. This was then transformed to provide data about the number of people in each of those groups that were employed in enterprises that utilise cloud services.

The Figure below provides details of the number of people employed in enterprises buying cloud services used over the internet[174] for each of the 31 countries examined in the study. The highest number of people employed in these enterprises was 16.6 million workers in the UK. These workers represent 53 per cent of the UK workforce, which is 31.1 million.[175]

Germany has a larger workforce than the UK: 40.6 million in 2020. Yet, Germany had a smaller number of people working in enterprises using cloud (13.3 million) because the

173 The 31 countries include EU27 Member States, three EFTA countries (Iceland, Norway and Switzerland) and the UK. Liechtenstein is a very small EFTA country but due to an absence of Eurostat data for many of the statistics utilised in this study it is not included in analysis.

174 The figure includes all enterprises using low, medium and high cloud services.

175 Eurostat indicator LFSI_EMP_A (https://ec.europa.eu/eurostat/databrowser/view/lfsi_emp_a/default/table?lang=en)

overall proportion of enterprises connected to cloud services is lower. Thus, in Germany 33 per cent of employees are in enterprises buying cloud services used over the internet.

Figure 16: The number of people employed in enterprises buying cloud services used over the internet by country in 2020

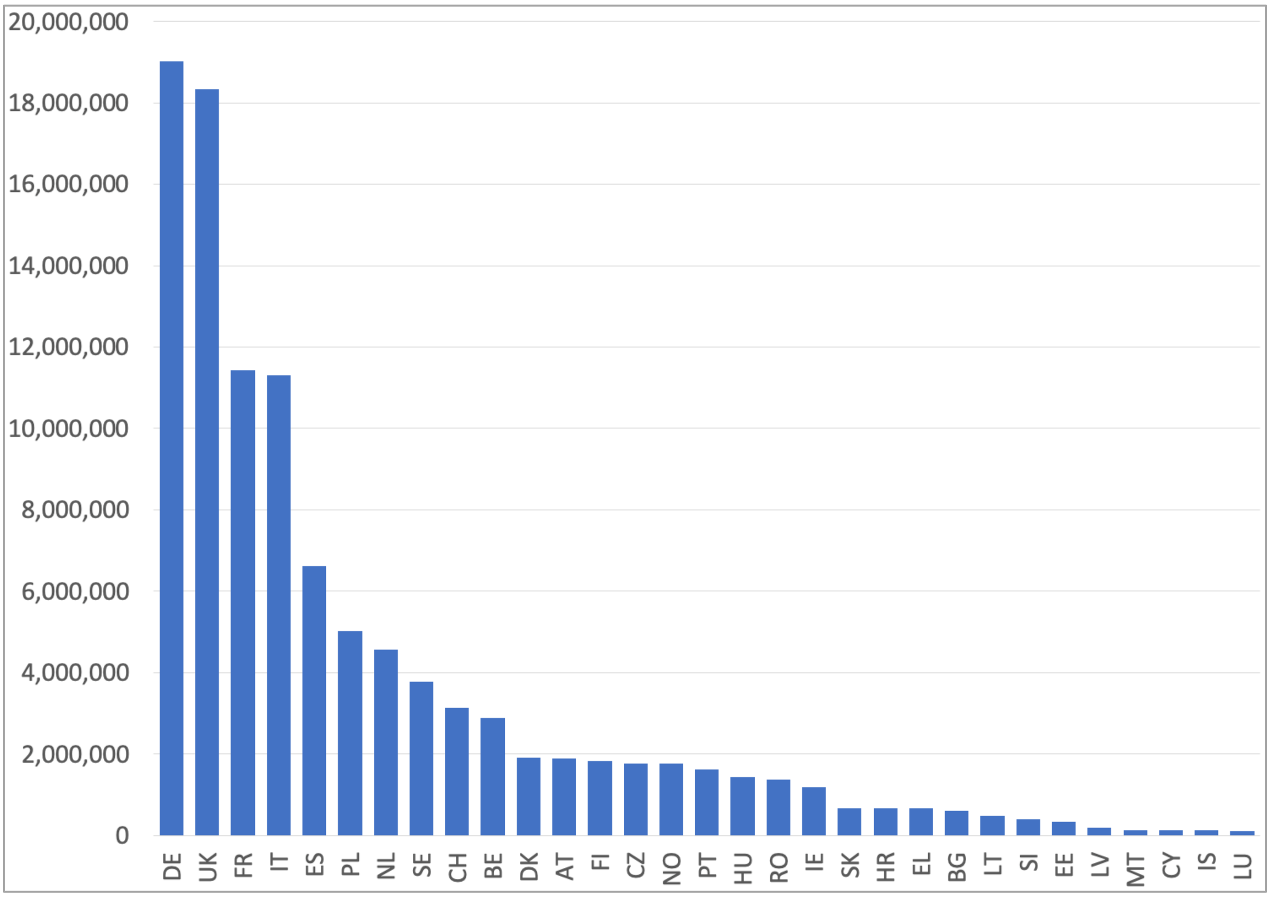


Table 6 provides a further breakdown of people employed in enterprises by sector and country, that buy cloud services used over the internet.[176]

In total, in EU27, EFTA countries, and the UK, it is estimated that in NACE sectors C to S there are 63.41 million people working in enterprises that buy cloud services used over the internet.[177] This represents 27.5 per cent of the working population - 231 million[178] - for the 31 countries examined in this study.

176 The Eurostat isoc_cicce_use data set does not provide data for Agriculture (NACE A) and Mining (B). Eurostat enterprise and cloud data had several missing cells for sectors and countries. Eurostat does not provide data for Liechtenstein (population 38,000) - an EFTA country. Values in missing cells were extrapolated from appropriate data 'pro-rata' for the same country where appropriate. Data is not provided for Finance and Insurance Services (K). ICT (J) connectivity levels are used to extrapolate connectivity levels for Finance. The only public sector NACE sector is administrative and support services (N). This level was therefore extrapolated to NACE public sectors P to S.

177 The figure includes all enterprises using low, medium and high cloud services.

178 Eurostat Ibid. Indicator LFSI_EMP_A (https://ec.europa.eu/eurostat/databrowser/view/lfsi_emp_a/default/table?lang=en)

Table 6:The number of people employed in enterprises by sector and country that buy cloud services used over the internet in 2020

	All NACE Sectors (C to S)	NACE Sectors (C to M)	NACE Sectors (N to S)
EU27 Member States	44%	41%	50%
EFTA	70%	69%	72%
UK	58%	55%	61%
Total	47%	44%	52%
Belgium	61%	54%	71%
Bulgaria	21%	19%	26%
Czechia	35%	33%	40%
Denmark	69%	68%	71%
Germany	45%	44%	48%
Estonia	53%	51%	57%
Ireland	55%	51%	62%
Greece	19%	16%	25%
Spain	37%	31%	48%
France	44%	41%	49%
Croatia	43%	42%	46%
Italy	53%	49%	64%
Cyprus	34%	33%	35%
Latvia	24%	24%	24%
Lithuania	38%	36%	44%
Luxembourg	43%	39%	50%
Hungary	34%	32%	36%
Malta	54%	50%	60%
Netherlands	58%	51%	68%
Austria	46%	43%	50%
Poland	34%	32%	41%
Portugal	37%	34%	42%
Romania	21%	20%	24%
Slovenia	43%	41%	48%
Slovakia	27%	28%	26%
Finland	75%	72%	81%
Sweden	76%	73%	80%
Iceland	65%	64%	67%
Norway	68%	65%	72%
Switzerland	71%	71%	72%
United Kingdom	58%	55%	61%

Table 7: The number of people employed in private enterprises by sector and country that buy cloud services used over the internet in 2020. NACE sectors C to M

	C Manufacturing	D Electricity & Gas	E Water Supply	F Construction	G Wholesale & Retail	H Transportation	I Accomm & Food	J ICT	L Real Estate	M Prof, Scientific & Technical
EU27 Member States	49%	54%	52%	31%	39%	47%	32%	46%	29%	31%
EFTA	76%	81%	68%	61%	69%	71%	66%	70%	59%	64%
UK	60%	70%	63%	42%	60%	63%	55%	53%	49%	48%
Total	52%	57%	52%	36%	43%	50%	39%	47%	33%	34%
Belgium	68%	83%	58%	42%	50%	66%	36%	55%	49%	38%
Bulgaria	24%	34%	31%	17%	14%	22%	15%	23%	13%	12%
Czechia	40%	42%	38%	21%	29%	39%	23%	35%	17%	21%
Denmark	75%	67%	70%	58%	69%	74%	62%	67%	56%	63%
Germany	51%	48%	48%	33%	43%	48%	35%	47%	31%	37%
Estonia	61%	66%	59%	40%	51%	53%	51%	52%	51%	33%
Ireland	52%	52%	53%	35%	53%	52%	51%	59%	39%	45%
Greece	21%	30%	23%	11%	15%	19%	14%	15%	16%	11%
Spain	39%	46%	52%	20%	29%	34%	24%	43%	11%	24%
France	50%	61%	55%	30%	39%	52%	29%	48%	33%	29%
Croatia	48%	58%	51%	34%	43%	48%	34%	42%	28%	25%
Italy	58%	70%	67%	37%	43%	62%	40%	57%	26%	34%
Cyprus	36%	63%	34%	26%	32%	40%	39%	42%	25%	27%
Latvia	27%	35%	31%	20%	23%	30%	23%	26%	16%	13%
Lithuania	44%	44%	43%	31%	33%	41%	31%	35%	16%	20%
Luxembourg	41%	39%	40%	35%	38%	45%	35%	41%	21%	39%
Hungary	43%	53%	47%	19%	25%	41%	23%	30%	18%	18%
Malta	49%	49%	49%	48%	50%	57%	51%	54%	37%	43%
Netherlands	59%	56%	66%	43%	54%	62%	42%	52%	44%	41%
Austria	54%	44%	47%	39%	44%	48%	32%	42%	42%	29%
Poland	41%	54%	39%	20%	28%	34%	22%	33%	21%	19%
Portugal	38%	43%	49%	28%	31%	41%	28%	41%	32%	26%
Romania	25%	32%	27%	15%	16%	21%	15%	23%	12%	14%
Slovenia	51%	54%	50%	29%	41%	42%	31%	35%	25%	22%
Slovakia	35%	42%	39%	15%	23%	31%	24%	28%	24%	17%
Finland	82%	82%	73%	64%	72%	71%	67%	77%	54%	62%
Sweden	82%	84%	81%	65%	73%	76%	69%	76%	65%	65%
Iceland	74%	82%	74%	48%	68%	77%	65%	65%	33%	49%
Norway	74%	77%	70%	57%	67%	68%	63%	70%	40%	57%
Switzerland	76%	83%	67%	64%	70%	73%	67%	70%	68%	68%
United Kingdom	60%	70%	63%	42%	60%	63%	55%	53%	49%	48%

(Predominantly private sector)

Table 8: The number of people employed in public enterprises by sector and country that buy cloud services used over the internet in 2020. NACE sectors N to S

	N Administrative and Support Services	P Education	Q Human, Health and Social Work	R Arts, Entertainment, Recreation	S Other Service Activities
EU27 Member States	50%	50%	50%	50%	50%
EFTA	72%	72%	72%	72%	72%
UK	61%	61%	61%	61%	61%
Total	52%	52%	52%	52%	52%
Belgium	71%	71%	71%	71%	71%
Bulgaria	26%	26%	26%	26%	26%
Czechia	40%	40%	40%	40%	40%
Denmark	71%	71%	71%	71%	71%
Germany	48%	48%	48%	48%	48%
Estonia	57%	57%	57%	57%	57%
Ireland	62%	62%	62%	62%	62%
Greece	25%	25%	25%	25%	25%
Spain	48%	48%	48%	48%	48%
France	49%	49%	49%	49%	49%
Croatia	46%	46%	46%	46%	46%
Italy	64%	64%	64%	64%	64%
Cyprus	35%	35%	35%	35%	35%
Latvia	24%	24%	24%	24%	24%
Lithuania	44%	44%	44%	44%	44%
Luxembourg	50%	50%	50%	50%	50%
Hungary	36%	36%	36%	36%	36%
Malta	60%	60%	60%	60%	60%
Netherlands	68%	68%	68%	68%	68%
Austria	50%	50%	50%	50%	50%
Poland	41%	41%	41%	41%	41%
Portugal	42%	42%	42%	42%	42%
Romania	24%	24%	24%	24%	24%
Slovenia	48%	48%	48%	48%	48%
Slovakia	26%	26%	26%	26%	26%
Finland	81%	81%	81%	81%	81%
Sweden	80%	80%	80%	80%	80%
Iceland	67%	67%	67%	67%	67%
Norway	72%	72%	72%	72%	72%
Switzerland	72%	72%	72%	72%	72%
United Kingdom	61%	61%	61%	61%	61%

(Predominantly public sector)

Conclusion

This section introduced the standardisation process to estimate the number of employees in enterprises buying cloud services used over the internet. Eurostat data about various enterprise and workforce sizes was used to highlight the need for standardisation to indicate how size contribution in terms of employees does not always reflect their contribution to the workforce. Belgium was used to provide a simple overview of the approach.

It was highlighted that a more granular approach was used in developing the analysis to create tables in this section. Analysis was further enhanced by using the standardisation methodology to undertake analysis for each of the NACE sectors studied (C to S) for each enterprise size band in each country studies for every year for which data was provided by Eurostat (2016 to 2020) or for forecasts of future trends from 2021 to 2030).

Further details of the rationale for the approach and how more than 30,000 computations were undertaken using the standardisation methodology and be found in the Methodological Annex.

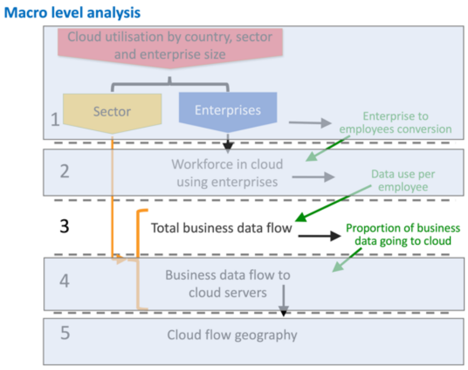
3.1.4 Stage Three: Data flow per employee

The third stage of the study focused on internet traffic generated by employees. ITU statistics from 2016 to 2020 were used to estimated internet traffic in the 31 countries investigated in the study. Analysis used ITU and Cisco internet traffic growth rates to forecast internet traffic until 2030.

The focus of this study is enterprise data flowing to cloud and edge data centres. Cisco statistics were used to find the percentage of internet traffic generated by enterprises.

Estimates of the enterprise internet traffic generated by enterprises in each country provided the basis for analysis to estimate average internet traffic (calculated as GB/month) per employee.

3.1.4.1 Data use by employees



This third stage of analysis utilises ITU, Cisco and TeleGeography statistics to estimate the total internet traffic for enterprises. This enables average traffic flow per employee to be calculated. The information acquired is then used in analysis when estimating enterprise data flowing to cloud and edge data centres in Stage 4.

Our research team could not find information on the average internet traffic *per employee* in Europe[179]. Therefore, our starting point for estimating this figure was to examine annual internet traffic[180] and then to use this information to estimate the average volume of internet traffic used by enterprises.

Using Belgium as an example helps to show the analysis undertaken in this stage of the research. Further details are provided below and in the methodological annex.

Stage one analysis presented Eurostat data that revealed that in 2018 the Belgian workforce was comprised of 4.795 million people. Analysis described later in this section estimates that enterprise generated internet traffic in Belgium in 2018 was 1.6 EX

179 EU27 Member States, EFTA countries and the UK.

180 ITU describe 'traffic flows' as 'traffic generated by fixed-broadband subscribers measured at the end user access point by adding upload and download traffic'. Data coverage became more extensive in 2016.

(exabytes). Dividing the size of the workforce (4.795 million) into Belgian internet traffic flow in 2018 (1.6 EX per annum) provides an average monthly internet traffic flow of 29.8 GB per employee. These workforce averages are used in the next stage to estimate the proportion of average employee data that is flowing to main cloud and edge data centres in enterprises buying cloud services used over the internet.

A handful of studies provide estimates of all internet traffic in Europe. The relevant studies include International Telecommunication Union (ITU) statistics and data from Cisco and TeleGeography for annual internet traffic in Europe.

International Telecommunication Union: The ITU, originally the International Telegraph Union, is a specialised agency of the United Nations that is responsible for collecting data on information and communication technologies. Internet traffic information is collected from national telecommunication/ICT ministries and regulatory authorities. The ITU should therefore be a robust and reliable source of data for the context of this study.

The ITU provides data for fixed broadband traffic (exabytes (EX) per annum). The metadata states that the data relates to traffic generated by fixed-broadband subscribers measured at the end user access point.[181] The year 2018 is the latest year for which ITU provide data; likewise, this is also the latest year for which Eurostat cloud utilisation information is available.

ITU statistics are provided for 14 of the EU27 Member States and Switzerland[182]. As the next section describes, ITU statistics were extrapolated pro-rata to provide forecasts for total internet traffic for the 31 countries in this study, see Table 9. Table 10 uses the data to provide forecasts for internet traffic generated by enterprise. The growth rate for the ten EU27 Member States where information was available from ITU for 2016 to 2019 was 22 per cent (CAGR).[183]

Two studies provide complementary insights to internet traffic in Europe but it should be pointed out that neither study provides a source or methodology for their estimates. The studies are:

> ***Cisco:*** Cisco[184] estimates EU27 internet traffic of 270 EX a year (775,000 TB/day, and 23,593,000 TB/month) in 2018.[185] They forecast CAGR of 22 per cent to 2022. No source is provided for these estimates. If the Cisco forecast for growth was maintained, study team analysis estimates that internet traffic would reach 2,938 EX

181 ITU metadata further stated the value is measured 'adding up and download traffic'. No other metadata information was provided. Data coverage became more extensive in 2016. Metadata is provided by ITU.

182 Belgium, Bulgaria, Croatia, Czechia, Denmark, Germany, Greece, Hungary, Ireland, Italy, Portugal, Romania, Slovakia and Spain.

183 Intuitively this figure seems high, but this could be because of considerable growth in the use of video applications (Netflix, Amazon Prime, YouTube, Hulu, Peacock, Disney etc.) which have very high bandwidth requirements. Growth in audio streaming applications (Spotify, Apple Music, Amazon Music, Deezer, Tidal etc.) could also be a contributory factor.

184 'Cisco visual networking index: Forecasts and trends 2017 to 2022' (https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-741490.html) and Cisco slides (particularly slide 8) can be found at (https://www.cisco.com/c/dam/m/en_us/network-intelligence/service-provider/digital-transformation/knowledge-network-webinars/pdfs/1213-business-services-ckn.pdf)

185 Cisco forecast internet traffic in 'Western Europe of 49.9 EX per month by 2022, growing at a CAGR of 22 per cent'. Study team analysis 'reversed' the CAGR estimate of data flows to calculate 270 EX a year in 2018 for Western Europe. In effect this was reverse extrapolation, extrapolating the CAGR rate to a previous year (2018). This then enables comparison with 2018 data provided by the ITU and TeleGeography. Page 3 of Cisco Visual networking Index (2017 to 2022) document states that 'By 2022, We will have 380 million Internet users (89 percent of population), up from 358 million (85 percent of population) in 2017'. This equates to a total Western Europe population size of 427 million. The EU27 population size is 444 million. A discrepancy exists but for ease of comparison Western Europe is assumed to mean EU27.

a year in 2030 (8,450,000 TB/day, and 256,726,000 TB/month). Cisco does not provide an insight to how they forecast growth of 22 per cent.

TeleGeography: TeleGeography[186] estimates EU27 internet traffic of 267.3 EX in their 2018 internet map. The source and methodology for this estimate is not provided. Forecasts of future growth are also not provided. The map also provides estimates of internet traffic in Asia for 2018 (79.5 EX), Canada, USA (90.3 EX) and Latin America (45.6 EX).

Together, the three sources provide relatively similar estimates for internet traffic in 2018: ITU 225 EX a year, Cisco 270 EX and TeleGeography 267 EX. The difference between the estimates is approximately 20 per cent. Since the ITU estimates are derived from country-level data, rather than (Cisco and TeleGeography's) world-wide estimates, the study team uses ITU estimates for further analysis in this study. Sensitivity analysis[187] will enable different estimates for European internet traffic to be used in the analysis. It is interesting that both the ITU study and Cisco study forecast internet traffic growth of 22 per cent per annum. Since there is agreement between the two sources this figure (22 per cent) is used in the analysis in the remainder of this section to forecast internet traffic growth from 2020 to 2030.

Tender Specifications stipulate that the objective of the study is 'to provide a mapping and analysis of enterprise data flowing to cloud and edge data centres within and across EU Member States and EFTA countries'.[188] The study is also required to examine enterprise data flowing to cloud and edge data centres in private and public sectors (NACE C to M and N to S, respectively). It was therefore important for our team to find evidence or estimates for the proportion of internet traffic generated by enterprises (in NACE C to S) compared to the proportion from other sources - predominantly households.

The first step in estimating internet traffic generated by enterprises was analysis to estimate internet traffic from all sources (i.e household and enterprise internet traffic). This analysis is described below.

As noted above, the ITU provides annual information about internet traffic. This data was available for 14 of the EU27 Member States and Switzerland. To address the gaps, extrapolation methods were undertaken to calculate internet traffic volume in countries where ITU data was not available. The multiple regression analysis, using machine learning methods to estimate internet traffic in the 17 countries for which ITU did not provided information are described in the methodological annex.

The methodological annex also describes how Cisco's estimate of 22 per cent growth was used to forecast growth in internet traffic[189]. This analysis estimates that **305 EX/year of internet traffic is flowing in EU27 Member States in 2020**, see Table 9.

Predictions by the study team estimate there will be 825 EX/year of internet traffic flowing in EU27 Member States in 2025. In 2030, forecasts estimate that there will be 2,229 EX/year of total internet traffic. The Table overleaf provides an overview of total internet traffic forecasts from 2016 to 2030 for the 31 countries included in the study. These total internet traffic forecasts will include traffic for households and enterprises.

186 TeleGeography 2018. 'Global internet map 2018' (https://global-internet-map-2018.telegeography.com/).

187 Sensitivity analysis in the model developed for the study enables alternative values for many components to be changed from default values to examine the effect of different assumptions on the outcomes and impacts investigated.

188 Tender Specifications page 11.

189 'Cisco visual networking index: Forecasts and trends 2017 to 2022'.

Table 9: Internet traffic (EX/year) in the 31 countries studied 2016 to 2030

	2016	2017	2018	2019	2020	2025	2030
EU27	138	168	205	250	305	825	2,229
EFTA	8	10	12	14	18	48	128
UK	30	36	44	54	66	178	481
Total	176	214	261	318	389	1,050	2,838
Belgium	4.5	5.5	6.7	8.1	9.9	26.8	72.4
Bulgaria	1.9	2.4	2.9	3.5	4.3	11.7	31.5
Czechia	3.6	4.3	5.3	6.5	7.9	21.3	57.6
Denmark	3.6	4.4	5.4	6.6	8.1	21.8	58.8
Germany	27.0	32.9	40.2	49.0	59.8	161.6	436.7
Estonia	0.7	0.8	1.0	1.2	1.5	4.0	10.8
Ireland	2.1	2.6	3.1	3.8	4.7	12.6	34.0
Greece	2.5	3.1	3.8	4.6	5.6	15.2	41.1
Spain	14.4	17.6	21.5	26.2	31.9	86.3	233.3
France	19.4	23.7	28.9	35.3	43.1	116.4	314.7
Croatia	0.8	1.0	1.3	1.5	1.9	5.1	13.7
Italy	15.1	18.5	22.5	27.5	33.6	90.7	245.1
Cyprus	0.3	0.3	0.4	0.5	0.6	1.6	4.4
Latvia	0.8	0.9	1.1	1.4	1.7	4.6	12.5
Lithuania	1.2	1.4	1.8	2.2	2.6	7.1	19.2
Luxembourg	0.4	0.5	0.5	0.7	0.8	2.2	6.0
Hungary	2.7	3.3	4.0	4.8	5.9	16.0	43.1
Malta	0.2	0.3	0.3	0.4	0.5	1.3	3.5
Netherlands	8.3	10.1	12.3	15.0	18.3	49.4	133.6
Austria	3.5	4.2	5.2	6.3	7.7	20.8	56.1
Poland	8.3	10.1	12.3	15.0	18.3	49.5	133.8
Portugal	3.3	4.0	4.9	5.9	7.2	19.6	52.8
Romania	3.7	4.6	5.6	6.8	8.3	22.4	60.6
Slovenia	0.8	1.0	1.2	1.5	1.8	4.8	13.0
Slovakia	1.4	1.8	2.1	2.6	3.2	8.6	23.3
Finland	2.7	3.3	4.0	4.9	5.9	16.1	43.4
Sweden	4.6	5.6	6.8	8.3	10.1	27.3	73.9
Iceland	0.2	0.3	0.3	0.4	0.5	1.3	3.6
Norway	2.7	3.3	4.0	4.9	6.0	16.2	43.7
Switzerland	5.0	6.1	7.5	9.1	11.1	30.0	81.2
United Kingdom	29.7	36.2	44.2	53.9	65.8	177.8	480.7

Source: Study team further analysis and extrapolation from ITU data Indicator i135tfb from the ITU WTID subscription data set.

To replicate the methodology to estimate internet traffic per employee in Belgian it is necessary to analyse data from Table 9 and estimate the proportion of total internet traffic in the table that is generated by enterprises.

Cisco estimates that in 2020, fixed broadband devices used by enterprises comprised 24.8 per cent of internet traffic. The remaining traffic (75.2 per cent) was attributed to households. Note that Cisco do not provide a source or methodology for this estimate, nor could any other sources be found. Only Cisco provides this information. Using the Cisco figure of 24.8 per cent of internet traffic generated by enterprises in 2020 it is possible to estimate that enterprise internet traffic is 75.7 EX/year in 2020[190].

Cisco predict the proportion of total internet traffic that is used by enterprises is steadily increasing. In 2023 they predict 26 per cent of traffic will be generated by enterprises[191].

190 Table 9 estimates that total internet traffic in 2020 is 305 EX/year. If this figure is multiplied by the percentage of internet traffic generated by enterprise (24.8 per cent) on can estimate that enterprise generated internet traffic in 2020 is 75.7 EX/year.

191 Cisco annual internet report page 6 (https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.pdf). Cisco forecast that in 2023 consumer share of devices will be 74 per cent. Consumer growth is expected to grow at 9.1 per cent CAGR and business growth at 12 per cent. Using the Cisco estimate for 2023, our team extrapolated 'backwards' an estimate of 24 per cent for the business share in 2018.

If this rate of increase in internet traffic generated by enterprises (24.8 per cent in 2020 and 26 per cent in 2023) is sustained the study team forecast that 26.9 per cent of internet traffic will be generated by enterprise in 2025 and 29.2 per cent in 2030.

In 2025, the study team have forecasted that 26.9 per cent of internet traffic can be attributed to enterprises, thus enterprise internet traffic in 2025 is estimated to be 222 EX/year (825 EX/year (from Table 9) x 26.9 per cent). In 2030, the study team estimate that there will be 2,229 EX/year of total internet traffic and 29.2 per cent will be attributed to enterprises, thus enterprise internet traffic is 651EX/year (2,229 EX/year x 29.2 per cent)[192].

Table 10 provides forecasts for internet traffic generated by enterprise from 2016 to 2030 for the 31 countries included in the study.

Table 10: Forecast of internet traffic (EX/year) generated by enterprise 2016 to 2030

	2016	2017	2018	2019	2020	2025	2030
EU27	32.0	39.7	49.2	61.0	75.7	222.0	651.0
EFTA	1.8	2.3	2.8	3.5	4.4	12.8	37.5
UK	6.9	8.6	10.6	13.2	16.3	47.9	140.4
Total	40.7	50.6	62.6	77.7	96.4	282.6	828.9
Belgium	1.0	1.3	1.6	2.0	2.5	7.2	21.1
Bulgaria	0.5	0.6	0.7	0.9	1.1	3.1	9.2
Czechia	0.8	1.0	1.3	1.6	2.0	5.7	16.8
Denmark	0.8	1.0	1.3	1.6	2.0	5.9	17.2
Germany	6.3	7.8	9.6	12.0	14.8	43.5	127.5
Estonia	0.2	0.2	0.2	0.3	0.4	1.1	3.2
Ireland	0.5	0.6	0.8	0.9	1.2	3.4	9.9
Greece	0.6	0.7	0.9	1.1	1.4	4.1	12.0
Spain	3.3	4.2	5.2	6.4	7.9	23.2	68.1
France	4.5	5.6	6.9	8.6	10.7	31.3	91.9
Croatia	0.2	0.2	0.3	0.4	0.5	1.4	4.0
Italy	3.5	4.4	5.4	6.7	8.3	24.4	71.6
Cyprus	0.1	0.1	0.1	0.1	0.1	0.4	1.3
Latvia	0.2	0.2	0.3	0.3	0.4	1.2	3.6
Lithuania	0.3	0.3	0.4	0.5	0.7	1.9	5.6
Luxembourg	0.1	0.1	0.1	0.2	0.2	0.6	1.7
Hungary	0.6	0.8	1.0	1.2	1.5	4.3	12.6
Malta	0.0	0.1	0.1	0.1	0.1	0.3	1.0
Netherlands	1.9	2.4	2.9	3.7	4.5	13.3	39.0
Austria	0.8	1.0	1.2	1.5	1.9	5.6	16.4
Poland	1.9	2.4	3.0	3.7	4.5	13.3	39.1
Portugal	0.8	0.9	1.2	1.4	1.8	5.3	15.4
Romania	0.9	1.1	1.3	1.7	2.1	6.0	17.7
Slovenia	0.2	0.2	0.3	0.4	0.4	1.3	3.8
Slovakia	0.3	0.4	0.5	0.6	0.8	2.3	6.8
Finland	0.6	0.8	1.0	1.2	1.5	4.3	12.7
Sweden	1.1	1.3	1.6	2.0	2.5	7.4	21.6
Iceland	0.1	0.1	0.1	0.1	0.1	0.4	1.0
Norway	0.6	0.8	1.0	1.2	1.5	4.4	12.8
Switzerland	1.2	1.4	1.8	2.2	2.8	8.1	23.7
United Kingdom	6.9	8.6	10.6	13.2	16.3	47.9	140.4

The introductory section about the analysis undertaken in this stage of analysis used Belgium as a case study to describe how average internet traffic per month per employee was calculated. This method of dividing enterprise generated internet traffic by the size of

192 EFTA: 2020 total internet traffic is 17.2 EX/year, enterprise internet traffic is 4 EX/year (17.2 x 24.8 per cent). 2025 total internet traffic is 47.2 EX/year, enterprise traffic is 13 EX/year (47.2 x 26.9 per cent). 2030 total internet traffic is 128.3 EX/year, enterprise internet traffic is 38 EX/year (128.3 x 29.2 per cent). UK: 2020 total internet traffic is 65.8 EX/year, enterprise internet traffic is 16 EX/year (65.8 x 24.8 per cent). 2025 total internet traffic is 177.8 EX/year, enterprise internet traffic is 48 EX/year (177.8 x 26.9 per cent). 2030 total internet traffic is 480.7 EX/year, enterprise internet traffic is 140 EX/year (480.7 x 29.2 per cent).

the workforce in a country was replicated across the 31 countries included in the study to provide the GB/month internet traffic per month per employee estimates presented in Table 11.

In 2020 in EU27 Member States average internet traffic per employee is 34 GB/month. This is expected to grow 2.7 times to 94 GB/month in 2025. A similar level of growth (2.8 fold) is expected to increase average internet traffic per employee to 264 GB/month in 2030.

Table 11: Estimates of internet traffic per employee (GB/month) 2016 to 2030

	2016	2017	2018	2019	2020	2025	2030
EU27	15	18	22	27	34	94	264
EFTA	22	28	34	42	52	142	395
UK	19	24	29	36	45	123	342
Total	16	19	23	29	36	100	280
Belgium	20.3	24.8	29.8	36.4	45.5	127.5	358.4
Bulgaria	13.3	15.6	19.4	23.4	30.2	82.4	231.0
Czechia	14.3	17.4	21.4	26.6	33.4	91.6	256.8
Denmark	27.2	33.4	40.8	49.7	62.7	171.4	477.4
Germany	13.5	16.5	20.4	25.2	30.9	85.6	238.7
Estonia	21.3	25.9	32.1	39.2	50.5	137.8	389.1
Ireland	20.2	24.6	29.6	35.8	45.1	114.9	304.4
Greece	14.1	17.1	20.8	25.3	31.8	86.6	239.4
Spain	16.2	19.5	23.6	28.8	37.0	97.3	265.7
France	15.1	18.6	22.9	28.4	35.4	99.9	285.3
Croatia	10.8	13.0	16.0	19.7	24.8	69.1	195.4
Italy	13.7	16.8	20.7	25.6	32.6	91.2	261.0
Cyprus	15.2	18.0	21.2	25.6	32.0	87.5	238.9
Latvia	17.9	22.0	26.8	33.3	42.5	121.4	352.9
Lithuania	18.0	22.4	26.9	34.1	43.4	119.6	340.7
Luxembourg	29.6	34.6	42.3	50.6	62.3	160.2	420.9
Hungary	12.6	15.4	19.0	23.4	29.2	76.3	206.8
Malta	21.0	24.7	28.1	33.2	40.7	97.3	242.6
Netherlands	20.2	24.5	29.8	36.3	45.3	125.4	349.6
Austria	16.8	20.7	25.4	31.3	39.1	109.2	308.5
Poland	10.6	12.9	15.9	19.7	24.6	68.8	195.1
Portugal	14.5	17.5	21.3	26.2	33.4	90.7	252.8
Romania	9.0	10.9	13.5	16.9	21.5	62.4	184.0
Slovenia	18.1	21.2	25.9	32.4	40.3	111.8	313.8
Slovakia	11.9	14.6	17.8	22.0	27.9	74.8	206.8
Finland	22.4	27.6	33.3	40.9	51.8	148.0	425.6
Sweden	19.1	23.1	28.3	34.8	44.1	119.8	334.0
Iceland	23.5	29.2	34.8	43.3	53.8	140.6	378.9
Norway	21.1	26.2	31.9	39.0	49.2	139.0	396.9
Switzerland	22.6	27.9	34.2	42.4	52.6	145.6	408.6
United Kingdom	19.4	23.9	29.3	36.0	45.0	122.9	342.2

These average figures for internet traffic per employee each month can be used in the final stage of analysis to estimate the volume of data flowing from enterprises to cloud and edge data centres. **However, first of all, the proportion internet traffic flowing from enterprises to cloud and edge data centres is required.** This is investigated in the next penultimate stage of analysis described in the next section.

Conclusion

This section utilised, enhanced, and validated ITU data and undertook analysis using multiple regression methods to estimate internet traffic for each country examined in the study.

Both the ITU and Cisco forecast internet traffic growth of 22 per cent per annum. Since there is agreement between the two sources this figure (22 per cent) was used in the analysis in the remainder of this section to forecast internet traffic growth from 2020 to 2030.

Estimates provided by Cisco concerning the proportion of internet traffic generated by enterprise were used to estimate the enterprise internet traffic from 2018 to 2030.

The final stage of analysis estimated internet traffic per employee by dividing internet traffic generated by enterprise in each country by the number of employees in countries.

These estimates provide the starting point for the next section which examines the proportion of total enterprise internet traffic that might be flowing to cloud data centres.

3.1.5 Stage Four: Enterprise data flowing to cloud and edge data centres

This stage describes the methodology developed to undertake the penultimate stage of macro analysis – the estimation of enterprise data flows to cloud and edge data centres by enterprises buying cloud computing services used over the internet.

Stage 4 utilises information from the two preceding stages. Stage 2 provided standardised data about the number of employees in enterprises buying cloud computing services used over the internet. This information was provided for NACE sectors and different sizes of enterprise (employees).

Enterprise internet traffic information calculated in Stage 3, particularly the estimates of average internet traffic (calculated as GB/month) per employee, was also utilised in stage 4. In stage 4, the number of employees in a sector or enterprises of a particular size (stage 2 analysis) was multiplied by average internet traffic per employee (stage 3 analysis). This analysis made it possible to estimate total enterprise internet traffic for the sector or enterprises of a particular size.

The final element of stage 4 analysis estimated the percentage of internet traffic for a sector or enterprise of a particular size that was flowing to cloud service providers.

Enterprise router information, for enterprises buying cloud computing services used over the internet, was analysed to find the average percentage of enterprise internet traffic flowing to cloud service providers.

Enterprise internet traffic consist of the consumption and exchange of data to achieve operational activities. For example, some enterprise internet traffic might consume information about research publications or potential markets for product development activities. Traffic also consists of data flow from the enterprise when emails and other communications are sent to suppliers and customers. Some of the enterprise internet traffic will be sent to cloud data centres and received from cloud data centres.

The major objective for this study is to geographically map and estimate the *volume* of enterprise data flowing to cloud and edge data centres from enterprises buying cloud computing services used over the internet.

3.1.5.1 Router management and monitoring information

Our study has developed an innovative new approach by collecting information from internet routers.[193] The stated major objective 'an estimation of the proportion of enterprise internet traffic that flowed to cloud and edge data centres from enterprises buying cloud computing services used over the internet' has not previously been a topic of study. No insights have been found in other studies; and so our study brings new insight into discussion.

Internet traffic monitoring software is included in many enterprise routers. The software enables users to check connectivity, examine traffic levels and plan future capacity, examine network outages, and understand security issues. In this study router information was used to find the volumes of data flowing to cloud services providers.

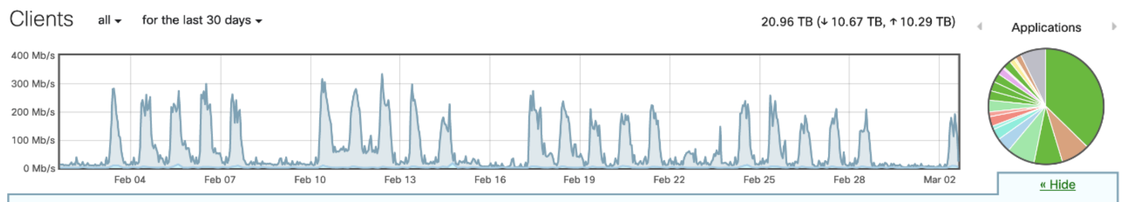
The figure below provides a screenshot of data that can be obtained from a Cisco router. Access to the IP addresses of all websites contacted from a router is valuable for examining internet traffic destinations and origins. Through IP look-up tables, it is generally possible to obtain the:

- latitude and longitude of the IP address;
- country of the IP address (generally with 99 per cent accuracy);[194]
- region/state of the IP address (90 per cent accuracy);
- city of the IP address (81 per cent accuracy).

As Figure 17 shows, the names and applications for many IP addresses are provided by a router monitoring software. For example, items 8 (SharePoint, 885.9 GB), 13 (DropBox, 523.5 GB) and 17 (Windows Office 365, 885.9 GB) are all cloud services. Router information therefore provides a useful method of monitoring the amount of internet traffic flowing through a router to a cloud data or edge centre. Thus, router management provides an innovative method for examining enterprise data flowing to cloud and edge data centres.

193 As its name implies, a router 'routes' traffic between the device in an enterprise or household and the internet.
194 https://whatismyipaddress.com/ip-lookup.

Figure 17: An example of management information provided by an enterprise router



Applications details

#	Description	Group	Usage	% Usage	Group usage	Group % usage
1	Miscellaneous secure web	–	7.92 TB	37.3%	7.92 TB	37.3%
2	UDP	–	1.70 TB	8.0%	1.70 TB	8.0%
3	Miscellaneous secure web - 212.250.159.189	–	1.68 TB	7.9%	1.68 TB	7.9%
4	Miscellaneous web	–	1.67 TB	7.9%	1.67 TB	7.9%
5	Non-web TCP	–	887.61 GB	4.1%	887.61 GB	4.1%
6	CDNs	–	714.47 GB	3.3%	714.47 GB	3.3%
7	YouTube	Video	668.25 GB	3.1%	886.47 GB	4.1%
8	Sharepoint	Productivity	592.78 GB	2.7%	885.92 GB	4.1%
9	Miscellaneous secure web - 212.250.159.139	–	583.99 GB	2.7%	583.99 GB	2.7%
10	Miscellaneous secure web - 192.168.150.2	–	580.01 GB	2.7%	580.01 GB	2.7%
11	Miscellaneous secure web - 192.168.144.248	–	543.24 GB	2.5%	543.24 GB	2.5%
12	Miscellaneous secure web - github.com	–	490.38 GB	2.3%	490.38 GB	2.3%
13	Dropbox	File sharing	430.35 GB	2.0%	523.53 GB	2.4%
14	Google HTTPS	–	360.74 GB	1.7%	360.74 GB	1.7%
15	Host-based email (POP3/IMAP/SMTP)	Email	317.19 GB	1.5%	426.18 GB	2.0%
16	microsoft.com	–	290.94 GB	1.3%	290.94 GB	1.3%
17	Windows Office365	Productivity	272.95 GB	1.3%	885.92 GB	4.1%
18	SSH	–	263.04 GB	1.2%	263.04 GB	1.2%
19	Netflix	Video	120.42 GB	0.6%	886.47 GB	4.1%
20	Software updates	Software & anti-virus updates	96.68 GB	0.4%	113.29 GB	0.5%
21	Spotify	Music	95.80 GB	0.4%	178.30 GB	0.8%
22	Facebook	Social web	95.74 GB	0.4%	207.35 GB	1.0%
23	iTunes	Music	76.03 GB	0.3%	178.30 GB	0.8%

The information provided in Figure 17 was obtained in February 2020 from the Dock Innovation Centre in Leicester, UK; it is comprised of 55 small businesses.[195] Dock uses Cisco routers which provide details (in aggregate anonymously) of applications and websites used. The Centre recorded 7.90 TB of internet traffic in February 2020. 1.29 TB of this internet traffic was enterprise data flowing to cloud data and edge centres.[196] This means that 17.2 per cent of internet traffic flowed to cloud and edge data centres.

The study team contacted over 200 innovation centres and cluster workplaces between September and December 2020 in the 31 countries being examined to seek similar router information. The response rate was poor. Many replied that due to lockdowns they and many of their tenants were working from home and results would not be representative of the pre-Covid-19 situation examined previously. Fourteen responses were received and seven had Cisco routers[197]. During discussions respondents again stressed how working conditions were very different from normal. Despite assurances of anonymity and technical support to navigate the router dashboards, only three further responses were received. This means that **the average level of internet traffic flowing to cloud and edge data centres was recorded at 17.6 per cent.**

In addition to router analysis, the micro-level analysis questionnaire also included questions about yearly data flow and the amount of such to cloud and edge data centres[198]. For the

195 Dock Innovation Centre in Leicester, UK (https://www.dockleicester.co.uk/).

196 Relevant cloud apps listed in router output include SharePoint, DropBox, Windows Office 365, Microsoft OneDrive, iCloud, Google drive etc.)

197 The remaining seven used a variety of other routers.

198 See section 3.2 which describes an interview survey with enterprises. This provided 61 complete and usable replies, prom the survey of 127 enterprises.

61 enterprises providing information about these questions, the average level of internet traffic flowing to cloud and edge data centres was 16.6 per cent.

The overall average across the two methods was 17.2 per cent; this figure was therefore used in analysis. We believe that in the future, router analysis should be pursued. With time and investment, it would be possible to create remote access to sufficient routers in order to offer considerable insights into internet traffic and data flowing to cloud service providers on a regular basis.

Applying the figure of 17.2 per cent of data flowing to cloud among enterprises utilising cloud data services, it is possible to forecast data flows between 2018 and 2030. The methodology is relatively simple. Belgium will be used as an example again to describe the methodology.

Stage 1 analysis estimated the number of people working in enterprises buying cloud services used over the internet; this was 2,896,000 people in Belgium in 2020.

In Belgium, in 2020, the average amount of data used by Belgian employees was 45.5 GB/month (see Table 11 in the previous section). This figure was used as a multiplicand with the 2,896,000 people working in enterprises buying cloud services used over the internet to estimate the amount of data that would be used by these enterprises. Analysis estimated the figure 142,800 TB/month of data (1.63 EX/year)[199].

As noted above, we estimate that 17.2 per cent of this data will flow to cloud and edge data centres - this is 24,560 TB/month (0.28 EX/year) in Belgium[200].

3.1.5.2 Data flowing to main cloud and edge data centres

Section 2 presents the results for this analysis for the 31 countries analysed between 2016 and 2030. For that year range, Eurostat statistics and ITU data is used to underpin the analysis. Linear extrapolation techniques are used to forecast the constituent elements of the analysis to 2030.[201]

Section 2 shows that in 2020 the volume of data flowing to cloud and edge servers from EU27 Member State enterprises buying internet cloud services was 504,760 TB/month (5.78 EX/year). The figure for EFTA countries in 2020 is 51,210 TB/month (0.59 EX/year). In the UK the estimate is 152,810 TB/month (1.75 EX/year).

EU27 Member State enterprise data flowing to cloud and edge data centres is forecast to increase 15-fold from the 2020 level to 7,669,830 TB/month (87.8 EX/year) in 2030. From 2020 to 2025, data flowing to cloud and edge data centres is forecast to increase 4-fold to 2,201,000 TB/month (25.2 EX/year). Likewise, from 2020 to 2030 for EFTA, a 12-fold increase is predicted, 616,680 TB/month (7.1 EX/year) and from 2020 to 2025 a 4-fold increase is predicted, 193,390 TB/month (2.2 EX/year). The UK is expected to increase to 2,291,070 TB/month (26.2 EX/year) in 2030 and to 630,940 TB/month (7.2 EX/year) in 2025. The faster growth rate in EU27 Member States is due to its relatively low starting position in terms of the number of enterprises buying cloud computing services used over

199 2,896,000 x 45.5 GB/month =142,800 TB/month

200 17.2 per cent of 142,800 TB/month = 24,560 TB/month

201 Simple linear extrapolation for some components of the model provided unrealistic elements, for example percentages over 100 per cent. To prevent these errors ‘s-shaped’ curve components were introduced in extrapolation to prevent overly large or small percentage values arising. Further details of the methodology are provided in the methodological annex.

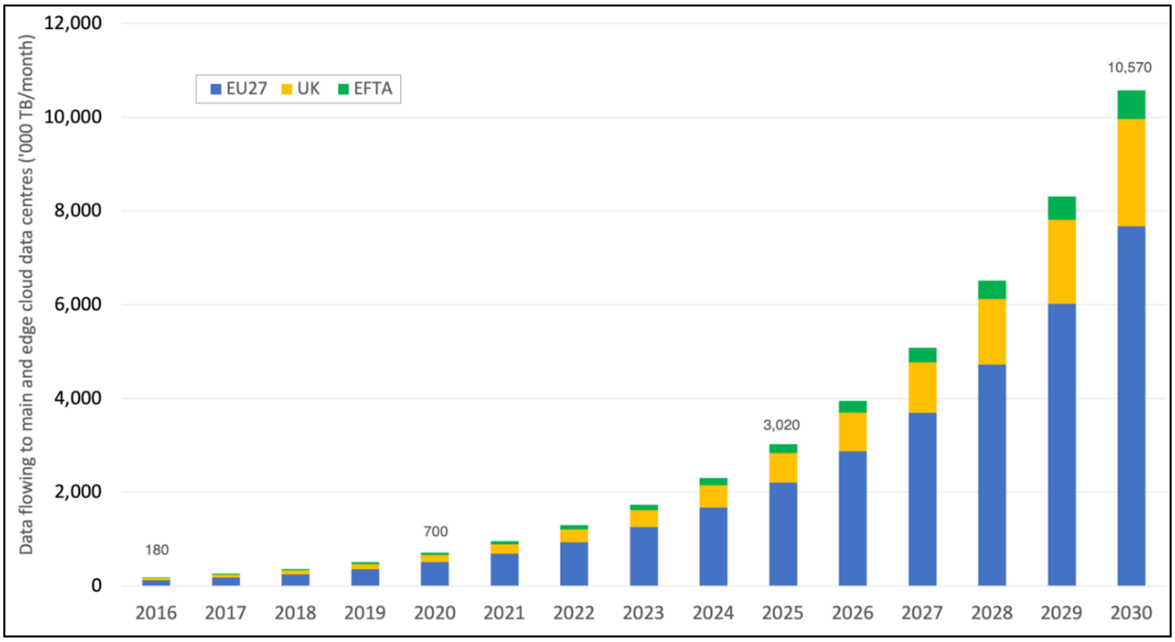
the internet. Uptake will increase among enterprises and data volumes will increase among those that are connected.

Table 12: Data flows to cloud and edge servers (TB/month) 2016 to 2030

GEO (Labels)	2016	2017	2018	2019	2020	2025	2030
EU27	122,527	176,587	245,471	352,301	504,763	2,201,058	7,669,835
EFTA	14,049	20,075	26,976	37,382	51,211	193,390	616,687
UK	44,652	62,084	86,398	114,204	152,815	630,940	2,291,077
Total	181,228	258,746	358,845	503,887	708,790	3,025,388	10,577,600
Belgium	6,677	9,900	12,949	17,755	24,567	94,965	307,673
Bulgaria	543	920	1,287	1,864	2,704	12,447	53,605
Czechia	2,075	3,246	4,682	6,613	9,290	44,533	164,557
Denmark	6,229	8,884	12,143	16,741	23,023	89,898	281,185
Germany	20,383	29,448	41,634	61,929	93,474	424,732	1,583,855
Estonia	555	838	1,210	1,979	3,061	14,383	46,130
Ireland	2,873	3,903	5,238	7,476	10,595	43,096	148,570
Greece	856	1,276	1,946	2,522	3,296	15,916	65,304
Spain	12,033	16,810	22,388	31,356	43,096	186,759	682,122
France	18,983	27,514	39,204	54,189	75,039	335,685	1,180,543
Croatia	633	1,020	1,395	1,953	2,667	13,121	48,180
Italy	10,689	15,257	21,905	37,261	58,974	286,944	931,104
Cyprus	137	232	323	464	683	3,343	12,864
Latvia	281	453	646	975	1,451	7,250	30,524
Lithuania	774	1,309	1,799	2,563	3,620	16,513	59,092
Luxembourg	365	474	693	996	1,403	5,410	20,133
Hungary	1,458	2,252	3,181	4,515	6,350	29,050	100,920
Malta	211	306	458	661	945	4,639	15,851
Netherlands	11,896	17,030	24,724	31,934	41,459	165,765	550,383
Austria	2,168	3,415	4,650	7,035	10,388	47,586	170,350
Poland	3,508	5,283	7,294	11,888	18,470	86,978	298,755
Portugal	2,202	3,564	4,521	6,105	8,964	36,112	123,077
Romania	879	1,396	1,998	3,252	5,073	28,410	127,238
Slovenia	671	869	1,268	1,840	2,611	10,994	39,412
Slovakia	698	998	1,435	2,032	2,852	13,010	48,445
Finland	5,571	7,583	9,739	13,010	17,469	62,462	196,263
Sweden	9,179	12,405	16,760	23,392	33,238	121,056	383,699
Iceland	359	499	689	943	1,294	4,869	15,566
Norway	4,764	6,885	9,006	12,452	17,067	63,655	198,783
Switzerland	8,927	12,691	17,281	23,986	32,851	124,866	402,338
United Kingdom	44,652	62,084	86,398	114,204	152,815	630,940	2,291,077

Figure 18 shows a graphical representation of the numbers presented in Table 12.

Figure 18: Enterprise data flowing to cloud and edge data centres in TB/month 2016 to 2030



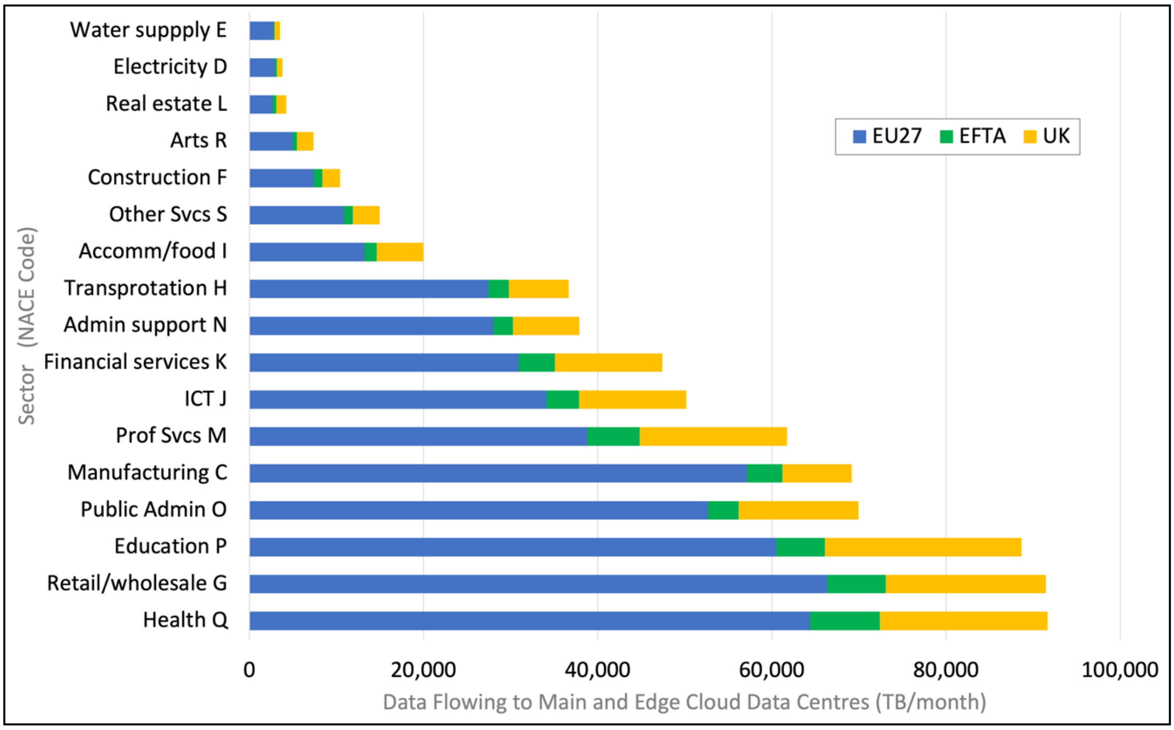
Aggregate data flows to cloud and edge data centres can be disaggregated by sector and enterprise size (employees). These data flows will be influenced by the number of employees in a sector or enterprise of particular size and the proportion of enterprises that buy cloud services used over the internet.[202]

Figure 19 provides an overview of the proportion of data flowing to cloud and edge data centres for NACE sectors C to S. Largest data flows to cloud for the 31 countries studied **come from the Health sector (NACE Q, 12.9 per cent of all flows, 91,600 TB/month, 1.05 EX/year); Retail and Wholesale (NACE G, 12.9 per cent, 91,400 TB/month; 1.05 EX/year); Education (12.5 per cent, 88,600 TB/month; 1.01 EX/year) and Public administration (9.9 per cent, 69,900 TB/month; 0.8 EX/year).**

Conversely, the smallest data flows arise from Water Supply (NACE E; 0.5 per cent, 3,500 TB/month), Electricity and Gas (NACE D; 0.5 per cent, 3,800 TB/month; 0.04 EX/year) and Real Estate (NACE L; 0.6 per cent, 4,200 TB/month; 0.05 EX/year).

202 For example, Real Estate (1.66 million employees in EU27 Member States in 2020 [Labour Force Survey data]) has a slightly smaller size of workforce to Water Supply (1.68 million employees)). Due to a higher proportion of enterprises buying cloud services used over the internet, Real Estate (43 per cent of enterprises buy cloud services; 4,210 TB/month) has more data flowing to cloud data servers in 2020 than Water Supply (30 per cent of enterprises buy cloud services; 3,506 TB/month).

Figure 19: Data flow to cloud and edge data centres by sectors in 2020



Annex 3 provides estimates of data flowing to cloud and edge data centres for NACE sectors C to S from 2016, with forecasts to 2030.

Section 2 examined the proportion of enterprises in different employee size bands. This highlights that in EU27 Member States most enterprises are small - 96.9 per cent of enterprises have fewer than 20 employees.[203] In 2020, 34 per cent of enterprises of this size purchased cloud computing services used over the internet.

Enterprises with '250 or more' employees comprise only 0.2 per cent of enterprises but they employ 35.1 per cent of the workforce. In 2020, 66 per cent of enterprises with 'over 250' employees purchased cloud computing services used over the internet. **This is almost twice the proportion of cloud service utilisation of smaller enterprises.**

It is therefore not surprising that the largest amounts of data flow come from enterprises with '250 or more employees'. In EU27 Member States in 2020, enterprises with '250 or more employees' created 61 per cent of all data flows to cloud (309,500 TB/month [3.54 EX/year] of total EU27 flows of 504,700 TB/month in 2020 [5.78 EX/year]).[204]

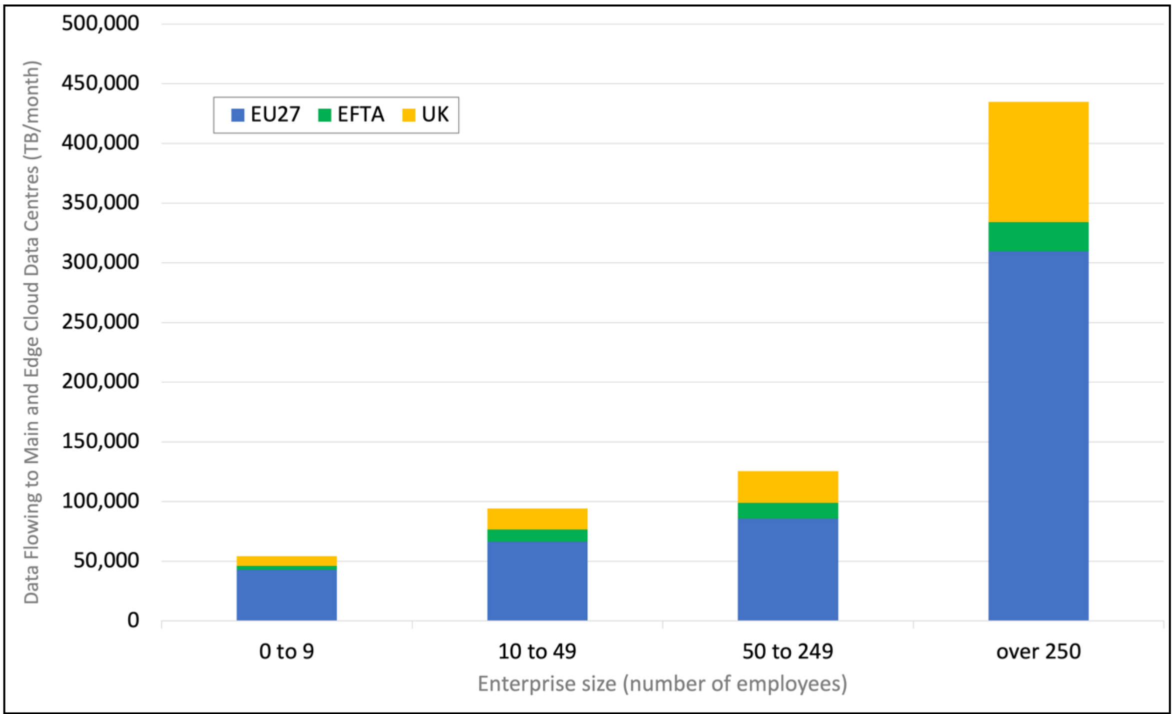
In EFTA countries in 2020, enterprises with '250 or more employees' created 48 per cent of all data flows to cloud (24,700 TB/month [0.023 EX/year] of total EU27 flows of 51,200 TB/month in 2020 [0.048 EX/year]).

In the UK in 2020, enterprises with '250 or more employees' created 66 per cent of all data flows to cloud (100,500 TB/month [0.095 EX/year] of total EU27 flows of 152,800 TB/month in 2020 [0.145 EX/year]).

203 Similar circumstances prevail in EFTA and the UK.

204 In EFTA countries data flows from enterprises with '250 or more' employees comprise 48 per cent of flows (24,700 TB/month of 51,200 TB/month). In the UK, data flows from enterprises with 250 or more employees comprise 66 per cent of flows (100,500 TB/month of 152,800 TB/month).

Figure 20: Data flow to cloud and edge data centres by different sizes of enterprise in 2020



Annex 3 provides estimates of data flowing to cloud and edge data centres for various enterprise sizes from 2016, with forecasts to 2030.

3.1.5.3 Data flowing to cloud and edge data centres as a proportion of all enterprise data traffic

Table 13 provides an insight into the magnitude of enterprise data flowing to cloud and edge data centres. Specifically, Table 13 shows this as a proportion of total internet traffic from enterprises (in the middle column) and total internet traffic from households and enterprise for the EU27, EFTA and the UK.205 In EU27 Member States, the 5.78 EX/year (505,060 TB/month) of enterprise data flowing to cloud and edge data centres represents 7.6 per cent of the 75.7 EX/year (6,614,000 TB/month) of total internet enterprise traffic. It also shows 1.9 per cent of total internet enterprise and consumer traffic of the 305 EX/year (26,651,000 TB/month) flowing in the EU27.[206]

The higher levels of uptake of cloud computing services in EFTA countries in proportion to EU27 cause the percentage of enterprise data flowing to cloud and edge data centres for EU27 and EFTA countries in aggregate terms to be higher than when only considering EU27 Member States. The UK has higher levels of enterprises buying cloud services used

205 This is primarily an estimate for day-to-day flows from enterprises to cloud providers. It does not take account of one-off uploads and downloads when changing cloud service providers. But this consideration may be negligible since Flexera estimate that 93 per cent of enterprises have a multi-cloud strategy. Data flows associated with change in providers are therefore likely to be decreased since flows to one provider might decrease at the same rate as growth to the new provider. Flexera Cloud computing trends: 2020 state of cloud report (https://www.flexera.com/blog/industry-trends/trend-of-cloud-computing-2020)

206 The calculation uses data about enterprise data flowing to cloud and edge data centres and divides the figure by the internet traffic within the geographical location for enterprises (in the middle column) and all internet traffic (enterprises and households) in the right column.

over the internet than EU27 Member States, but lower levels than EFTA countries. Results in Table 12 are therefore between these two areas.

Table 13: Enterprise data flowing to cloud and edge data centres as a proportion of internet traffic to the three geographical areas studied in 2020

Region	Enterprise data flowing to cloud and edge data centres	Enterprise data flowing to cloud and edge data centres as a proportion of enterprise internet traffic	Enterprise data flowing to cloud and edge data centres as a proportion of (enterprise and consumer) internet traffic
EU27	5.78 EX/year, 505,060 TB/month	7.6 per cent	1.9 per cent
EFTA	0.59 EX/year, 51,210 TB/month	13.4 per cent	3.3 per cent
UK	1.75 EX/year, 152,810 TB/month	10.7 per cent	2.6 per cent
EU27, EFTA and UK	8.1 EX/year, 708.790 TB/month	8.4 per cent	2.1 per cent

Eurostat[207] statistics reveal that 64 per cent of enterprises buying cloud services use public cloud delivery mode (meaning data which are in physical data centres of private cloud providers).[208] Cisco also estimates that 63.9 per cent of cloud workloads and compute instances used public cloud delivery mode with data located in physical data centres of private cloud providers in 2018.[209] 64 per cent of cloud enterprises equates to 1.8 EX/year, (157,100 TB/month) in EU27 Member States, 0.2 EX/year (17,250 TB/month) in EFTA countries and 0.63 EX/year (55,300 TB/month) in the UK in public cloud.

Furthermore, Cisco estimates that the proportion of data going to public cloud is increasing at approximately five per cent per annum. Therefore, they estimated that 73 per cent of cloud data flows will utilise public cloud data centres by 2021. Eurostat data between 2014 to 2018 suggests stability in public cloud use at about 64 per cent.

Eurostat data states that in 2018, 46 per cent of EU27 Member States enterprises use private cloud data centres.[210] Since the total percentage of Eurostat data for public and private adds up to 110%, it could be assumed that some enterprises, probably ten per cent, use hybrid cloud delivery mode meaning both public and private cloud data centres[211]. Cisco has a lower estimate of 36.1 per cent of cloud workloads and compute instances in private cloud data centres in 2018.

Table 14 provides an insight to enterprise data flowing to cloud data centres in the countries examined in the study. The first column presents an estimate of enterprise data flowing to cloud data centres. The second column presents Eurostat data about the proportion of enterprises buying cloud computing services used over the internet which utilise *public*

207 Indicator isoc_cicce_use (https://ec.europa.eu/eurostat/web/products-datasets/-/isoc_cicce_use)

208 The Eurostat metadata description is 'cloud services delivered from shared data centres of service providers'.

209 Cisco. 2018. 'Cisco global cloud index'. p3. (https://virtualization.network/Resources/Whitepapers/0b75cf2e-0c53-4891-918e-b542a5d364c5_white-paper-c11-738085.pdf) Cisco estimate that cloud workloads and compute instances in public cloud data centres in 2016 were 58 per cent. This is estimated to increase at 4.95 per cent per annum to 73 per cent in 2021. Analysis suggests that in 2018 the proportion would be 63.9 per cent.

210 The Eurostat metadata description is 'cloud services delivered from data centres of service providers exclusively reserved for the enterprise'. Eurostat indicator isoc_cicce_use last updated 24 February 2020

211 Eurostat statistics for public and private cloud use are volatile, few clear trends are visible even for individual countries. It is possible the data is inaccurate.

cloud services in 2018.[212] The third column presents Eurostat statistics for the *private* cloud equivalent data in 2018.[213] The fourth column provides an estimate of enterprise internet traffic in each country.[214] The final column provides an estimate of the proportion of enterprise data flowing to cloud data centres as a proportion of *all* enterprise internet traffic.

212 Eurostat data is for enterprises that purchase cloud services 'delivered from shared data centres of service providers' as a percentage of 'enterprises that buy cloud services'. As noted previously, Eurostat statistics for public and private cloud total more than 100. For ease of reference, data has been adjusted so that public and private cloud services total 100 per cent. Eurostat does not provide public and private cloud data for Switzerland. Accordingly, percentages were estimated as the average of other EFTA countries.

213 Eurostat data is for enterprises that purchase cloud services 'delivered from data centres of service providers exclusively reserved for the enterprise' as a percentage of 'enterprises that buy cloud services'. As noted previously Eurostat statistics for public and private cloud total more than 100. For ease of reference data has been adjusted so that public and private cloud services total 100 per cent. Eurostat does not provide public and private cloud data for Switzerland. Accordingly, percentages were estimated as the average of other EFTA countries.

214 This estimate includes data flows within the country, within EU27 and elsewhere in the world.

Table 14: Enterprise data flowing to cloud and edge data centres disaggregated by cloud delivery modes and as a proportion of other internet traffic in 2018

	Enterprise data flowing to cloud and edge data centres (TB/month)	Proportion of enterprise data flowing to public cloud and edge (%)[215]	Proportion of enterprise data flowing to private cloud and edge(%)	Total enterprise internet traffic (TB/month)	Enterprise data flowing to cloud and edge as proportion of enterprise internet traffic
EU27	245,471	58	42	17,914,358	1.4%
EFTA	26,976	63	37	1,032,529	2.6%
UK	86,398	72	28	3,863,116	2.2%
Belgium	12,949	48	52	581,588	2.2%
Bulgaria	1,287	61	39	253,549	0.5%
Czech Rep.	4,682	62	38	462,692	1.0%
Denmark	12,143	57	43	472,719	2.6%
Germany	41,634	58	42	3,509,653	1.2%
Estonia	1,210	77	23	87,159	1.4%
Ireland	5,238	72	28	273,604	1.9%
Greece	1,946	69	31	330,187	0.6%
Spain	22,388	60	40	1,875,119	1.2%
France	39,204	49	51	2,529,362	1.5%
Croatia	1,395	52	48	110,301	1.3%
Italy	21,905	50	50	1,969,735	1.1%
Cyprus	323	59	41	35,096	0.9%
Latvia	646	59	41	100,345	0.6%
Lithuania	1,799	73	27	153,992	1.2%
Luxembourg	693	55	45	47,988	1.4%
Hungary	3,181	58	42	346,661	0.9%
Malta	458	67	33	27,933	1.6%
Netherlands	24,724	58	42	1,073,646	2.3%
Austria	4,650	69	31	451,114	1.0%
Poland	7,294	64	36	1,075,585	0.7%
Portugal	4,521	66	34	424,731	1.1%
Romania	1,998	74	26	487,116	0.4%
Slovenia	1,268	57	43	104,324	1.2%
Slovakia	1,435	71	29	186,939	0.8%
Finland	9,739	73	27	349,139	2.8%
Sweden	16,760	61	39	594,080	2.8%
UK	86,398	72	28	3,863,116	2.2%
Iceland	689	53	47	28,650	2.4%
Norway	9,006	66	34	351,383	2.6%
Switzerland	17,281	59	41	652,496	2.6%

The first column of data in Table 14 enables the relative importance of enterprise data flowing to cloud and edge data centres from different countries. The black line (at 10 per cent) in the bar graph in Figure 21 indicates the 13 countries[216] that comprise 2 per cent or

215 As noted in previous footnotes Eurostat information does not usually total 100 per cent for public and private cloud within a country. In this table data has been revised to ensure that the sum of the percentages of public and private cloud total '100 per cent'. Some differences may arise due to rounding.

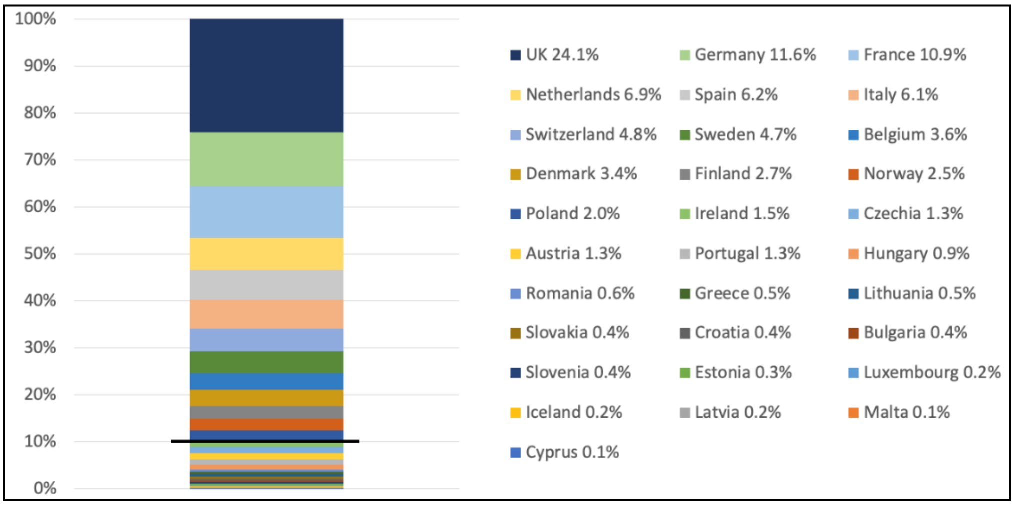
216 The 13 countries are UK, Germany, France, Netherlands, Spain, Italy, Switzerland, Sweden, Belgium, Denmark, Finland, Norway, Poland.

more of enterprise data flowing to cloud and edge data centres. These countries are above the black line. Collectively, the 13 countries comprise 89.6 per cent of all enterprise data flowing to cloud and edge data centres.[217]

The United Kingdom with 86,398 TB/month of enterprise data flowing to cloud and edge data centres has the highest levels of data flow. Indeed, UK data flows represent 24.1 per cent of all enterprise data flowing to cloud and edge data centres amongst the 31 countries included in the study. Germany with 41,634 TB/month of enterprise data flowing to cloud and edge data centres has the second highest levels of data flow. German enterprise data flows represent 11.6 per cent of all enterprise data in the 31 countries analysed.

The leading EFTA country is Switzerland which has 41,634 TB/month of enterprise data flowing to cloud and edge data centres. Swiss enterprise data flows represent 4.8 per cent of all enterprise data in the 31 countries analysed

Figure 21: Total enterprise data flowing to cloud and edge data centres for each country in 2020 as a percentage of total enterprise data flowing to cloud and edge data centres for all countries included in the analysis



Conclusion

This penultimate stage of the analysis developed an innovative method of estimating the overall volume of enterprise data flowing to cloud infrastructures. Here we can see how each stage set the grounding for the next. Analysis was built upon information provided in Stage 2 about the standardised estimate of the number of employees in enterprises buying cloud services used over the internet in 2018. Standardisation was important because it provided Stage 4 with a more robust basis for calculating enterprise data flowing to cloud in this penultimate stage. Stage 2 created estimates for the number of employees in enterprises buying cloud services used over the internet. These estimates were central to analysis in this stage

Stage 3 estimated enterprise internet traffic generated by enterprises in each country provided the basis for analysis to estimate average internet traffic (calculated as GB/month) per employee.

217 Leading EU Member State enterprise data flowing to cloud and edge data centres as a proportion of EU27 flows reveal that Germany has 17.0%, France 16.0%, Netherlands 10.1%, Spain 9.1%, Italy 8.9%, and Sweden 6.8%. In total these six Member States have 67.9 per cent of EU27 enterprise data flowing to cloud and edge data centres.

Insights, from router management information and micro-level questionnaires, about the proportion of enterprise internet traffic flowing to cloud and edge data centres, could then be applied to stag 3 output. The average figure of 17.2 per cent of enterprise internet traffic flowing to cloud and edge data centres (from this stage) was applied to the estimate average internet traffic per employee (from stage 3) for the number of employees in enterprises buying cloud services used over the internet (from stage 2).

The average (17.2 per cent) was then applied to information about enterprise internet traffic from 2018 to 2030 generated. It was then possible to estimate and forecast the amount of enterprise data flowing to cloud and edge data centres from enterprises buying cloud computing services used over the internet from 2018 to 2030 in Stage 4.

3.1.6 Stage Five: Enterprise data flowing to cloud data centres and edge infrastructures

This final stage of the methodology examines the geographical flows of enterprise data flowing to cloud and edge data centres by enterprises buying cloud computing services used over the internet.

The enterprise cloud flow geography utilises information from stage 4 about the amount of enterprise data flowing to cloud and edge data centres by enterprises buying cloud computing services used over the internet in each of the 31 countries analysed.

Analysis makes the assumption that the amount of enterprise data flowing to cloud and edge data centres will be proportional to the market share of the main cloud service providers. In each country estimates were made of data flowing to cloud and edge data centres for the ten main cloud services providers examined in section 2.6. This analysis estimated enterprise data generated by a country.

The locations of main cloud and edge data centres were identified in section 2.6. For each of the 31 countries, analysis identified the nearest main cloud and edge data centre for the ten main cloud services providers. Analysis made the assumption that 80 per cent of enterprise data flowed to main cloud data centres and 20 per cent flowed to edge data centres. On the basis of these assumptions, it was then possible to model geographical cloud data flows and examine the volume of enterprise data flowing to different countries. This analysis estimated enterprise data flowing to cloud and edge data centres in countries that possessed cloud infrastructure.

Figure 22 below provides an overview of the locations of 99 cloud and edge data centres. The 26 main *cloud* data centres were located in 11[218] of the 31 countries examined in the study. The 73 *edge* data centres were located in 19[219] of the countries studied. Lastly, 12[220] of the countries studied did not have main cloud or edge data centres.

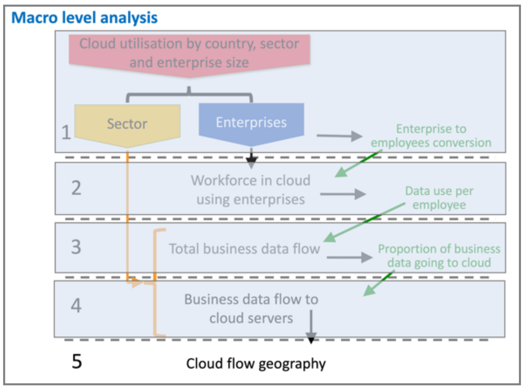


Figure 22: Locations of the main cloud and edge data centres in EU27 Member States, EFTA and the UK

	Cloud Provider 1	Cloud Provider 2	Cloud Provider 3	Cloud Provider 4	Cloud Provider 5	Cloud Provider 6	Cloud Provider 7	Cloud Provider 8	Cloud Provider 9	Cloud Provider 10
AT		E	E		E	E			E	E
BE		E	M							
CZ		E								
CY		E	E		E	E				
DE	M	M + 4E	4E	M	3E	M		M	M	E
DK		E			E					
ES		E	E							
FI		E	M							
FR		M + 2E	2E	M	2E		M + 2E			M
IE		M + E	M + E		M + E					
IT		E	E	M	E					
MT		E	E							
NL		E	M + E	M	M + E	E				
PL		E					E			
SE		M + E	E		E					
CH		E	E		2E	M				
IS	E	E	E		E					
NO		E		M	2E					
UK	E	M + 4E	E	M	E	M			M	E

Interviews with some of the leading cloud service providers confirmed that routing for enterprise data flowing to cloud and edge data centres is undertaken automatically to the 'best available data centre'. This automated and complex process of directing huge volumes of internet traffic will adhere to 'load balancing' principles. Additionally, the process probably aims to minimise traffic costs for cloud service providers.

Analysis suggests[221] that cloud service requests from an enterprise will be directed to the nearest data centre, but this is unlikely to be correct. The study team has undertaken data flow speed testing (ping times) to 16 websites[222] that were cached on the cloud and edge

218 Belgium, Switzerland, Germany, Finland, France, Ireland, Italy, Netherlands, Norway, Sweden, UK.

219 The 11 countries with main cloud data centres plus Austria, Czech Republic, Cyprus, Denmark, Spain, Iceland, Malta, Poland.

220 Bulgaria, Estonia, Greece, Croatia, Latvia, Lithuania, Luxemburg, Hungary, Portugal, Romania, Slovenia, Slovakia.

221 Cloudacademy. 2019. Disadvantages of cloud computing. https://cloudacademy.com/blog/disadvantages-of-cloud-computing/

222 Web sites were selected to provide coverage of at least four of the main cloud service providers where they were hosted. Sites included Adobe.com cached at Microsoft Azure cloud data centres, Amazon.com AWS, Apple.com AWS, Bing.com Microsoft Azure, Ebay.co.uk Google Cloud, Facebook.com AWS, Google.com Google Cloud, Instagram.com AWS, Microsoft.com Microsoft Azure, Office.com Microsoft Azure, Reddit.com AWS, Tmall.com Alibaba Cloud, Twitch.tv AWS, Twitter.com Google Cloud, Yahoo.com Microsoft Azure, Youtube.com Google Cloud.

data centres of the leading cloud service providers. Analysis from the 31 countries in this study showed that ping times imply relatively low levels of routing to edge data centres. Ping times indicated that most traffic was routed to the primary cloud and edge data centres.[223] The assumption is therefore made that **80 per cent of enterprise data flowing to cloud from a country is to main cloud data centres; the remaining 20 per cent flows go to edge data centres.**

Assumptions also had to be made about the distribution of enterprise data to the main cloud and edge data centres of the main cloud service providers.

One could assume the distribution of cloud data flows in each country is undertaken pro-rata with the market share of the main cloud providers in a given country. Table 14 highlights that there is no agreement in market research reports about the global market shares of cloud service providers in Q1 2021. We were unable to find market shares using ratio or interval data[224] for EU27, EFTA, the UK, or the 31 individual countries investigated in this study.

Table 15: Global market share for leading cloud service providers Q1 2021

Cloud provider	Canalys market share (%)	Statista market share (%)	Market share adopted in this analysis
Cloud Provider 2	32%	32%	33%
Cloud Provider 5	19%	20%	17%
Cloud Provider 3	7%	9%	8%
Cloud Provider 1	5%	6%	5%
Cloud Provider 4		5%	5%
Cloud Provider 6		2%	3%
Cloud Provider 7			1%
Cloud Provider 8			1%
Others	37%	26%	27%

Source: Canalys, 2021. Statista, 2021 [225]:

The final column Table 15 provides details of the percentages of data flowing to different cloud service providers that were used in the analysis for this study.[226]

Once again, Belgium is used to demonstrate how analysis was undertaken. Previous analysis has highlighted that Belgium had 24,560 TB/month (0.28 EX/year) flowing to cloud

223 Analysis shows a strong statistical relationship (R2 = 0.75) between distance from cloud data centre locations and access speed. Distance of a location from cloud services explains 52.9 per cent of variance in connection speeds. Tech4i2. 2020. 'Covid19 impact on internet use' (https://admin.tech4i2.com/storage/thoughts/23/files/Tech4i2-Covid19-and-internet-Europe-and-N-America-4pdf.pdf)

224 Ratio and interval data enable mathematical analysis because the relative difference between data points is known. Ratio data is provided on a scale that not only provides the order of data points, it also makes the difference between data points known along with information about the value of true zero. Interval data is a numerical scale where the order of data points is known as well as the difference between data points (percentages are a good example of interval data)

225 Canalys quoted at https://www.parkmycloud.com/blog/aws-vs-azure-vs-google-cloud-market-share. Statista https://www.statista.com/chart/18819/worldwide-market-share-of-leading-cloud-infrastructure-service-providers

226 The market share adopted for the study in the right hand column was based on global market share information. In general an average value was taken. Cloud Provider 7 and Cloud Provider 8 are relatively large providers in their own countries (France and Germany respectively) but both are small in European terms and were thus allocated one per cent market shares. The total market share from known sources in table 13 is 73 per cent. Thus 27 per cent is unknown. These unknown providers are not listed in the two sources, thus we cannot name them. It is highly probable they are local providers that have less than one per cent market share. Since we do not know the names of these providers we cannot know the locations of their main and edge data servers.

servers. The second column of Table 15 shows the proportion of data flowing to each of the service providers according to the market percentage in the left column.

Table 16: Share of Belgium enterprise data flowing to different cloud service providers in 2020 and the locations of the nearest provider main cloud and edge facilities

Cloud service provider	Market share	% of total country cloud data flow (24,560 TB/month)	Nearest main cloud data centre for 80% of flow	Nearest edge facility for 20% of flow
Cloud Provider 2	33%	8,100	Germany	Netherlands
Cloud Provider 5	17%	4,170	Netherlands	Netherlands
Cloud Provider 3	8%	1,960	Belgium	Netherlands
Cloud Provider 1	5%	1,220	Germany	UK
Cloud Provider 4	5%	1,220	Germany	Germany
Cloud Provider 6	3%	737	Germany	Netherlands
Cloud Provider 7	1%	245	France	France
Cloud Provider 8	1%	245	Germany	Germany
Others	27%	6,630	-	-

Thus, 33 per cent of total Belgium enterprise cloud data (24,560 TB/month) is expected to flow to Cloud Provider 2 (CP2); this equates to 8,100 TB/month. This method is replicated for the proportion of enterprise data flowing to the other cloud service providers.

As noted previously, 80 per cent of the amount of data flowing to Provider 1 (6,480 TB/month, 0.07 EX/year) is allocated to the nearest Provider 1 main data server. As shown in the table, Germany is the nearest main Provider 1 data centre to Belgium. See column three for the nearest main data centres for each of the cloud service providers. The rightmost column lists the locations of the nearest edge server for each of the providers. The nearest edge provider to Belgium is in the Netherlands. Thus, the main German cloud data centre of Provider 1 will receive approximately 6,480 TB/month (80 per cent of the Belgian enterprise data flowing to Provider 1 - 8,100 TB/month multiplied by 80 per cent). The remaining 20 per cent (1,620 TB/month, 0.02 EX/year) is allocated to Provider 1 edge servers in the Netherlands.

This process is replicated for the other seven cloud services providers for Belgium. Then, the overall approach is then replicated across all 31 countries for the five years from 2016 to 2021.[227]

Having replicated the methodology explained in the Belgian example for all 31 countries included in the analysis, it was possible to examine the amount of data flowing from each country studied to main and edge data centres. This data and an overview of results was provided earlier in the study. It is replicated here so that it sits alongside (and can be more easily compared with) information about enterprise data flowing into the cloud and edge data centres in the 31 countries studied.

227 Due to the huge growth in edge data centres, analysis beyond 2021 would provide erroneous results.

Table 17: Enterprise data flowing to cloud and edge data servers from the 31 countries examined in the study 2016 to 2021 (TB/month)

GEO (Labels)	2016	2017	2018	2019	2020	2021
EU27	122,527	176,587	245,471	352,301	504,763	681,754
EFTA	14,049	20,075	26,976	37,381	51,212	68,727
UK	44,652	62,084	86,398	114,204	152,815	205,393
Total	181,228	258,746	358,845	503,887	708,790	955,875

Belgium	6,677	9,900	12,949	17,755	24,567	32,836
Bulgaria	543	920	1,287	1,864	2,704	3,683
Czechia	2,075	3,246	4,682	6,613	9,290	12,933
Denmark	6,229	8,884	12,143	16,741	23,023	31,262
Germany	20,383	29,448	41,634	61,929	93,474	122,275
Estonia	555	838	1,210	1,979	3,061	4,210
Ireland	2,873	3,903	5,238	7,476	10,595	13,866
Greece	856	1,276	1,946	2,522	3,296	4,751
Spain	12,033	16,810	22,388	31,356	43,096	58,165
France	18,983	27,514	39,204	54,189	75,039	103,965
Croatia	633	1,020	1,395	1,953	2,667	3,813
Italy	10,689	15,257	21,905	37,261	58,974	81,518
Cyprus	137	232	323	464	683	953
Latvia	281	453	646	975	1,451	2,015
Lithuania	774	1,309	1,799	2,563	3,620	5,008
Luxembourg	365	474	693	996	1,403	1,751
Hungary	1,458	2,252	3,181	4,515	6,350	8,770
Malta	211	306	458	661	945	1,338
Netherlands	11,896	17,030	24,724	31,934	41,459	57,630
Austria	2,168	3,415	4,650	7,035	10,388	14,185
Poland	3,508	5,283	7,294	11,888	18,470	24,704
Portugal	2,202	3,564	4,521	6,105	8,964	11,438
Romania	879	1,396	1,998	3,252	5,073	7,236
Slovenia	671	869	1,268	1,840	2,611	3,501
Slovakia	698	998	1,435	2,032	2,852	3,894
Finland	5,571	7,583	9,739	13,010	17,469	22,615
Sweden	9,179	12,405	16,760	23,392	33,238	43,439
Iceland	359	499	689	943	1,294	1,727
Norway	4,764	6,885	9,006	12,452	17,067	22,917
Switzerland	8,927	12,691	17,281	23,986	32,851	44,083
United Kingdom	44,652	62,084	86,398	114,204	152,815	205,393
Unknown	48,932	69,861	96,888	136,049	191,373	258,086

Table 18: Enterprise data flowing to cloud and edge data servers in the 31 countries examined in the study 2016 to 2021 (TB/month)

	2016	2017	2018	2019	2020	2021
EU27	109,214	156,579	217,237	307,682	435,847	588,108
EFTA	4,059	5,789	7,860	11,082	15,506	20,883
UK	19,023	26,516	36,860	49,073	66,064	88,798
Total	132,296	188,885	261,957	367,837	517,417	697,789
Belgium	5,524	7,986	11,136	16,231	23,486	31,790
Bulgaria	0	0	0	0	0	0
Czechia	183	280	404	571	801	1,111
Denmark	623	888	1,214	1,674	2,302	3,126
Germany	35,943	51,997	72,720	105,202	151,968	180,747
Estonia	0	0	0	0	0	0
Ireland	16,234	22,629	31,153	41,779	56,576	75,730
Greece	0	0	0	0	0	0
Spain	1,167	1,671	2,207	3,072	4,269	5,707
France	14,248	20,553	28,535	39,586	55,001	75,152
Croatia	0	0	0	0	0	0
Italy	3,369	4,857	6,836	10,689	16,051	45,827
Cyprus	25	42	59	83	119	167
Latvia	0	0	0	0	0	0
Lithuania	0	0	0	0	0	0
Luxembourg	0	0	0	0	0	0
Hungary	0	0	0	0	0	0
Malta	17	25	38	54	77	110
Netherlands	21,654	31,242	43,643	61,609	86,963	117,891
Austria	1,014	1,547	2,170	3,229	4,701	6,495
Poland	362	550	756	1,176	1,774	2,386
Portugal	0	0	0	0	0	0
Romania	0	0	0	0	0	0
Slovenia	0	0	0	0	0	0
Slovakia	0	0	0	0	0	0
Finland	1,775	2,472	3,271	4,533	6,309	8,316
Sweden	7,077	9,840	13,096	18,193	25,448	33,552
Iceland	45	63	87	119	163	218
Norway	1,551	2,187	2,893	4,011	5,567	7,388
Switzerland	2,464	3,539	4,880	6,952	9,776	13,278
United Kingdom	19,023	26,516	36,860	49,073	66,064	88,798
Unknown	48,932	69,861	96,888	136,049	191,373	258,086

Table 18 provides an overview of the amount (TB/month) of enterprise data flowing to cloud service providers in different countries 2016 to 2021[228]. The unknown data flow to cloud and edge data centres at the foot of the table is estimated by using the estimate of the proportion (27 per cent) of data flowing to unknown providers.

Table 18 records values of zero for 12 countries because they are not thought to possess cloud or edge data centres for the eight main cloud service providers listed in Figure 10. The 27 per cent of data flowing to 'other' unknown cloud service providers (whose cloud

228 Analysis focuses on 2016 to 2021 because information about cloud infrastructure in these years is known. Many commentators predict considerable growth in cloud edge data centres in the next few years. Extrapolation to create forecasts in the future, when the distribution of infrastructure will change considerably, would create spurious results.

infrastructure locations is unknown) are included so that the magnitude of data flowing to smaller providers (that collectively appear to have 27 per cent market share) can be appreciated.

By far the largest amount of enterprise data flows are served by cloud and edge data centres in Germany. In 2020, Germany received 151,968 TB/month (1.74 EX/year) of cloud data flows from other countries. This represents 30.7 per cent of cloud data flows to EU27 Member States and 25.9 per cent of cloud data flows across the 31 countries studied.

The EU27 Member State receiving the second highest inflow of data served by its cloud and edge data centres in 2020 was the Netherlands (86,963 TB/month, 1 EX/year).

Both these countries (Germany and the Netherlands) received more enterprise data flowing to their cloud and edge data centres in 2020 than the UK (88,798 TB/month, 1.02 EX/year). In 2020, the UK received 12.7 per cent of total data flowing to cloud and edge facilities from the 31 countries studied. EFTA countries received 3 per cent of total data flowing to cloud and edge facilities (15,506 TB/month, 0.18 EX/year).

The final table in this section examines the net enterprise data flow for the 31 countries examined in the study. This is found by subtracting data inflows to a country in Table 18 from data outflows presented in Table 18. For example, the net data flow in Belgium in 2020 is 1,080 TB/month (0.01 EX/year). Using the Belgian example, the analysis involves subtracting 23,486 TB/month data inflows to their cloud and edge data centres located in Belgium from the 24,567 TB/month. Thus, there is net data outflow of 1,080 TB/month.

Table 19 presents net enterprise data flow figures for the 31 countries examined in the study. Three countries have net enterprise data inflows to their cloud and edge data centres[229]. These three are depicted in bold in Table 19 they all have positive values. All other countries have net data outflows, minus figures in Table 19. Minus figures indicate there is more enterprise data flowing out of the country to cloud and edge data centres in other countries than data flowing into the country (if the country has cloud or edge data centres; as noted earlier 12 countries do not have this infrastructure).

229 The ‘total’ figure in the fourth row of Table 19 is the same number (TB/month) as the ‘unknown’ data flows at the foot of Table 18. The negative figure arises because, as noted earlier the locations of the cloud or edge infrastructure for the smaller cloud services providers that comprised 27 per cent of the market could not be allocated to a country.

Table 19 Net inflow or outflow of enterprise data flowing to cloud and edge data servers for the 31 countries examined in the study 2016 to 2021 (TB/month)

	2016	2017	2018	2019	2020	2021
EU27	-13,312	-20,008	-28,234	-44,619	-68,916	-93,647
EFTA	-9,990	-14,286	-19,116	-26,299	-35,706	-47,844
UK	-25,629	-35,568	-49,538	-65,131	-86,751	-116,595
Total	-48,932	-69,861	-96,888	-136,049	-191,373	-258,086
Belgium	-1,153	-1,914	-1,814	-1,524	-1,080	-1,046
Bulgaria	-543	-920	-1,287	-1,864	-2,704	-3,683
Czechia	-1,892	-2,966	-4,278	-6,042	-8,489	-11,822
Denmark	-5,606	-7,996	-10,929	-15,067	-20,720	-28,136
Germany	***15,560***	***22,549***	***31,086***	***43,273***	***58,494***	***58,472***
Estonia	-555	-838	-1,210	-1,979	-3,061	-4,210
Ireland	***13,361***	***18,726***	***25,915***	***34,304***	***45,981***	***61,864***
Greece	-856	-1,276	-1,946	-2,522	-3,296	-4,751
Spain	-10,866	-15,139	-20,182	-28,284	-38,827	-52,458
France	-4,735	-6,961	-10,669	-14,603	-20,038	-28,813
Croatia	-633	-1,020	-1,395	-1,953	-2,667	-3,813
Italy	-7,319	-10,400	-15,069	-26,572	-42,923	-35,691
Cyprus	-112	-191	-264	-381	-563	-786
Latvia	-281	-453	-646	-975	-1,451	-2,015
Lithuania	-774	-1,309	-1,799	-2,563	-3,620	-5,008
Luxembourg	-365	-474	-693	-996	-1,403	-1,751
Hungary	-1,458	-2,252	-3,181	-4,515	-6,350	-8,770
Malta	-193	-281	-421	-607	-868	-1,228
Netherlands	***9,758***	***14,213***	***18,919***	***29,675***	***45,504***	***60,261***
Austria	-1,155	-1,869	-2,480	-3,806	-5,686	-7,690
Poland	-3,146	-4,733	-6,538	-10,712	-16,696	-22,318
Portugal	-2,202	-3,564	-4,521	-6,105	-8,964	-11,438
Romania	-879	-1,396	-1,998	-3,252	-5,073	-7,236
Slovenia	-671	-869	-1,268	-1,840	-2,611	-3,501
Slovakia	-698	-998	-1,435	-2,032	-2,852	-3,894
Finland	-3,796	-5,111	-6,468	-8,477	-11,160	-14,299
Sweden	-2,102	-2,566	-3,664	-5,199	-7,790	-9,887
Iceland	-314	-436	-602	-825	-1,131	-1,510
Norway	-3,213	-4,698	-6,113	-8,441	-11,500	-15,529
Switzerland	-6,463	-9,152	-12,401	-17,034	-23,075	-30,805
United Kingdom	-25,629	-35,568	-49,538	-65,131	-86,751	-116,595

The country with the highest net data inflow to its cloud and edge data centres in 2020 was Germany. German net enterprise data inflow to its cloud and edge data centres in 2020 was 58,494 TB/month (0.67 EX/year). Germany has a main cloud data centre for Cloud Provider 2, but unlike the two countries in second and third place for net inflows (Ireland and the Netherlands) it does not have main cloud data centres for the second and third largest cloud service providers (Cloud Provider 3 and Cloud Provider 5).

Ireland has the second highest net enterprise data inflow to its cloud and edge data centres (45,981 TB/month, 0.53 EX/year). Ireland has a relatively small population (5 million in 2020) and only has 10,595 TB/month (0.12 EX/year) of enterprise data flowing to cloud and edge data centres. However, it is unique in having main cloud data centres for the three cloud service providers with the largest market share (Cloud Provider 2, Cloud Provider 3, and Cloud Provider 5, see Figure 10).

The Netherlands has the third highest net enterprise data inflow to its cloud and edge data centres in 2020 (45,504 TB/month (0.52 EX/year). The Netherlands has main cloud data centres for Cloud Provider 3 and Cloud Provider 5 (and four edge data centres). However, unlike Ireland, it does not have a main cloud data centre for Cloud Provider 2 the cloud service provider with the largest market share.

The EU27 Member State with the largest net outflow of enterprise data to cloud and edge data centres in other countries is Italy. In 2020, Italy was thought to have one main cloud data centre and three edge centres, see Figure 10. As a result, net enterprise data outflow from Italy is 42,923 TB/month (0.49 EX/year).

Of the 31 countries examined largest net enterprise data out flows were recorded by the UK. The UK has quite a lot of data infrastructure, including a main Cloud Provider 2 cloud data centre. But it does not have main cloud data centres for the second and third largest cloud service providers (Cloud Provider 3 and Cloud Provider 5). As a result the net enterprise data out flow in the UK in 2020 was 86,751 TB/month (1 EX/year).

Figure 23 provides a graphical representation of the geography for net inflow or outflow of enterprise data flowing to cloud and edge data servers for the 31 countries examined in the study in 2020.

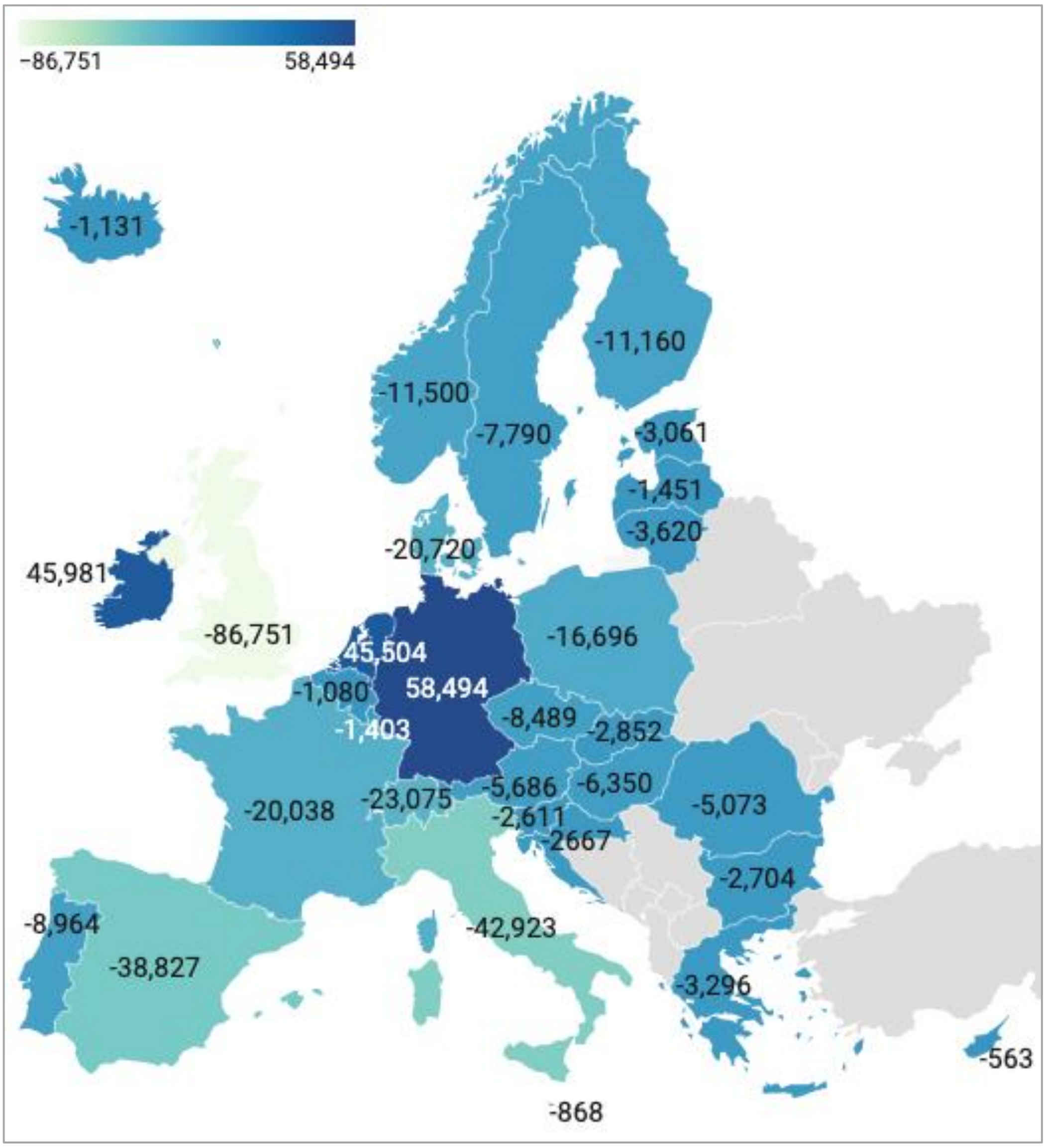


Figure 23 Data flows geography: Net inflow/outflow map for data flowing to cloud and edge data centres for the 31 countries examined in the study (TB/month) in 2020

Conclusion

This final stage of the analysis developed a further innovative method of investigating the details of enterprise data flows to cloud and edge infrastructures across the EU27, EFTA and the UK. The culmination of Stage 4 was to **estimate volume and origin of data flows to cloud and edge servers in TB/month for 2016 to 2030.**

Furthermore, in stage 5 the analysis sets the destination of cloud data flows by allocating the volume of data flows to cloud and edge service providers' infrastructures according to their market shares across the EU27MS, EFTA and the UK. Analysis then painstakingly investigated the locations of the nearest main cloud and edge date centre for each cloud service provider in all 31 countries. Then, the appropriate market share cloud data flow volumes were allocated to the relevant countries.

After completing the analysis, it was possible to examine how much data was flowing to cloud and edge data centres in each country that possessed cloud and edge infrastructure in the 31 countries examined in the study.

3.2 Collecting primary data on the dimensions of cloud data flows

In parallel to the development of the macro-level approach, described above, the micro-level approach was designed to complement the analysis to get insights on data types (persona and non-personal data). It aimed at providing more detailed information based primarily on interviews with the infrastructure networks and cloud and edge service providers and a survey of cloud data users. This approach is done at micro-level, using primary data from the interviews and the survey.

The objective of the **interviews** was to target primarily cloud service providers and infrastructure networks. The aim was to get feedback on the methodology of the study and insights on the services provided by cloud providers (SaaS, PaaS, IaaS, Edge Services, added value cloud services, such as AI, blockchain, big data analytics) in Europe as well as volume of data stored in cloud infrastructures and cloud capacity.

The objective of the **survey** was to target primarily cloud data users: enterprises from the different sectors targeted in the study and public entities (including public administrations and academia) departing from a first survey developed by the European Commission as part of the European Data Flow Monitoring Initiative[230] and published on the EU survey tool in the summer of 2019. The aim was to focus on the perspective of the users and provide insights on cloud data types that could not be captured by the macro-level approach. The survey gathered 174 responses.

The micro-level approach allowed to gather estimates on **data types:** the survey estimates that the total data stored in cloud infrastructure is 41% personal data and 59% non-personal data.

The survey provided also insights on *enabling technology* used to store data and *functional area*s, namely the most cloud intensive productive activities of enterprises. As an example, the survey findings highlighted that in 2020, the percentage between cloud and edge infrastructure used to store data was approximately 73 per cent cloud vs 27 per cent edge.

230 https://ec.europa.eu/digital-single-market/en/european-data-flow-monitoring-initiative

Moreover, the survey findings show that the most cloud intensive productive activity is research and development. concerning, corporate activities within an enterprise, data storage and back-up are the most cloud intensive activities, followed by client relationship management. However, due to the low reliability of the findings at country level, it was decided to not use the findings on functional areas and enabling technology. Concerning enabling technology findings from the survey they did not align with the other findings taken as assumption in this analysis based on the modelling ping test. Concerning the functional areas, due to low answer rate of the survey, it was decided to not include as an assumption in the modelling and only focus on the Eurostat data on low, medium and high level of cloud service. Hence, only the findings on data types were used.

Due to lack of responsiveness or knowledge of interviewees, data on the overall volume stored, cloud centre capacity and services provided could not be gathered through the interviews. Survey data, except for the dimension of data types (i.e share of personal and non-personal data store in cloud infrastructure), was not deemed reliable enough to feed into the analysis. This is why only five dimensions using mostly findings from the macro-level approach are presented in the report. Lessons learnt from this exercise are explained more in detail in Annex 5.

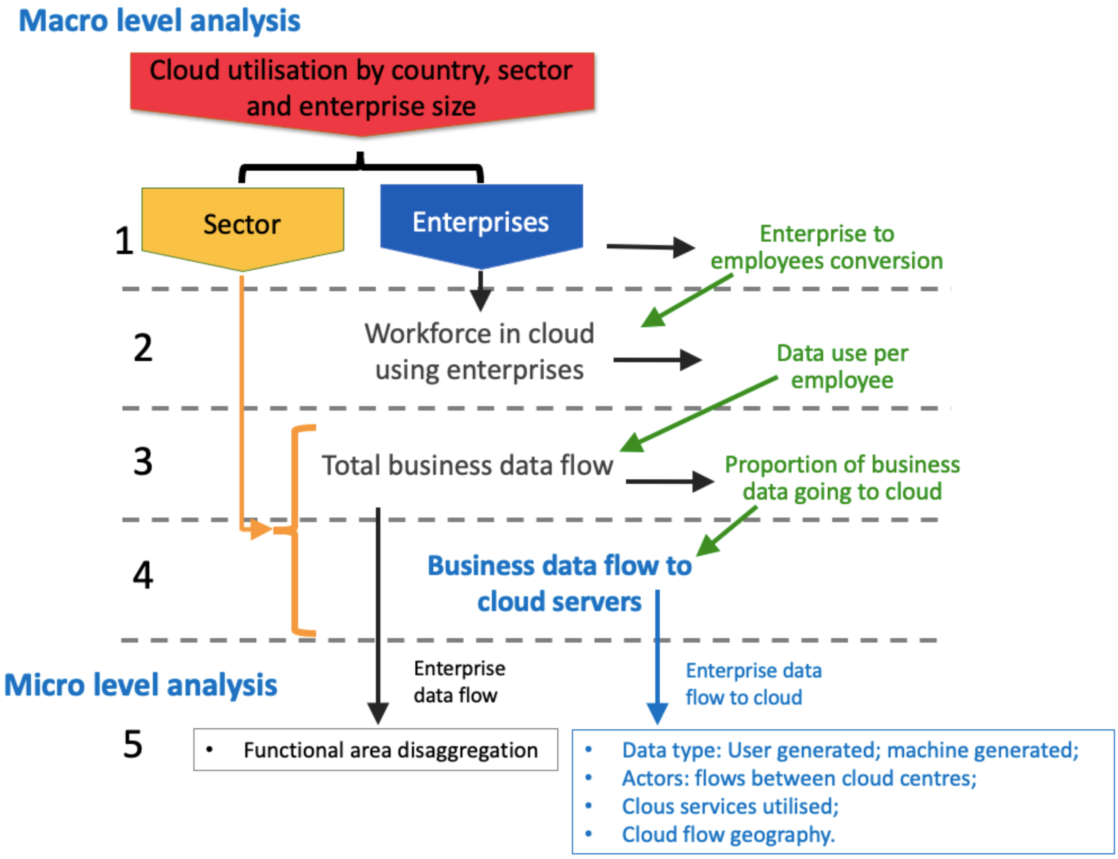
A data collection methodology encompassing a complementary macro-level and a micro-level approach was designed to collect data to measure the different dimensions of data flows and provides a typology of data flows per country on this basis.

The macro-level approach enabled the examination of data flows by estimating data flows based on four characteristics corresponding to the first four dimensions of the typology of data flows: by country, by sector, by enterprise size and nature of cloud services utilised.

By providing an estimate of the overall volume of data flows, the macro-level analysis also built a framework onto which it was possible to bolt findings from other sources, e.g. other reports or studies, the interviews and the survey conducted in the context of this study, and provide insights on other dimensions of data flows, such as geography, sector, enterprise size and cloud services. The objective of the micro-level approach was to gather data, using the interviews and the survey, on data types that could not be captured by the macro-level analysis.

This rationale is explained in Figure 24.

Figure 24: How micro-level anlaysis supports the macro-level approach in the study



Stages 1 to 4 describe how the macro-level analysis produced baseline estimates for enterprise data flows (see the black bold arrow) and enterprise flows to cloud (see the blue bold arrow). In the micro-level analysis, Stage 5, the baseline estimates can be used to provide quantitative insights to additional dimensions of data flows.

The table below shows which data collection source and approach was chosen to estimate the different dimensions of data flows.

Table 20: The study dimensions and insights provided by micro-level and macro-level analysis

Data flow dimension	Data collection source	Approach chosen	Objective
Geography	Secondary data provides information about the size of enterprises buying low, medium and high cloud services used over the internet. Enterprise size bands 0 to 9, 10 to 49, 50 to 249 and 250 or more employees	Macro-level	To map enterprise data flowing to cloud data centres from and to EU countries, EFTA and UK
Sector	Secondary data provides information about enterprises buying cloud services used over the internet in NACE sectors C to S	Macro-level	To map enterprise data flowing to cloud data centres by NACE sectors (C to S)

Data flow dimension	Data collection source	Approach chosen	Objective
Enterprise size	Secondary data provides information about enterprises buying cloud services used over the internet in 31 countries	Macro-level	To map enterprise data flowing to cloud data centres by enterprise size
Nature of cloud services utilised	Secondary data provides information on the nature of cloud service utilised (low-medium-high) by enterprises Primary data provides information on the percentage of cloud applications used	Macro-level and micro-level	To map enterprise data flowing to cloud data centres by low/medium/high To provide an estimation on the main services used
Data type	Primary data (survey) provides information on the percentage of data stored per type (personal or non-personal)	Micro-level	To provide estimation on data stored by data type (personal/non-personal)

For this exercise, the data collected in the micro-level analysis was not reliable enough to be extrapolated and back a solid estimation of data flows at country-level and for all identified dimensions of data flows. Some reliability issues were identified also in the survey. This could be attributed to the respondents' lack of knowledge about their cloud usage (generation), usage, and the details of their contracts. Another reason for the lack of reliability of the answers is the questionnaire itself. In particular, answers to open questions were particularly difficult to analyse as too diverse (see annex 5, detailed methodology for collecting primary data)

However, the usefulness of a micro-level approach based on stakeholder data collection to complement the findings of the macro-level approach predominantly using secondary sources remains valid. For example, in the study, the macro-level analysis found that 12,620 TB/day of enterprise data was flowing to cloud servers. The survey found that on average 27 per cent of survey respondents used Software as a Service (SaaS) (Dimension 4). If survey respondents were representative of all EU27 Member State enterprises, this could suggest that 3,400 TB/day of enterprise data flowing to cloud servers is associated with SaaS (12,620 x 27 per cent). A summary of survey respondents' overview is provided in annex 10. In this way, the more general macro-level insights can be further disaggregated by micro-level insights provided from survey results and provide data at country-level in future exercises. As lesson learnt, in order to make the **survey** exercise more useful, the following points could be improved:

- Increase the number of respondents (at least 2,400) and build the sampling approach with a professional survey provider that can guarantee reliable responses through a more restrictive panel.
- The sample could be extended in order to cover more exhaustively public administrations and academia.
- Limit the number of questions to a minimum, focusing on the aspects that cannot be grasped by the macro-level approach. The macro-level approach has shown success to provide reliable estimations of Dimensions 3 - Actors (Enterprise data to cloud services), Dimension 6 - Sectors and Dimension 7 - Geography. The survey

should therefore focus on the remaining dimensions, omitting those aspects that are difficult to grasp for cloud users (e.g. the location of their cloud data centres).

- Use only or mostly close-ended questions. For instance, in Q.12, give a list of 10 the most used applications and let the respondent choose between them. An open text can be added if the respondent answers ‘Other’.
- The survey target should still be cloud users, however the questions should focus gathering estimation on proportions rather than volumes of data used. It is easier for cloud users to work on proportions rather than giving estimations on the volumes of data.

4. A methodology for estimating cloud and edge data centre capacity and utilisation

The previous chapter focused on the flows of data to cloud and edge data centres. This chapter takes a closer look at the cloud and edge data. It examines the location of main cloud and edge data centres and provides estimates of cloud data capacity and the amount of capacity utilised by enterprises buying cloud services used over the internet.

4.1 Location of the cloud and edge data centres

Section 2 described why main cloud data centres (or hypercentres) are fairly easy to define. They are huge, cost millions of euros and are generally only built by the main cloud service providers. It is therefore relatively easy to find where they are located.

Conversely, edge computing is rapidly evolving, and edge data centres are becoming increasingly difficult to define and therefore identify where they are located. Edge computing is becoming more of a continuum of edge data centres, with different computing capabilities supplied by a variety of cloud data centre providers. For example, new technology such as micro data centres provide edge services with more limited computing capabilities. In addition, new providers, such as telecommunications companies, are using their data centres to provide cloud services at the edge.

There are few information sources[231] that provide location details for the main cloud data centres and edge data centres.[232] There is generally agreement between the sources about the locations of main cloud data centres. But there is far less agreement about the location of edge data centres. This is almost certainly due to the lack of a clear definition of edge data centres so far. The smaller physical size of cloud edge centres, points of presence and micro data centres has not helped observers to find them and partnership agreements with IXPs and telecommunication companies can obscure the identification and location of edge data centres.[233]

The emergence of smaller providers, often led by telecommunications businesses offering cloud services at national or regional level, also creates complexity.[234]

Finding the location of main cloud and edge data centres is made a little easier by the fact that some of the leading cloud service providers share information about the locations of their main cloud data centres and sometimes edge centres.[235] Conversely, some smaller cloud providers, only provide locations of 'data centres', without distinguishing between main cloud and edge centres.

This study has therefore analysed all available information and created a 'best estimate' of the locations of main cloud and edge data centres for the main cloud service providers

231 TeleGeography (https://global-internet-map-2021.telegeography.com). CloudScene (https://cloudscene.com/browse/on-ramps). DataCentreKnowledge (https://www.datacenterknowledge.com/cloud/telegeography-maps-world-s-cloud-data-centers)

232 Including historic information about AWS cloud data centres on WikiLeaks (https://wikileaks.org/amazon-atlas/map/)

233 Tsidulko J. 2020. VMWare, AWS now giving each other preferential treatment

234 Synergy. 2021. European cloud providers struggle to reverse market share losses (https://www.srgresearch.com/articles/european-cloud-providers-struggle-reverse-market-share-losses)

235 Cloud Provider 2, Cloud Provider 7, Cloud Provider 1, Cloud Provider 8, Cloud Provider 3, Cloud Provider 4, Cloud Provider 5, Cloud Provider 6, Cloud Provider 9, Cloud Provider 10.

operating in Europe,[236] see Figure 25. These locations have been shared with the ten main cloud service providers and they have been asked to provide details of any errors[237].

Figure 25: Locations of the large cloud service providers' main cloud and edge data centres in EU27 Member States, EFTA and UK in 2021

	Cloud Provider 1	Cloud Provider 2	Cloud Provider 3	Cloud Provider 4	Cloud Provider 5	Cloud Provider 6	Cloud Provider 7	Cloud Provider 8	Cloud Provider 9	Cloud Provider 10	Total Main	Total Edge	Total
AT		E	E		E	E			E	E	0	6	**6**
BE		E	M								1	1	**2**
CZ		E									0	1	**1**
CY		E	E		E	E					0	4	**4**
DE	M	M + 4E	4E	M	3E	M		M	M	E	6	12	**18**
DK		E			E						0	2	**2**
ES		E	E								0	2	**2**
FI		E	M								1	1	**2**
FR		M + 2E	2E	M	2E		M + 2E			M	4	8	**12**
IE		M + E	M + E		M + E						3	3	**6**
IT		E	E	M	E						1	3	**4**
MT		E	E								0	2	**2**
NL		E	M + E	M	M + E	E					3	4	**7**
PL		E					E				0	2	**2**
SE		M + E	E		E						1	3	**4**
CH		E	E		2E	M					1	4	**5**
IS	E	E	E		E						0	4	**4**
NO		E		M	2E						1	3	**4**
UK	E	M + 4E	E	M	E	M			M	E	4	8	**12**
Total Main	**1**	**5**	**4**	**6**	**2**	**3**	**1**	**1**	**2**	**1**	**26**		
Total Edge	**2**	**26**	**17**	**0**	**18**	**3**	**3**	**0**	**1**	**3**		**73**	
Total	**3**	**31**	**21**	**6**	**20**	**6**	**4**	**1**	**3**	**4**			**99**

'M' indicates a main data centre. 'E' indicates an edge data centre. The number indicates the number of main and edge centres.

This list above has been used when investigating data flows to cloud service providers' infrastructures in this study. More specifically, when estimating the destination of the data flows from one country to another.

4.2 Cloud and edge data centre capacity and utilisation

Having examined what cloud is and where cloud and edge data centres are located, it is relevant for this study to try and answer the difficult question 'how much cloud business storage capacity is there?'. Answers to this question will be useful to examine utilisation levels for cloud computing to see how close to full capacity they might be operating and whether further storage capacity appears to be required. After many approaches to cloud service providers, it became clear that for commercial reasons cloud service providers did not want to provide this information.

By utilising information from relatively reputable sources highlighted below, it is possible to estimate cloud data centre capacity and utilisation in Europe.

4.2.1 Analysis to estimate cloud and edge data centre capacity

Four pieces of information from Cisco and ZDNet reports about data centre capacity (this will include cloud and edge data centres) help to provide some insights. The four snippets include:

236 The ten cloud service providers examined have market shares of more than one per cent. Most telecommunications companies (with the exception of Deutsche Telekom) in Europe have a market share of less than one per cent across Europe and have therefore been excluded from analysis.

237 Two responses were received. Information from all relevant sources was triangulated and used to create a 'definitive' list of the 26 main cloud data centres and 73 edge centres in figure 23.

1. The ***Cisco Global Cloud Index: Forecast and Methodology 2016 to 2021*** provides information about the storage capacity of data centres.[238] The study states 'that the installed storage capacity in global data centres will grow nearly 4-fold from 2016 to 2021, growing from 663 EX in 2016 to 2,662 EX (2.6 ZB) by 2021'.
2. ***ZDNet*** report that 'Cloud data centre traffic will represent 95 per cent of total data centre traffic by 2021; compared to 88 per cent in 2016'.[239]
3. The Cisco study estimates that in 2016 18.6 per cent of global cloud traffic was generated in EU27 Member States.[240] By 2021 this is expected to reduce to 17.9 per cent of global cloud traffic.[241]
4. Cisco also estimate that 70 per cent of installed storage capacity is used by business (the other 30 per cent is used by consumers).[242]

By examining the information above for 2016, it is possible to make a tentative estimate of cloud and edge data centre storage capacity in EU27 Member States. The assumptions and analysis are provided below:

a) ZDNet asserts that in 2016 cloud data centre traffic comprised 88 per cent of all data centre traffic. We therefore make the assumption that cloud data centre capacity will be approximately the same percentage (88 per cent) of the global storage capacity of data centres of 663 EX (bullet point one above).[243] This equates to ***global cloud data centre storage capacity of 583 EX in 2016***. Using the same calculations this figure increases to 2,529 EX in 2021;[244]

b) It is assumed that if the EU27 proportion of global cloud data traffic is 18.6 in 2016 (bullet three above), it is not unreasonable to assume that the EU27 Member States will have 18.6 per cent of the global cloud data centre storage capacity of 583 EX. This equates to ***EU27 cloud data centre storage capacity of 109 EX in 2016***. Using the same calculations this figure increases to 453 EX in 2021;[245]

c) Bullet point 4 above notes that 70 per cent of storage capacity is used by businesses. This suggests ***76 EX of business cloud data centre storage capacity in EU27 Member States in 2016.***[246] This figure increases to 317 EX of business cloud data centre storage capacity in 2021.

The preceding analysis has focused on cloud and edge data centre storage capacity in EU27 Member States. These estimates can be extrapolated to the remaining four countries in the study by looking at the relative population sizes of the additional countries (Iceland, Norway, Switzerland, and the UK), see the below Table.[247]

238 Cisco. 2018. 'Cisco Global Cloud Index: Forecast and methodology 2016 to 2021'. p20-21 (https://virtualization.network/Resources/Whitepapers/0b75cf2e-0c53-4891-918e-b542a5d364c5_white-paper-c11-738085.pdf)

239 Ranger S. 2018. 'Cloud computing will virtually replace traditional data centres within three years' (https://www.zdnet.com/article/cloud-computing-will-virtually-replace-traditional-data-centers-within-three-years). This study utilises secondary sources.

240 Section 3.2.4 notes that in other papers Cisco suggest the population is 427 million (this was the size of the EU27 population in c2020). The EU27 population size in 2016 was 444 million (https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Population_and_population_change_statistics). A discrepancy exists but for ease of comparison in Cisco studies Western Europe is assumed to mean EU27.

241 Cisco. 2018. Ibid. page 40

242 Cisco. 2018. Ibid.

243 663 x 88 per cent equals

244 2,662 EX multiplied by 95 per cent.

245 2,529 EX multiplied by 17.9 per cent.

246 109 EX multiplied by 70 per cent

247 Due to rounding totals may not tally. Previous assumptions and estimates have been undertaken at EU27 level, obviously across 27 countries. Estimates made at a fine level of granularity for EFTA (three countries) and the UK must

Table 21: Estimates of cloud and edge data centre storage capacity in EU27 Member States, EFTA and the UK in 2016 and 2021

	Cloud data centre storage capacity 2016 (EX)	Cloud data centre storage capacity 2021 (EX)	Business cloud data centre storage capacity 2016 (EX)	Business cloud data centre storage capacity 2021 (EX)
EU27	109	453	76	317
EFTA	3.5	14.5	2.4	10.2
UK	16	68	11	47
Total	**129**	**535**	**90**	**374**

The assumptions and method to make these calculations is clearly provided above to enable users to undertake their own calculations or sensitivity analysis.

It is possible to extrapolate these estimates and provide forecasts for 2025 and 2030. A large caveat must be expressed about extrapolating the assumptions and estimates in the calculations above. Nonetheless, in 2025 EU27 cloud data centre storage capacity is forecast[248] to be 1,390 EX and business cloud and edge data centre storage capacity would be 973 EX in 2025, this is 2.1 fold increase on business storage capacity in 2021. In 2030, EU27 cloud data centre storage capacity increases to 5,356 EX and business cloud data centre storage capacity is estimated to be 3,749 EX. This is 8.3 fold increase on business storage capacity in 2021. Table 21 uses the same method described above to extrapolate forecasts for EFTA and the UK.

be treated with extreme caution. Aggregated figures are more likely to be correct than those for smaller numbers of countries.

248 Using linear extrapolation methods. These are described in the methodological annex.

Table 22: Estimates of cloud and edge data centre storage capacity in EU27 Member States, EFTA and the UK in 2025 and 2030

	Cloud data centre storage capacity 2025 (EX)	Cloud data centre storage capacity 2030 (EX)	Business cloud data centre storage capacity 2025 (EX)	Business cloud data centre storage capacity 2030 (EX)
EU27	1,390	5,356	973	3,749
EFTA	45	172	31	120
UK	207	798	145	559
Total	**1,642**	**6,326**	**1,149**	**4,428**

4.2.2 Cloud and edge data centre utilisation

Consideration of cloud and edge data centre capacity utilisation provides an insight to future capacity needs and possible cloud data centre investment requirements. If utilisation of cloud capacity is low, future growth might be able to be accommodated in the unused storage capacity.

The Cisco Global Cloud Index: Forecast and methodology 2016 to 2021 states that globally data stored in data centres will grow 4.6-fold from 286 EX in 2016 to 1,331 EX in 2021.[249]

In comparison with the information provided in bullet point 1 of the previous section, this suggests that ***43 per cent of global cloud storage capacity was utilised in 2016*** (663 EX divided by 286 EX). Comparison of the same two sources suggests cloud data centre utilisation **increases from 43 per cent in 2016 to 50 per cent in 2021**.

Tentative linear extrapolation of this trend would suggest that in 2025 56 per cent of cloud data centre capacity would be utilised and this figure increases to 65 per cent in 2030. Once again it must be stressed that these figures are based on assumptions from relatively reputable sources presented at the beginning of section 4, but their methodologies have not been disclosed. They should be thus used with caution.

249 Cisco. 2018. Ibid. page 4

5. Lessons learnt

Replicability of the methodology

This study provides an innovative method of investigating where the enterprise data flowing to cloud and edge infrastructures would flow. The methodology could be re-used in future years to monitor trends in terms of data flows within Europe, between EU27 and EFTA countries and the UK.

The findings could be useful evidence to support the evaluation of the Regulation on a framework for the free flow of non-personal data in the EU free flow of data, for example to assess the extent of the removal of barriers preventing the free flow of data across Europe and support decision making. Moreover, the mapping of cloud data flows across the EU, and the identification of their origin and destination, will allow as well to identify the key cloud data flow corridors where potential investments might be needed for extra cloud data storage.

Lastly, the study aims to lay down the basis for future studies, going beyond the European scene and trying to address the challenges of estimating the value of data flows and also estimating extra-EU cloud data flows.

Add how edge can be better added to this methodology going forward as expanding the methodology.

Models

There is a common aphorism in statistics - 'All models are wrong, but some are useful', usually attributed to George Box[250]. Since all models are wrong, scientists cannot obtain a 'correct' one by excessive elaboration.

We adhere to these principles in developing the cloud data flows model described in Chapter 3. We try to keep the model simple yet robust through the use of official statistics and statistics from more reputable private sector organisations.

The model consists of a **first stage of the analysis** where Eurostat data from 2016 to 2020 is used to provide forecasts about the characteristics of enterprises buying cloud computing services used over the internet until 2030. Analysis and forecasts provided forecasts for the percentage of enterprises buying cloud computing services used over the internet (see Figure 13), the percentage of enterprises buying cloud computing services in different NACE sectors (see Figure 14) and forecasts for cloud service adoption by enterprises of different sizes 2016 to 2030 (see Figure 15).

In the **second stage** of the model, cloud enterprise data is standardised to provide insights to the number of employees in enterprises buying cloud services used over the internet.

The **third stage macro analysis** provides insights to internet traffic per employee which can then be used when estimating the amount of enterprise data flowing to cloud and edge data centres (in stage four). Estimates for internet traffic per employee are presented in Table 11.

250 https://mathshistory.st-andrews.ac.uk/Biographies/Box/

Stage 4 utilises information from the two preceding stages. Stage 2 provided standardised data about the number of employees in enterprises buying cloud computing services used over the interne[251]t. The primary information used from stage 3 concerns estimates of average internet traffic (calculated as GB/month) per employee.

Stage 4 analysis provided estimates and forecasts for total enterprise internet traffic for the NACE sectors (see Figure 19) and enterprises in different size bands (employees), see Figure 20.

Lastly in the **fifth stage,** the model presented the results of research about locations of cloud and edge data centres for each of the 31 countries included in the study. Analysis examined the geographical flows of enterprise data flowing to cloud and edge data centres by enterprises buying cloud computing services used over the internet. Table 16 provided estimates of enterprise data flowing to cloud and edge data servers in the 31 countries examined in the study 2016 to 2021. Figure 23 provided a choropleth map of the net inflow and outflow of data flowing to cloud and edge data centres for the 31 countries examined in the study (TB/month) in 2020

Main data gaps

Commercial confidentiality

As already indicated above, one of the key obstacles during the course of the project has been the lack of responsiveness of the stakeholders approached for the primary data collection.

Cloud service provision is a very competitive market for providing services directly to enterprises and for hosting content that can be accessed by citizens and enterprises.

Many providers use commercial confidentiality as a reason for not providing fulsome responses to questions.

Covid-19

There was no baseline information about enterprise data flows to cloud prior to Covid-19 lockdowns in the EU. Indeed, one reason for the study was to provide these insights.

Lockdowns commenced in March 2020, three months after the start of the study and continue. Most interviews and surveys with stakeholders took place after lockdowns, when many more people will have been working from home.

It is unknown whether results will be representative of more normal times. On balance we believe results will be comparable with pre-Covid times and with the post-Covid era to come. With the exception of a few sectors (e.g. hospitality and some retail) most industries have been undertaking similar volumes of work during lockdown. It is simply the locations where that work might have taken place that will have changed. Most employees are domiciled relatively close to their place of work; few people travel across international borders to their place of work. The results of this study, focusing at national level, should therefore be robust even with Covid-19 happening.

251 The modelling in this paper is largely based on Eurostat ‘enterprise’ statistics. Eurostat describe an ‘enterprise’ as ‘an organisational unit producing goods or services which has a certain degree of autonomy in decision-making. An enterprise can carry out more than one economic activity and it can be situated at more than one location’ (https://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Enterprise). The Eurostat definition of an ‘enterprise’ does not differentiate between the intermediate or ultimate country ownership of an enterprise.

Information and statistics

Many statistics provided by commercial businesses are not underpinned by robust methodologies. Indeed, when approached some commercial providers were not able to provide methodologies describing how information was obtained or how forecasts were made.

National statistics offices provide a great deal of robustly collected and cleaned information. But some data sets have relatively large amounts of data missing. For example, data relating to the number of employees working in enterprises of different sizes is sparse in some sectors.[252]

Eurostat data about the enterprises purchasing cloud services used over the internet is comprehensive but collection omits NACE sectors K (Finance) and the public service sectors NACE O to S (public administration and defence, Education, Human health and social work, Arts and entertainment and other services).[253]

Technical information

Understanding the volumes of data consumed by an enterprise and the volumes flowing to cloud data centres requires considerable technical skill and knowledge. The person to whom one talks in enterprises about the usefulness or efficiency gains stemming from cloud computing does not generally have the necessary technical skills or access to sources of information about data flows stemming from data or cloud service consumption.

The most robust source of information about internet traffic is the management software associated with some routers[254]. This can provide detailed information for enterprises (using the router) about data flows and volumes of internet traffic of different websites and to cloud service providers. Initial examination of router management software suggests it could be a useful source of information in the future. This could be achieved by obtaining sufficient permission from enterprises to view their management software results. Sample sizes for the number of enterprises from which to collect router information will depend on the focus of future studies.

252 This deficiency necessitates the extrapolation of the proportion of employees in companies of different sizes at national level or in similar NACE sectors to enable missing values to be calculated. Sectors with sparse information include electricity and gas (NACE D) water supply (NACE E) and Financial services (NACE K).

253 Eurostat provides data about enterprises purchasing cloud services used over the internet by NACE sector and company size (employees). This deficiency necessitates the extrapolation of the proportion of enterprises purchasing cloud services at national level or in similar NACE sectors to enable missing values to be calculated.

254 Cisco routers, particularly those using Miraki software, provide good insights to internet traffic flows. Some other routers provide similar information, others do not. See the methodological annex.

6. Conclusions

This study examines the volume of enterprise data flowing to cloud data and edge centres across the EU27, EFTA countries and the UK in each of these countries that possessed cloud data and edge centres.

a. Overall data flows

Overall, in the EU27 member states, the flow of **data stemming from enterprises[255] buying cloud services used over the internet** was 504,763 TB/month (5.8 EX/year) in 2020. By 2030 the amount is forecast to be 15.2 times greater at 7,669,835 TB/month (87.8 EX/year) and by 2025 the amount is forecast to be 4.4 times greater (than the 2020 value) at 2,201,058 TB/month (25.2 EX/year).

Across the 31 countries examined[256], **data flows to cloud and edge data centres by enterprises buying cloud services used over the internet** were estimated to be 708,790 TB/month (8.1 EX/year) in 2020. By 2030 the amount is forecast to be 14.9 times greater at 10,577,600 TB/month (121.1 EX/year). By 2025 the amount forecast is to be 4.3 times greater (the 2020 value) at 3,025,388 TB/month (34.6 EX/year).

b. Country level analysis

One element of the analysis was to estimate enterprise data flows (TB/month) to the cloud and edge data centres of the ten main cloud service providers in countries that possess cloud infrastructure between 2016 and 2021[257]. **The largest volume of enterprise data flows is served by cloud and edge data centres in Germany**. In 2020, Germany received 151,968 TB/month (1.74 EX/year) of cloud data flows from other countries. This represents 30.7 per cent of cloud data flows to EU27 Member States and 25.9 per cent of cloud data flows across the 31 countries studied. The EU27 Member State receiving the **second highest inflow of data served by its cloud and edge data centres in 2020 was the Netherlands** (86,963 TB/month, 1 EX/year).

Both these countries (Germany and the Netherlands) received more enterprise data flowing to their cloud and edge data centres in 2020 than the UK (66,064 TB/month, 0.76 EX/year). In 2020, the UK received 12.8 per cent of total data flowing to cloud and edge facilities from the 31 countries studied. EFTA countries received 3 per cent of total data flowing to cloud and edge facilities (15,506 TB/month, 0.18 EX/year).

Analysis discovered that in 2020, the highest outflow of data (data flowing from a country to a cloud and edge data centre) was generated by the United Kingdom (152,815 TB/month, 1.75 EX/ year), followed by Germany (93,474 TB/month, 1.07 EX/year), and France (75,039 TB/month, 0.86 EX/year).

255 In this study, enterprises are defined as private NACE sectors (C to M) and public NACE sectors (N to S). Enterprises consist of mostly businesses (see footnote 8) but also some public sectors (N Administrative and Support Services, P Education, Q Human Health and Social Work, R Arts and Entertainment and Recreation, S Other Service Activities).

256 Liechtenstein is not included in the analysis due to a lack of data on Eurostat

257 Analysis focuses on 2016 to 2021 because information about cloud infrastructure in these years is known. Many commentators predict considerable growth in cloud edge data centres in the next few years. Extrapolation to create forecasts in the future, when the distribution of infrastructure will change considerably, would create spurious results.

c. Net data flows by country

A further stage of analysis examined *net* enterprise data flowing to cloud and edge data centres for the 31 countries examined in the study. This was found by subtracting data inflows to a country from data outflows generated by the country. Three countries have net enterprise data inflows to their cloud and edge data centres. All other countries have net data outflows.

The country with the highest *net* data inflow to its cloud and edge data centres in 2020 was Germany. German net enterprise data inflow to its cloud and edge data centres in 2020 was 58,494 TB/month (0.67 EX/year). **Ireland** has the second highest net enterprise data inflow to its cloud and edge data centres (45,981 TB/month, 0.53 EX/year). The **Netherlands** has the third highest net enterprise data inflow to its cloud and edge data centres in 2020 (45,504 TB/month (0.52 EX/year).

The EU27 Member State with the largest net outflow of enterprise data to cloud and edge data centres in other countries is Italy; 42,923 TB/month (0.49 EX/year). Of the 31 countries examined, the largest net enterprise data out flows were recorded by the **UK**; 86,751 TB/month (1 EX/year).

d. Cloud service adoption and forecasts to 2030

In terms of sectors, in 2020, the largest data flows to cloud come from the Health sector (NACE Q, 12.9 per cent of all flows, 91,600 TB/month; 1.048 EX/year), **Retail and Wholesale** (NACE G, 12.9 per cent, 91,400 TB/month; 1.046 EX/year) and **Education** (12.5 per cent, 88,600 TB/month; 1.01 EX/year). In 2025, it is expected that Retail and Wholesale will have 13.1 per cent of all flows (390,300 TB/month; 4.47 EX/year), the Health sector will have 12.3 per cent of all flows (377,900 TB/month; 4.32 EX/year), and Education will have 12.1 per cent of all flows (367,800 TB/month; 4.21 EX/year). In 2030, it is expected that Retail and Wholesale will have 13.2 per cent of all flows (1,370,670 TB/month; 15.69 EX/year), Health will have 12.1 per cent of all flows (1,293,560 TB/month; 14.80 EX/year), and Education will have 11.9 per cent of all flows (1,262,730 TB/month; 14.45 EX/year).

Moreover, in 2020 the largest volume of data flowing to main cloud and edge data centres comes from enterprises with 250 or more employees. In the EU27 Member States, enterprises with 250 or more employees triggered 61 per cent of all data flows to cloud (309,500 TB/month [3.54 EX/year] of total EU27 flows of 504,700 TB/month [5.78 EX/year] in 2020).[258] EU27 Member States enterprises with less than 10 employees account for 8.5 per cent of all data flows to cloud (42,870 TB/month; 0.49 EX/year). EU27 Member States enterprises with 10 to 49 employees achieved 13.2 per cent of all data flows to cloud (66,560 TB/month; 0.76 EX/year). EU27 Member States enterprises with 50 to 249 employees obtained 17 per cent of all data flows to cloud (85,800 TB/month; 0.98 EX/year).

By 2030 it is expected that 65 per cent of EU27 enterprises with 10 to 49 employees will buy cloud services used over the internet. This is almost twice the number in 2020 (34 per cent) and higher than the 52 per cent of EU27 enterprises with 10 to 49 employees who are expected to buy cloud services used over the internet by 2025[259]. **46 per cent of EU27 enterprises with 50 to 249 employees bought cloud services used over the internet in 2020**. In 2025 and 2030, this is expected to increase to 68 per cent and 83 per

258 In EFTA countries, data flows from enterprises with 250 or more employees comprise 48 per cent of flows (24,700 TB/month of 51,200 TB/month). In the UK, data flows from enterprises with 250 or more employees comprise 66 per cent of flows (100,500 TB/month of 152,800 TB/month).

259 EFTA: 2020 – 63 per cent, 2025 – 90 per cent, 2030 – 98 per cent. UK: 2020 – 41 per cent, 2025 – 72 per cent, 2030 – 95 per cent.

cent, respectively[260]. **For EU27 enterprises with over 250 employees, in 2020, 66 per cent bought cloud services used over the internet.** In 2025 this is expected to increase to 90 per cent, and then to 97 per cent by 2030[261].

In terms of services[262], the proportion of enterprises buying low and medium levels of cloud services used over the internet is declining. An increasing number of enterprises buy high level cloud services used over the internet. In 2020 19 per cent of EU27 enterprises used low levels of cloud services, 30 per cent used medium levels of cloud services and 51 per cent used high levels of cloud services. In 2025, it is forecasted that 15 per cent of cent of enterprises will be using low levels of cloud services, 22 per cent using medium levels of cloud services, and 63 per cent will be using high levels of cloud services. By 2030, the situation is expected to change significantly - 72 per cent of enterprises are expected to be using high levels of cloud services, 16 per cent are forecast to use medium services and only 12 per cent will be purchasing low levels of cloud services.

Concerning data types, 41 per cent of the total data stored in cloud infrastructure is personal data and 59 per cent is non-personal data. Of these 41 per cent of personal data, 11 per cent is generated by a user/individual. Concerning non-personal data, out of 59 per cent, around 22 per cent is generated by a machine. Within machine-generated data, 19 per cent is industrial (productive) data, namely used as a direct input into the production process from a supply chain perspective and the remaining 3% are indicated as “other data” by survey respondents.

Overall conclusion

To conclude**, this study provides an innovative method for estimating the volume and types of enterprise data flowing to cloud infrastructures and for investigating where data flows across the EU.** The report provides a holistic approach and an integrated view of enterprise data flowing to cloud data centres and edge centres within the EU; between the EU and the UK; and between EU and EFTA countries. It captures several dimensions of the data flows, namely the sectors[263] (NACE sectors C-N) triggering the flows, enterprise size (small, medium, large enterprises),[264] services,[265] and data-intensive enterprise activities[266] and data types (personal/non-personal).[267] The study was not able to investigate data flows between the cloud and edge data centres of the businesses providing cloud services. These exchanges might arise so that businesses providing cloud service

260 EFTA: 2020 – 81 per cent, 2025 – 97 per cent, 2030 – 99 per cent. UK: 2020 – 63 per cent, 2025 – 92 per cent, 2030 – 98 per cent.

261 EFTA: 2020 – 90 per cent, 2025 – 98 per cent, 2030 – 100 per cent. UK: 2020 – 77 per cent, 2025 – 87 per cent, 2030 – 97 per cent.

262 Eurostat definitions of low, medium and high cloud computing (CC) services are as follows; low CC services are email, office software, storage of files, medium CC services include low CC services plus hosting of the enterprises’ database, high CC services are accounting software applications, CRM software, computing power.

263 C Manufacturing, D Electricity Gas, Steam and air conditioning supply, E Water Supply, F Construction, G Wholesale and retail trade, H Transportation and storage, I Accommodation and food service activities, J Information and Communication, K Financial and insurance activities, L Real estate activities, M Professional, scientific and technical activities, N Administrative and support services

264 Eurostat definition: small enterprises: 10-49 persons employed; medium-sized enterprises: 50-249 persons employed; small and medium sized enterprises (SMEs): 1-249 persons employed; large enterprises: 250 or more persons employed.

265 See ‘Definitions’ in Section 1.2.

266 ‘Data intensive activities’ refers to data stored in cloud infrastructures and generated by activities that are highly dependent on the use, production or/and provision of the cloud technology, such as industrial data, which is one of the typical data to feed into customer relationship management (CRM) activities. For more information, please refer to Annex 1.

267 ‘Data type’ refers to personal and non-personal data. For more information, please refer to Annex 1.

can back up data or operate more efficiently. Furthermore, the study did not investigate data flowing to private cloud infrastructure on company premises.

The methodology presented here can be used in future to monitor data flow trends across the European Union (the EU), between the EU and EFTA countries and the UK in support of EU policy, trade and investment decisions. Furthermore, the findings provide useful evidence to support the evaluation of the Regulation of the free flow of data (for example to assess the extent of the removal of barriers preventing the free flow of data across Europe) and the tools for a continuous analysis of data flows and the economic development of the EU's data processing sector. Moreover, the mapping of cloud data flows across the EU, and the identification of their origin and destination, will allow to identify the key cloud data flow corridors where potential investments might be needed for extra cloud data storage.

Finally, the study lays down the basis for future work, going beyond the European scene and trying to address the challenges of estimating the value of data flows and also extra-EU cloud data flows.

Annex 1: Designing a typology of data and data flows

Annex 1 explains in more detail the starting point of the methodology built for mapping data flows, which was to develop a robust typology of cloud-based data flows. As mentioned in the report, the study team made an overview of the existing literature review of data flows, in order to design a tailored typology of data flows to be mapped by the study.

This was done through three main research tools: literature review, scoping interviews and expert workshop. The findings are highlighted in this Annex to provide an overview of the work that was done worldwide in terms of measurement and identification of the key dimensions of data flows or other types of flows that could help approximate cloud-based data flows.

The study team selected studies which highlighted the main features of data types and flows. It should be noted that the scope of the literature review encompassed different types of flows (cloud-based data flows, other types of data flows and other types of flows) that are not necessarily cloud-based data flows, due to the lack of literature on cloud-based data flows. However, other types of flows are still relevant to identify key dimensions of flows if properly contextualised for cloud-based data flows.

The outcome of the exercise was the development of preliminary typology of data flows to be used for the study. Different dimensions of data flows were explored to then design a typology of data flows. Seven dimensions were initially identified: *1) data types, 2) technology enabling data flows, 3) actors using cloud services, 4) functional areas related to activities within an entity, 5) services, 6) sectors, 7) geography*.

A summary of the key findings of data flows are outlined below.

1. Typology of data flows

1.1. Key takeaways

The main features identified that enable a satisfactory level of granularity to disentangle data flows for the designed typology are outlined below:

- **Technology used** to exchange data (such as edge computing, cloud computing), technology services originating data flows (such as big data analytics) **and technology infrastructures** (such as internet-connected machinery, data centres)[268] to generate and store cloud-based data are dimensions to characterise data flows outlined by different authors. The authors of the study consider it as a useful approach to assess a) *the actors' relationship*, b) *the data type exchange*, (for example whether it is machine generated in case of internet-connected machinery)

268 Koloch G., Grobelna K et al., 'Data utilisation intensity and economic performance- a diagnostic analysis', 2017. UNCTAD, 'Digital economy report, value creation and capture: implications for developing countries', 2019. McKinsey, Globalisation in transition: the future of trade and value chains, 2019.

and; c) *intra company flows*, since it captures different purposes of the flows within a company (such as back-office consolidation). This is especially relevant for cloud-intensive companies, which are most likely to be large in size and of multinational nature given that cloud computing is used by 40% of large enterprises, while only by 17% of SMEs in the EU.[269]

Moreover, this categorisation is also of relevance to capture flows stemming from private cloud infrastructures, which otherwise may not be captured when looking at the provision of cloud services by public cloud services providers. Cloud services can be managed on premise and data centres can be privately owned and/or managed; hence the technology dimension must also capture flows that do not necessarily stem from public cloud service management but also from private cloud.

- Another key feature highlighted is the **relationship between the different actors** (vertical, horizontal and external data sharing),[270] which is helpful to disentangle *flows within the data flows value chain*. This categorisation is also useful in the context of the study approach aiming at analysing the cloud-based dimensions in the data flow value chain, by highlighting the relations within a sector, between different companies, whether it is a flow among companies interacting in a customer or supplier relationship type (i.e. vertical), among organisations involved in the same stage across the value chain, or originating flows with organisation outside the sector.
- The **relationship between the end-user and the type of transaction**[271] is also highlighted by scholars to capture a) the overall non-commercial data traffic, b) the types of transactions and services between buyers and sellers and c) the ones delivered to and from end-users. According to several authors, a flow of data is associated to an economic transaction, that could be not only commercial, but from which an economic gain could still stem from (such as corporate internal efficiency gains).

  This aspect is also useful to characterise *flows by types of actors such as among businesses, between business and consumer,* deriving from the analysis of the type of transactions triggering a data exchanged. For example, it would be useful to capture the cloud-based data flows stemming from the use of a company digital application (such as for mobility purpose) by a user via a connected device.

1.2. Gaps identified

The identified gaps from the literature review, scoping interviews and expert workshop with regard to data flows are the following:

- Public entities and associated data flows are not identified, such as government to end-users ones, government to government ones, government to business ones are missing aspects in the different studies.
- A sectorial approach that enables the mapping of cloud-based data flows among industrial sectors is missing. Instead, the existing literature is focusing on ICT at

269 Eurostat 2018, Digital Economy and Society Statistics, available at: https://ec.europa.eu/eurostat/statistics-explained/index.php/Digital_economy_and_society_statistics_-_enterprises#Use_of_cloud_computing_services.

270 Deloitte, 'Realising the economic potential of machine-generated, non-personal data in the EU' Report for Vodafone Group, 2018.

271 Koloch G., Grobelna K et al., 'Data utilisation intensity and economic performance- a diagnostic analysis', 2017. Mantelero, A., 'Cloud computing, trans-border data flows and the European Directive 95/46/EC: applicable law and task distribution', Journal for Law and Technology.

large, while there are several industrial sectors, such as manufacturing, electricity, utilities, wholesale and retail sectors showing upper-medium dependence towards cloud computing services (Eurostat 2018)[272] that are not yet mapped. This is therefore a necessity given the real dependency of those sectors towards cloud computing.

- An analysis on data flows stemming from edge computing is also missing. This is of particular relevancy given the EU Data Strategy,[273] released on 19 February 2020, envisions that by 2025 more than 80% of the processed data will be taking place at the edge while only 20% in central cloud infrastructures. This tendency is further confirmed when assessing the growth of the IoT market to reach up to 71 billion devices by 2025 worldwide. Machine to Machine connections will also constitute half of the global connected devices by 2025,[274] showing the necessity to carefully map edge flows in the study to bring a realistic overview of current data flows across Europe.
- The different categories do not either provide a breakdown between intra-corporate cloud-based data flows that identify the different internal corporate purposes for which the different type of cloud services is consumed in key business activities. Examples of cloud services are data back-up services, email exchanges, video conferencing set-up, HR data storage services, software security enablement, CRM and ERM services despite being among the most intensive cooperate functions relying on the consumption of cloud services.[275]
- Another missing aspect from the literature review is a process categorisation for actors differing from businesses, such as public entities and academia. This categorisation would allow for breaking down flows of data along the policy - making and research processes for instance.
- Last but not least, a major point is the measurement of cloud-based data flows as an unexploited research field within the literature review. The only existing measurement relates to data traffic at large, without disentangling flows stemming from either the direct and/or indirect consumption of cloud services or from the provision of cloud services and switching activities among cloud service providers (such as flows among data centres). This gap is mainly due to the still perceived novelty of cloud computing technology. However, the literature review is still a useful exercise to capture the highlighted dimensions in this section that could be adapted to cloud-based data flows from which indicators and proxies could then be derived from to ultimately measure those flows, once contextualised.

Summary of the key findings related to the design of the typology of data flows stemming from the literature review, expert workshop and interview is provided in the table below.

272 Available at: https://ec.europa.eu/eurostat/statistics-explained/index.php/Cloud_computing_-_statistics_on_the_use_by_enterprises#Enterprises_using_cloud_computing

273 Available at: https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020_en.pdf

274 Cisco (2018), 'Cisco Global Cloud Index: Forecast and Methodology 2016-2021' white paper. Available at: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/global-cloud-index-gci/white-paper-c11-738085.html

275 Desi 2019, 'Integration of Digital Technology by Enterprises', available at: https://ec.europa.eu/digital-single-market/en/integration-digital-technology

Table 23:Summary of data flows dimensions identified through literature review, scoping interviews and expert workshop

	Dimension	Usefulness
Useful inputs	Technology used	To capture the different technologies used to exchange data and thus triggering cloud-based data flows
	Relationships among actors	To characterise different cloud-based data flows among actors based on the types of relationships among actors, such as B2B, B2C, G2U, etc
	Transaction type	To capture both non-economic and economic nature of cloud-based data flows
Missing aspects	Sectorial approach	To identify the cloud-intensive sectors which generate cloud-based data flows
	Actors types	To capture the heterogeneity of actors involved in cloud-based data flows, such as cloud data users, providers and infrastructure network providers.
	Geographical	To capture intra and inter entities cloud-based data flows and the origin and destination of a flow
	Functional areas	To identify the business purposes triggering cloud-based data flows
	Service consumption, production and provision	To capture the types of cloud services consumed, produced or provisioned generating cloud-based data flows

2. Cloud intensive data types

The study team also analysed the different categorisations made by the different authors concerning data types. Indeed, the approach undertaken by most of the authors was to focus on analysing data types to then derive a generic typology of data flows (however not specific to cloud-based data flows). As stated in the methodological framework, the study team adopted an equally valid, however, more holistic approach by bringing the 'types of cloud intensive data'[276] as one of the seven dimensions which characterise cloud-based data flows. To this end the study team started by building first an objective typology of cloud-based data flows (based on the above seven described dimensions in section 1.2.3). The overall types of data highlighted in the literature were then examined into a matrix, to highlight which from those can qualify as '**cloud intensive data'** to feed into the typology of cloud-based data flows.

2.1. Key takeaways

The main features identified to enable a satisfactory level of granularity to disentangle cloud intensive data types for the cloud-based data flow typology are outlined below:

276 'Cloud-intensive data' refer to data generated by activities that are highly dependent to the use, production or/and provision of the cloud technology, such as industrial data, one of the typical data to feed into CRM activities.

- Personal[277] and non-personal data[278] are distinguished by some authors based on the categorisation of **machine-generated data versus user-generated data.**[279] Despite the fact that user-generated data can also be anonymised, and used for example as an output in an industrial process,[280] we consider the distinction between machine generated (anonymised input) and user generated as the most appropriate available proxies (while not perfect) to disentangle data flows from personal and non-personal data. This proxy will be used in our study to characterise the data types dimension.

  Data types are also disentangled by the **way data are collected** based on human added value input or process-mediated and machine collection,[281] which is important to factor into the study analysis considering the level of granularity needed for the study in particular to disentangle cloud intensive data types, from personal and non-personal, The distinction will be made by distinguishing data generated by human inputs (i.e. via social networks, comments, documents, videos, pictures, mobile data content, user-generate maps, emails which can be opinions and comments) from automated data generated by machine or processes-mediated. Process-mediated data relates to automated monitoring of the state of resources and recording business activity (i.e. anonymised medical records for public authorities and commercial transaction, banking records for businesses, workflow management). Lastly machine generated data are data from sensor, for example stationary sensors (home automation, weather/pollution sensors, traffic sensors, scientific sensors); mobile sensors (mobile phones locations, cars, satellite images) and data from computer systems (logs from IT systems, Web logs, which will include also data processed into cloud data centres).

  Another categorisation of the way data is generated is made by **distinguishing between data generated by networks, devices, and assets.**[282] This distinction is relevant for our study to factor in particular industrial cloud-based data flows stemming from IoT devices. The literature review defines 'device data' as either data generated by IoT devices stemming from sensors intended to gather data regarding external conditions, or data that are a by-product of a machine's core function. Furthermore, 'network data' and 'asset data' are also relevant to be captured into the study. In the literature review 'network data' relate to operational activity, system resilience, schedules, traffic flows and emissions, while 'asset data' relate to data licences, location, output, and usage, for example data generated from interconnected machinery.[283]

277 European Regulation 2016/679 of the European parliament and the Council, art. 4 defines personal data 'any information relating to an identified or identifiable natural person (…), directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person'

278 European Regulation 2018/1807 (COD) of the European Parliament and the Council, which defines non-personal data, as other data than personal, defined by article 4.1 of 2016/679.

279 Deloitte, 'Realising the economic potential of machine-generated, non-personal data in the EU' Report for Vodafone Group, 2018.

280 European Commission, COM(2019) 250 final, Commission Guidance on the Regulation on a framework for the free flow of non-personal data in the European Union.

281 Koloch G., Grobelna K et al., 'Data utilisation intensity and economic performance- a diagnostic analysis', 2017.

282 Deloitte, 'Realising the economic potential of machine-generated, non-personal data in the EU' Report for Vodafone Group, 2018.

283 Deloitte, 'Realising the economic potential of machine-generated, non-personal data in the EU' Report for Vodafone Group, 2018.

- The categorisation of the **production stages** (i.e. control, pre-production, supply chain management, production, post sales)[284] is also useful to capture the business purpose for which different types of data are created within the different functional areas of a business. For example, those data which are generated and exchanged for supply chain management purpose versus those for human resource management corporate needs.

2.2. Gaps identified

The gaps identified from the literature review, scoping interviews and expert workshop with regard to cloud-intensive data types are the following:

- Data activities related to other actors than businesses, such as public entities and academia, which are also users of cloud services, are not captured across the production stages of the public sector (i.e. public service delivery process) and academia (i.e. research and publication processes). Those only relate to business activities. Given the scope of the study, public entities and academia should be included.
- A distinction between industrial data used as direct input into the production process from a supply chain perspective and organisational data used as intermediary input for internal corporate activities, (such as human resources) is not captured either. The distinction is, however, necessary to disentangle on the one hand for the use of cloud-based data as either direct input into the production process for example for sensor monitoring device services in the manufacturing sector. On the other, cloud-based data as intermediary input into the internal corporate and/or governmental activities for example for performance management, fiscal, human resource and financial activities. This distinction would be necessary to feed into three of the seven dimensions that characterised cloud-based data flows being: cloud intensive data types (dimension 1), cloud intensive sectors (dimension 6) and types of functional activities (dimension 4).

To summarise the findings on the typology of cloud-intensive data types identified through the literature review, the expert workshop and interviews, a table is provided here below.

284 Kommerskollegium, ‘No transfer, No production, a Report on Cross-border Data Transfer Global Value Chains, and the Production of Goods’, 2015.

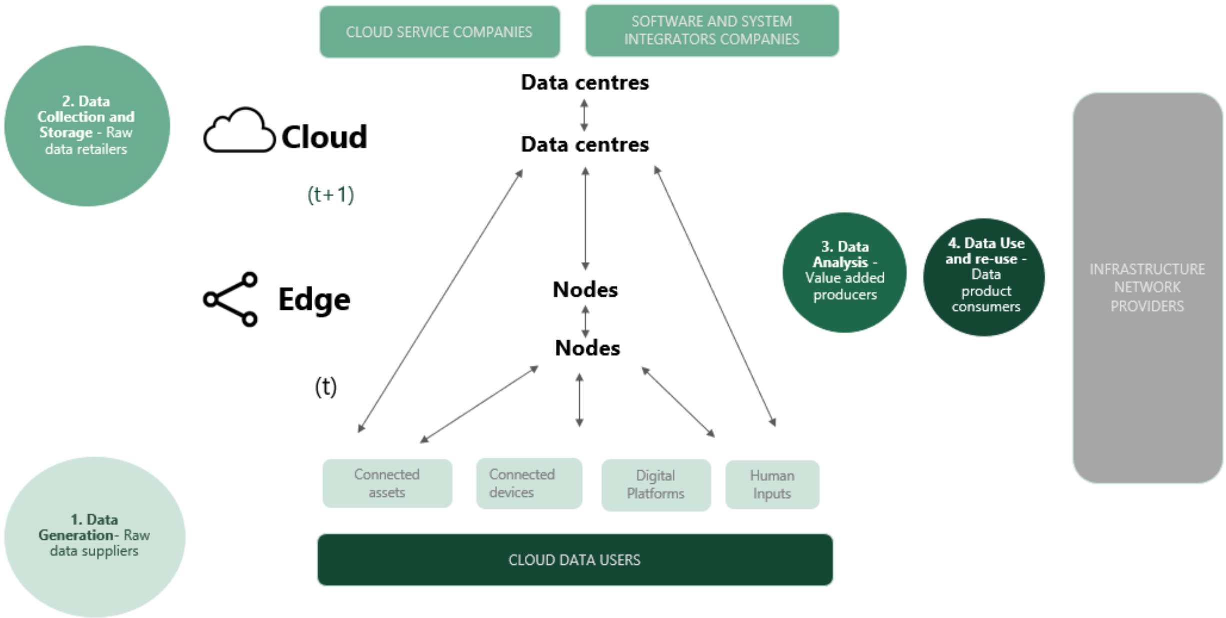
Table 24:Summary of the key highlights on the 'cloud-intensive data types' as a dimension of the cloud based data flow typology based on the literature review, scoping interviews and expert workshop

	Dimension	Usefulness
Useful inputs	Machine/ human generated	To disentangle personal and non-personal data
	The way data are collected	To disentangle personal and non-personal data
	Production stages	To capture the business activities among which data are flowing from to
Missing aspects	Data flow stemming from activities other than productive activities	To capture the types of cloud-intensive activities other than business activities (i.e. related to public administrations and academia)
		To capture data flows stemming from activities, other than the productive ones such as internal corporate and governmental process efficiency activities (such as data back-up, data storage cloud data security services, video conferencing an collaborative services… etc.)

The study team also explored existing methodologies to measure flows in other relevant sectors, such as the energy sector. Desk research was conducted to identify replicable elements to measure flows, in particular on the electricity sector, by looking at the transmission and distribution layers. However, no findings were applicable to our study, given that the energy infrastructure sector has a regulated transmission and distribution infrastructures, associated with mandatory advanced monitoring technologies which measure real-time thickness of electricity flows at both transmission and distribution level unlike the cloud computing sector.

Overview of typology of data flows, key actors A typology of flows applicable to cloud-based data – or **Typology of data flows**, was designed based on the main dimensions of data flows using the data flow value chain as main conceptualisation framework. To that end, the study team focused on the stages of the data flow value chain where cloud-based data flows could be identified to then analyse them by disentangling their diverse dimensions and actors involved, as the Figure below shows.

Figure 26: Illustration of the key actors and related flows



Source: VVA elaboration

Cloud-based data are generated (stage 1) to then be collected and stored in stage 2. Different flows may stem from data collection and storage with the purpose of aggregating and analysing cloud-based data (stage 3) and then being consumed for broader purposes (stage 4). Looking at the actors and their involvement, three main phases can then be outlined:

1. **Data flows from cloud data users to infrastructure storage** (edge or cloud): data generated by human inputs, connected assets and devices (IoT) are transmitted to edge nodes (i.e. router, switcher, base station), in case of edge computing technology, or/and to cloud infrastructures in case of cloud computing. Digital platforms in this context can be the intermediary between the generation and the technologies used for data flows. They enable the collection of the inputs, for example generated by connected devices or human inputs, and transmit to edge or cloud infrastructures.
2. **Data flows from infrastructure to infrastructure** (edge to cloud, edge to edge, cloud to cloud, cloud to edge): the edge nodes process data from the device and transmit the relevant data to cloud, cloud infrastructure can transmit data to other cloud infrastructures or edge ones.
3. **Data flows from infrastructure to user/device**: edge nodes transmit directly back to device, in case for real time application needs, or cloud infrastructure transmit data directly to the user.

After looking at flows from different phases, within the cloud-based data flows, the study team identified some key questions in the light of using a deductive approach to develop the Typology of data flows. The end goal of this deductive approach was to complement the findings from the data flow value chain and literature review, by further detailing the dimensions identified, when answering to the following questions.

- **What type of data are exchanged?** To answer the question, the dimension of **data types (dimension 1)** aims at describing which type of cloud-based data are exchanged. This dimension encompasses human generated data, namely data generated by human inputs, and machined generated data such as cloud-based data generated by connected assets and devices. Human generated versus machine generated is the proxy used in this study to distinguish personal data flow from non-personal data flows.

- **Which technology and service enable data flows?** The **technology dimension (dimension 2)** answers the question of how cloud-based data flows are exchanged, by disentangling flows by the type of technology used, focusing primarily on edge and cloud computing technologies for the purpose of this study. Moreover, the technology dimension gives some information on the time and latency requirements for sensitive data, which are exchanged through edge computing technology in case real-time exchange is required, while cloud computing technology is used for less time and latency sensitive data. The timing dimension, which would answer the question when cloud-based data flows occur, was not added in the typology as a dimension per se, due to the difficulty to provide a graphical representation of the timing of data and ultimately to include it in the mapping exercise. Moreover, the **service dimension (dimension 5)** departs from the type of services used, such as Software as a Service, Platform as a Service, Infrastructure as a Service, edge services, or cloud added value services (data analytics, AI services, blockchain …) which trigger cloud-based data flows.
- **Which are the actors and associated sectors triggering cloud-based data flows?** The **actor dimension (dimension 3)** describes which actors are generating cloud-based data and thus triggering cloud-based data flows. This is done by identifying the cloud data users, network infrastructure providers and cloud service companies as well as companies consuming cloud data from different sectors of the economy (such as the banking sector, energy, manufacturing…) for their business activities. The **sectorial dimension (dimension 6) is therefore also captured by** identifying cloud intensive sectors using cloud-based services to their sectorial activities that also generate cloud-based data flows.
- **From which activities are cloud-based data flows stemming from?** The **functional areas dimension (dimension 4)** aims at understanding the purpose or activity generating the flow, for example cloud-based data flows can be generated for the purpose of security activities, back-up data storage and human resources management. The functional area dimension is also key for characterising the data flows as commercial or non-commercial output. Indeed, some cloud-based data flows are stemming from activities that generate a commercial output (productive activities), for example purchasing and procurement activities. While other cloud-data based flows stem from activities without generating a commercial output, in the case of activities that are internal to a company (corporate activities), for example for data storage back-up and for data replicability, for enterprise security and cybersecurity.
- **Among which entities are cloud-based data flows happening?** The **actor dimension (dimension 3)** answers the question by capturing the organisational entity types among which a data flow is triggered. This can be intra-entity or inter-entity flows. Cloud-based data flows within a single company, public entity, academia, or cloud-based flows between entities. This dimension also defines whether the data flow is an industrial or a public entity flow, namely stemming from a business actor, or from a public one (including public administration and academic actors).

- **Which is the origin and destination of cloud-based data flows?** The geography dimension (dimension 7) aims also to characterise the geographical origin and destination of cloud-based data flows, which can happen within a **country or/and cross-border,** namely crossing multiple countries.

The seven dimensions of the Typology of data flows are not mutually exclusive per se, and they capture different levels of granularity of cloud-based data flows. For example, a flow between two sectors could be defined by the sectorial dimension (dimension 6), but also by the actors generating the flows (dimension 3), or by the functional area from which they originate from (dimension 4). A business entity working in the agriculture sector could generate a cloud-based data flow, by collecting the inputs from sensors and monitoring devices in the production. The connected device used in the production is transferring data real-time to edge nodes and back to the production site to monitor the quality and condition of production. This flow could be defined as non-personal data flow, since it is generated by a machine and cannot be used to identify a natural person (dimension 1). Moreover, the flow could be also defined by the underlying technology enabling it, namely an edge computing data flow (dimension 2). The cloud-based data flow could also be characterised as a flow internal to the business entity (dimension 3) and for operation and assembly purposes (dimension 4). The services used would be edge added value services (dimension 5), and the flow would also be defined as agriculture sector flow, since it is originated by a business from this sector (dimension 6). Going a more macro dimension, the flow could also be defined as 'within a country' (dimension 7) since the flow is happening between sensors in the production site to the edge nodes near to it and back to the production site.

The seven dimensions identified were reduced to five, which were actually measurable by the study team: *country, sector, enterprise size, services and type.*

Annex 2: Summary table of typology dimensions used to analyse data flows

Annex 2 summarises the dimensions of the Typology of data flows used in the study. The final framework was built with the objective of collecting data through secondary (macro-level approach) or primary (micro-level approach) sources on the Typology dimensions.

Table 25: Overview of typology data flows

What should the typology of data flows capture?	Dimensions	Approach	Dimension disentangled by
Which is the origin and destination of cloud-based data flows	Geography	Macro-level	• Enterprise data flowing to cloud data centres from and to countries
Which are the actors and associated sectors triggering cloud-based data flows? **Among which entities cloud based data flows are happening?**	Enterprise	Macro-level	• Enterprise data flowing to cloud data centres
	Sector		• Enterprise data flowing to cloud data centres by NACE SECTOR (C to N and O)
Which technology and service enable cloud-based data flows?	Services	Macro-level	Enterprise data flowing to cloud data centres by • Low use • Medium • High use
		Micro-level	• PaaS • SaaS • IaaS • Edge Services • Cloud to edge added value services
What type of cloud intensive data are exchanged?	Cloud-intensive data types	Micro-level	Enterprise data stored in cloud data centres by • User generated • Machine generated (anonymised)

Annex 3: Detailed methodology to examine enterprise cloud data flows

1. Introduction

A key objective for this study was 'to develop and test a new, self-sustained and replicable methodology to estimate data flows'. This methodological annex provides a more detailed overview of the five stage research process than was possible in the main body of the report.

1.1. Introduction and policy context

The main body of the report was required to present methodological elements of the research alongside qualitative and quantitative results obtained from analysis. To enable a better understanding of the methodology this section focuses only on the methodology. Qualitative and quantiative information is provided in this annex, but this is used to describe (usually in a step-by-step process) how components of the analysis were undertaken.

A further advantage of having a separate methdological index is that detailed and complex elements of the methodology can be described in more detail, without interrupting the presentation of key findings from analysis in the main report.

This annex focuses on each of the five stages of the methodology. The introduction to each stage of the methodology consists of a short description of the key activities undertaken.

In the second section of each stage, Belgium is used as an example to demonstrate the methodological approach. It is hoped that by focusing on the application of the methodology for one country the approach should be easier to understand than by describing analysis for all 31 countries included in the analysis.

Further details concerning the methodology are then described in the remainder of the methodological review of each stage of the study.

1.2. Developing a robust data foundation for modelling

Prior to this study there was not thought to be quantitaive studies or modelling work to investigate interent traffic or data flows to cloud and edge data centres. The lack of previous insights and methodologies was confirmed by participants at the first project workshop held on 6th February 2020 in Brussels.

This study is therefore thought to be the first to investigate and model data flowing to cloud and edge data centres by enterprises buying cloud computing services used over the internet.

There is a common aphorism in statistics – 'All models are wrong, but some are useful', usually attributed to George Box[285]. Since all models are wrong, scientists cannot obtain a 'correct' one by excessive elaboration.

We adhere to these principles in developing the model of enterprise data flowing to cloud and edge data centres in this study. We try to keep the model simple yet robust through the use of official statistics such as Eurostat. Where official statistics[286] are not available, we used statistics from more reputable private sector organisations. Where possible we tried to obtain information or insights from several reputable private sector organisations so that extreme values could be recognised. Where methodologies for their calculations were provided by reputable private sector organisations, these are described in the study. However, methodological descriptions were scarce. In these cases, we contacted the organisation to seek further information. Where replies were received, details are provided in this report.

Finally, like most research developing models, this study is founded on a number of assumptions. Where assumptions are made these are described in this annex. In developing the methodology, we have adopted Occam's razor or the 'law of parsimony'[287]. This is a general guideline to "keep it simple". When confronted with competing explanations for events, it is usually sensible to adopt the interpretation that pivots on the smallest number of supplementary assertions and assumptions.

1.3. Definition of an enterprise

This study has adopted the definition of an 'enterprise' used by Eurostat. Eurostat describe an 'enterprise' as 'an organisational unit producing goods or services which has a certain degree of autonomy in decision-making. An enterprise can carry out more than one economic activity and it can be situated at more than one location'[288]. The Eurostat definition of an 'enterprise' does not differentiate between the intermediate or ultimate country ownership of an enterprise

Eurostat use the term enterprise for organisational units in NACE sectors C to S. The term 'enterprise' thus includes organisational units usually considered to be the private sector (NACE C to M)[289] and NACE sectors N to S[290] commonly regarded as 'public sector'.

285 https://mathshistory.st-andrews.ac.uk/Biographies/Box/

286 By 'official statistics' we mean National Statistics offices or international statistics organisations such as Eurostat.

287 Ball P. 2016. The tyranny of simple explanations. The Atlantic. https://www.theatlantic.com/science/archive/2016/08/occams-razor/495332/

288 https://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Enterprise

289 NACE C. Manufacturing, D. Electricity, gas, steam and air conditioning supply, E. Water supply; sewerage, waste management and remediation activities, F. Construction, G. Wholesale and retail trade; repair of motor vehicles and motorcycles, H. Transportation and storage, I. Accommodation and food service activities, J. Information and communication, K. Financial and insurance activities, L. Real estate activities, M. Professional, scientific and technical activities

290 NACE N. Administrative and support service activities, O. Public administration and defence; compulsory social security, P. Education, Q. Human health and social work activities, R. Arts, entertainment and recreation, S. Other service activities

1.4. Extrapolation and s-shaped adoption curves

This study is producing information for the European Commission initiative focussing on 'The European Data Flow Monitoring'[291]. The study is required to provide a replicable methodology to estimate and monitor data flows within the EU27, EFTA countries, and the UK. One of its core objectives is to support the evaluation of the Free flow of non-personal data Regulation now and in the future.

Eurostat provides historic data but it does not provide forecasts for future trends for the variables used in this study. In order to provide forecasts for the future, extrapolation methods are utilised.

Extrapolation involves making statistical forecasts by using historical trends that are projected for a specified period of time into the future. It is only used for time-series forecasts. Extrapolation is fairly reliable, relatively simple, and inexpensive. Extrapolation is this methodology is useful when major changes are not expected; that is, causal factors are expected to remain constant. Causal factors are more likely to remain constant over the shorter term. The results of extrapolation are therefore more likely to be reliable over the short-term than longer term[292]. Extrapolation, which assumes that recent and historical trends will continue, produces large forecast errors if discontinuities occur within the projected time period.

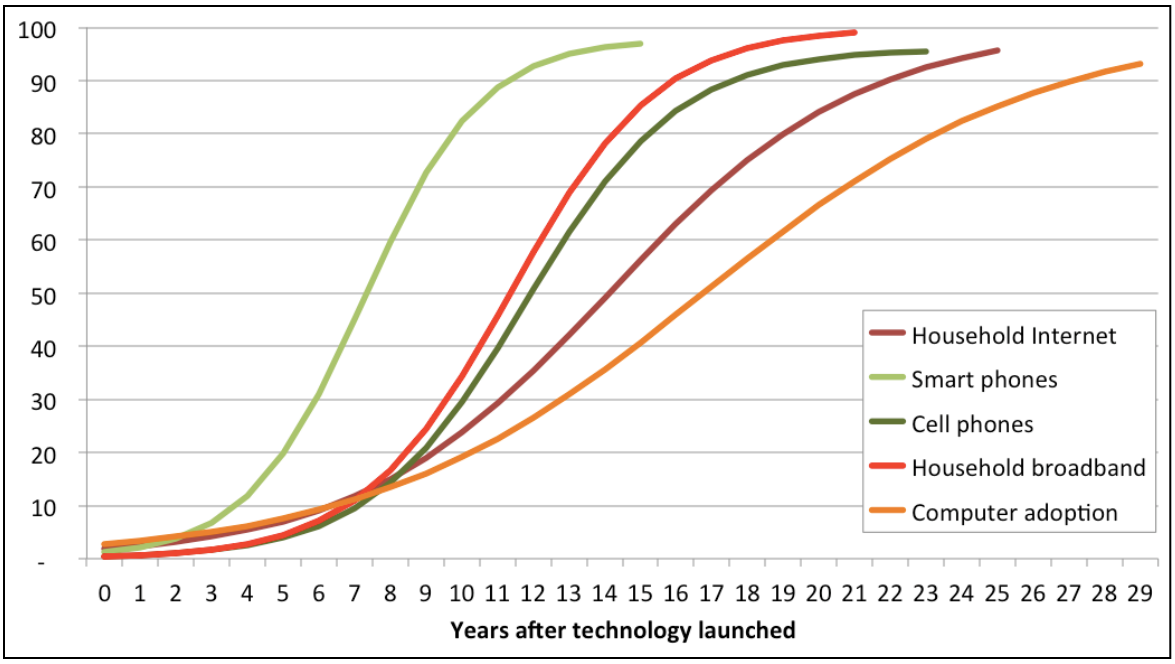
Forecasts in this study were made using the MS Excel inbuilt function. This method assumes that the change between two values is linear and that the margin of error is insignificant.

Linear extrapolation, using the MS Excel function, was used in making forecasts to 2030. Linear extrapolation is useful, but when used with percentage data it can produce forecasts in excess of 100 per cent. Eurostat data about enterprise purchase of cloud services used over the internet is provided as percentages. Clearly it is impossible for there to be more than 100 per cent of enterprises purchase of cloud services used over the internet.

It is well known that technology adoption often follows an s-shaped pattern, see Source: Tech4i2. 2015. SMART 2014/0012

Figure 27. S-shaped adoption curves are generated from Logistic models, Gompertz models and/or Fisher-Pry models. Adoption rates across the majority of the time period for a technology are relatively linear and consistent. But at the start and end of adoption periods the curves become more 's-shaped' in appearance.

291 'European Data Flow Monitoring'. https://digital-strategy.ec.europa.eu/en/policies/european-data-flow-monitoring
292 Glantz M. 2011. Projections and risk assessment.



Source: Tech4i2. 2015. SMART 2014/0012

Figure 27 S-shaped percentage adoption curves for five technologies in EU27 Member States

To adjust for linear extrapolation, forecasts of more than 100 per cent data from s-shaped adoption curves is utilised. The data represents the flatter parts of the s-shaped curve at the start and end technology adoption. The string of values below was entered in the appropriate cell when values exceeded 95 for those extrapolated to exceed 100 in 2030, see the data below.

95.0 96.0 96.8 97.5 98.1 98.6 99.0 99.3 99.5 99.6 99.7 99.8 99.

2. Stage One: Characteristics of enterprise buying cloud services 2016 to 2030

2.1. Introduction

The first stage of the study examined Eurostat data from 2016 to 2020 about enterprises buying cloud computing services used over the internet. The goal of the study was to the geographical flows of enterprise data flowing to cloud and edge data centres by enterprises buying cloud computing services used over the internet. It was therefore important at the start of the study to understand how many enterprises were using cloud computing services in the 31 countries investigated in the study. Enterprises not using cloud computing services could then be omitted from analysis.

The characteristics of enterprises buying cloud computing services was also important in enabling the study to have a more granular approach. Eurostat data enabled the analysis of enterprises buying cloud computing services by a number of enterprise characteristics. These included:

- Country (EU27, EFTA and UK);
- Sector (NACE C to S);
- Enterprise size ('0 to 9', '10 to 49', '50 to 249' and 'over 250').

In addition, Eurostat provided information about the nature and intensity of cloud services used enterprises buying cloud computing services.

Using the historic Eurostat data from 2016 to 2020 analysis used extrapolation methods to provide forecasts about the characteristics of enterprises buying cloud computing services used over the internet until 2030.

The starting point for the analysis of enterprise data flowing to cloud and edge data centres is the analysis of Eurostat statistics.[293] In particular, the statistics about enterprises buying cloud computing services used over the internet.

These statistics provide robust information, collected by national statistics offices, about the use of cloud services[294] between 2014 and 2018 by:

- Country (EU27, EFTA and UK);
- Sector (public and private sectors – NACE C to N);
- Enterprise size (in bands of '1 to 9', '10 to 19', '20 to 49', '50 to 249' and '250 or more' employees);
- Nature of cloud services utilised (low, medium, and high level).[295]

293 Eurostat indicator isoc_cicce_use. Data was extracted on 3rd March2021. On 21 January 2021 Eurostat updated the indicator; providing statistics for 2020, but not for 2019. Our team extrapolated estimates for 2019.

294 The Eurostat definition is 'Buy cloud computing services used over the internet'.

295 Low services include email and office software. Medium include file storage and database hosting (alongside those included as low). High includes accounting applications, CRM, computing power (alongside those included as low and medium). It will be reasonable to assume that higher level, more intense users, will exchange greater volumes of data.

2.2. Belgian example of the methodology

Belgium is used to demonstrate the methodological approach for this stage of the methodology. Stage one analysis is founded on two key statistics for Belgium:

- In 2018, Belgium had 635,500 enterprises, employing 4.565 million people;
- 40 per cent of enterprises (254,200; found by 635,500 enterprises x 40 per cent) in Belgium bought cloud computing services that are used over the internet.

As noted previously, Eurostat provide data for cloud service adoption by NACE sector and enterprise size. Thus, the preceding analysis was repeated many times for different sectors and enterprise sizes in Belgium.

2.3. Outputs and conclusion

Eurostat statistics about enterprises buying cloud computing services used over the internet. Analysis provides insights about cloud service adoption by country, sector, and enterprise size. Linear extrapolation methods used growth rates observed in the countries studied from 2014 to 2020 to extrapolate growth in enterprises buying cloud computing services used over the internet to 2030.

Enterprise insights are important, but as the next section highlights, enterprises can be any size from one to over a million people. Standardisation methods are therefore required to better estimate the number of people employed in enterprises buying cloud computing services.

3. Stage 2: Standardisation to estimate the workforce employed in enterprises buying cloud services

3.1. Introduction

Eurostat data and study estimates to 2030 of enterprises buying cloud computing services used over the internet undertaken in stage one provided a foundation for the study. However, a problem with enterprise statistics is that enterprises can range in size from one employee to many thousands of employees.

If there are two enterprises buying cloud services used over the internet and one had one employee and the other had 10,000 employees one would expect the larger enterprise to have much higher levels of data flowing to cloud and edge data servers than the enterprise with one employee.

To address this problem, and better understand the volumes of data used by enterprises of different sizes, it was necessary in stage 2 to use standardisation techniques to estimate the number of employees in enterprises buying cloud services used over the internet. Analysis provided estimates by country and NACE sectors.

The next section describes the application of the methodology in Belgium.

The method was enhanced and made more accurate by repeating the Belgium analysis for each of the NACE sectors studied (C to S) for each enterprise size band in each country studies for every year for which data was provided by Eurostat (2016 to 2020) and for forecasts of future trends (from 2021 to 2030). In total, this stage of the analysis required almost 30,000 computations.

3.2. Belgian example of the methodology

Stage one analysis presented Eurostat data that revealed that in 2018 the 635,500 enterprises in Belgium employed 4.795 million people. Eurostat data also revealed that 40 per cent of enterprises in Belgium bought cloud computing services used over the internet.

A simple method of standardisation would be to suggest that if 40 per cent of enterprises in Belgium bought cloud computing services used over the internet it would not be unreasonable to assume that 40 per cent of employees would probably be employed in enterprises buying cloud computing services used over the internet. This would equate to 1,918,000 people employed in enterprises buying cloud computing services used over the internet.

The figure of 1,918,000 employees is useful overall estimate, but it can be further refined and made more robust.

Eurostat provide statistics for the percentage of enterprises buying cloud computing services used over internet for four different sizes (employees) of enterprises. Greater accuracy can be achieved if the preceding method is applied to the four different sizes of enterprise. Once again, Belgium is used as an example.

Eurostat provide enterprise data by size bands for employees[296]. For example, in 2018 in Belgium:

Enterprises with '0 to 9' persons employed: There are 602,300 enterprises in Belgium with '0 to 9' persons employed. In total they employ 1.60 million employees. Eurostat data for 2018 reveals that approximately 14 per cent of enterprises in this size band use cloud services[297]. Analysis uses the previous assumption that 14 per cent of the workforce are likely to be employed in enterprises buying cloud computing services used over the internet. This would equate to **225,150** people employed in enterprises buying cloud computing services used over the internet.
Enterprises with '10 to 49' persons employed: There are 26,600 enterprises with '10 to 49' persons employing 834,100 employees in Belgium in 2018. Eurostat estimate that 36 per cent of enterprises in this size band use cloud services. Using the previous assumption that 36 per cent of the workforce are likely to be employed in enterprises buying cloud computing services used over the internet. This would equate to **300,270** people employed in enterprises buying cloud computing services used over the internet.
Enterprises with '50 to 249' persons employed: There are 4,100 enterprises in Belgium with '50 to 249' persons employed. In total they employ 673,600 employees. Eurostat data for 2018 reveals that 55 per cent of enterprises in this size band use cloud services. Analysis uses the previous assumption that 55 per cent of the workforce are likely to be employed in enterprises buying cloud computing services used over the internet. This would equate to **370,480** people employed in enterprises buying cloud computing services used over the internet.
Enterprises with 'more than 250' persons employed: There are 4,100 enterprises with 'over 250' persons employing 1.67 million employees in Belgium in 2018. Eurostat estimate that 79 per cent of enterprises in this size band use cloud services. Using the previous assumption that 79 per cent of the workforce are likely to be employed in enterprises buying cloud computing services used over the internet. This would equate to **1,326,560** people employed in enterprises buying cloud computing services used over the internet.

The above approach is more refined than the previous broad-based national analysis which estimated 1,918,000 people employed in enterprises buying cloud computing services used over the internet. When one totals the four estimates (for company size bands) above the figures estimate 2,222,460 people employed in enterprises buying cloud computing services used over the internet. This estimate is 304,460 (13.6 per cent) higher than the broad-based national analysis.

A more detailed approach, examining enterprise size bands, provides better and usually higher estimates of people employed in enterprises buying cloud computing services used over the internet than broad-based national analysis. Greater accuracy arises because in most economies a relatively small proportion of enterprises (4,100; 0.6 per cent of all enterprises in Belgium in 2018) with more than 250 employees, which are high adopters of buying cloud computing services used over the internet (79 per cent in Belgium), employ a relatively large proportion of the workforce (35 per cent).

296 Eurostat has data for EU 27MS, EFTA (excluding Liechtenstein) and UK

297 Eurostat data for this indicator (0-9 employees) is sparse, thus this per cent has been calculated through thorough estimations

3.3. Outputs and conclusion

This stage of the study used a relatively simple standardisation approach (nearly 30,000 times) to estimate the number of employees in enterprises buying cloud services used over the internet in 31 countries, in 17 NACE Sectors, for four sizes (workforce) of enterprise from 2016 to 2030.

Understanding the size of the workforce in enterprises[298] utilising cloud services is important in stage 4. Workforce sizes for enterprises buying cloud services used over the internet are multiplied by the estimates of internet traffic per employee found by stage 3. This multiplication calculates internet traffic for enterprises in different sectors and of different sizes (employees). Stage 5 estimates of the volume of internet traffic that is flowing to cloud and edge data servers. These percentage figures can then be utilised to estimate enterprise data flowing to cloud and edge data servers in different countries, sectors and of different sizes (employees) of enterprise.

298 Including the country, sector, and size of business in which they work.

4. Stage 3: Estimating internet traffic per employee

4.1. Introduction

The third stage of the study focused on internet traffic generated by employees. ITU statistics from 2016 to 2020 were used to estimated internet traffic in the 31 countries investigated in the study. Analysis used ITU and Cisco internet traffic growth rates to forecast internet traffic until 2030.

The focus of this study is enterprise data flowing to cloud and edge data centres. Cisco statistics were used to find the percentage of internet traffic generated by enterprises.

Estimates of the enterprise internet traffic generated by enterprises in each country provided the basis for analysis to estimate average internet traffic (calculated as GB/month) per employee.

4.2. Belgian example of the methodology

Eurostat data shows that in 2018 the Belgian workforce was comprised of 4.795 million people.

Stage 3 analysis, described later in this methodological description, estimates that internet traffic in Belgium in 2018 was 6.7 EX (exabytes).

Cisco estimates that in 2018, fixed broadband devices used by enterprises comprised 24.0 per cent of internet traffic. The remaining traffic (75.2 per cent) was attributed to households. Using the Cisco figure of 24.0 per cent of internet traffic generated by enterprises in 2018, it is possible to estimate that enterprise internet traffic is 1.6 EX/year (139,581 TB/month) in 2020.

Dividing the size of the workforce (4.795 million) into Belgian enterprise internet traffic flow in 2018 (139,581 TB/month, 142.9m GB/month) provides an average monthly internet traffic flow of 29.8 GB per employee.

These workforce averages are used in stage 4 to estimate the proportion of average employee data that is flowing to main cloud and edge data centres in enterprises buying cloud services used over the internet.

This method of dividing enterprise generated internet traffic by the size of the workforce in a country was replicated across the 31 countries included in the study to provide the GB/month internet traffic per month per employee estimates for all countries.

4.3. Analysis, ITU statistics and forecasts for internet traffic

The main report noted that ITU provides annual information about internet traffic. This data is available for 14 of the EU27 Member States and Switzerland. To address the gaps, extrapolation was required to calculate internet traffic volume in countries where ITU data was not available.

It would have been easy for the countries where internet traffic volume is known to calculate the average amount of internet traffic used per person or enterprise and then extrapolate that figure across all other countries.

However, using this simple method of applying the average amount of internet traffic used per enterprise (from ITU data) to countries where internet traffic volumes were unknown produced spurious results. On one or two occasions internet traffic volumes (per enterprise) in less developed countries, with low internet use,[299] were well above levels observed in countries with higher GDP and a longer history of more intensive internet use. Since this method appeared unsuitable a more robust methodology was adopted.

To provide a robust methodology to calculate missing values for the 13 EU27 Member States with missing values, multiple regression methods were undertaken iteratively over several rounds of analysis, using machine learning methods.

The dependent variable adopted in analysis was annual internet traffic in EX (this corresponds with the ITU data).

Independent variables utilized in multiple regression analysis to explain variance in internet traffic included population, population density, GDP per capita, workforce size, exports, imports, and percentage of individuals using the internet, percentage of businesses using the internet, and cloud usage. The preceding statistics and many others were obtained from the OECD, World Bank, and World Economic Forum[300]. The machine learning methods ran many hundreds of multiple regression calculations using different combinations independent variables.

Multiple regression equations were undertaken iteratively, adding and changing independent variables until the highest explanation of variance (R2 value) could be obtained. Problems that might have arisen from multicollinearity were resolved by removing variables that were highly correlated.

Eventually, a robust equation was developed that achieved an R2 value of 0.82. A high R^2 value (generally considered to be above 0.75) indicates that independent variables in a model explain most of the variation in internet traffic flows (the dependent variable).

The final equation created by the third stage of analysis is a linear equation in the form:

$$EX = -0.8236 + 0.0000004683(Population) + 0.3725(\%\ individuals\ internet\ use) - 0.3138(\%\ of\ businesses\ internet\ use).$$

The high R^2 value (0.82) indicates that the equation provides a good predictive tool. It was used to estimate internet traffic for the 13 EU27 Member States with missing internet traffic values.

299 The number of business connected to the internet was found using DESI variables (https://ec.europa.eu/digital-single-market/en/digital-economy-and-society-index-desi).

300 The study team used their judgement about independent variables that might be likely to have a causal link with internet traffic flow. In total approximately 50 variables were included in analysis. Machine learning repeated analysis many times with different combinations of independent variables.

The main report (section 3.1.4.1) noted that both the ITU study and Cisco study forecast internet traffic growth of 22 per cent per annum[301]. Since there is agreement between the two sources this figure (22 per cent) was used in the analysis in the remainder of this section to forecast internet traffic growth from 2020 to 2030.

The backcast multiplicand to estimate data flows in previous years is “0.180327”. This provides values that produce a 22 per cent increase in later years.

Cisco estimates that in 2020, fixed broadband devices used by enterprises comprised 24.8 per cent of internet traffic. The remaining traffic (75.2 per cent) was attributed to households. Cisco do not provide a source or methodology for this estimate, nor could any other sources be found. Using the Cisco figure of 24.8 per cent of internet traffic generated by enterprises in 2020 it is possible to estimate that enterprise internet traffic is 75.7 EX/year in 2020[302].

Cisco predict the proportion of total internet traffic that is used by enterprises is steadily increasing. In 2023 they predict 26 per cent of traffic will be generated by enterprises[303]. If this rate of increase in internet traffic generated by enterprises (24.8 per cent in 2020 and 26 per cent in 2023) is sustained, the study team forecast that 26.9 per cent of internet traffic will be generated by enterprise in 2025 and 29.2 per cent in 2030.

In 2025, the study team have forecasted that 26.9 per cent of internet traffic can be attributed to enterprises, thus enterprise internet traffic in 2025 is estimated to be 222 EX/year. In 2030, the study team estimate that there will be 2,229 EX/year of total internet traffic and 29.2 per cent will be attributed to enterprises, thus enterprise internet traffic is 651EX/year (2,229 EX/year x 29.2 per cent)[304].

4.4. Outputs and conclusion

Stage 4 analysis utilised, enhanced, and validated ITU data and undertook analysis using machine learning methods and multiple regression methods to estimate internet traffic for each country examined in the study.

Both the ITU and Cisco forecast internet traffic growth of 22 per cent per annum. Since there was agreement between the two sources this figure (22 per cent) was used in the analysis in the remainder of the stage to forecast internet traffic growth from 2020 to 2030.

301 Cisco forecast growth in data consumption of 22 per cent per annum to 2022. The study team extend this estimate to 2030 for forecasting purposes (Cisco visual networking index: Forecasts and trends 2017 to 2022. https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-741490.html and Cisco slides (particularly slide 8) can be found at https://www.cisco.com/c/dam/m/en_us/network-intelligence/service-provider/digital-transformation/knowledge-network-webinars/pdfs/1213-business-services-ckn.pdf

302 Total internet traffic in 2020 is 305 EX/year. If this figure is multiplied by the percentage of internet traffic generated by enterprise (24.8 per cent) on can estimate that enterprise generated internet traffic in 2020 is 75.7 EX/year.

303 Cisco annual internet report page 6 (https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.pdf). Cisco forecast that in 2023 consumer share of devices will be 74 per cent. Consumer growth is expected to grow at 9.1 per cent CAGR and business growth at 12 per cent. Using the Cisco estimate for 2023, our team extrapolated ‘backwards’ an estimate of 24 per cent for the business share in 2018.

304 EFTA: 2020 total internet traffic is 17.2 EX/year, enterprise internet traffic is 4 EX/year (17.2 x 24.8 per cent). 2025 total internet traffic is 47.2 EX/year, enterprise traffic is 13 EX/year (47.2 x 26.9 per cent). 2030 total internet traffic is 128.3 EX/year, enterprise internet traffic is 38 EX/year (128.3 x 29.2 per cent). UK: 2020 total internet traffic is 65.8 EX/year, enterprise internet traffic is 16 EX/year (65.8 x 24.8 per cent). 2025 total internet traffic is 177.8 EX/year, enterprise internet traffic is 48 EX/year (177.8 x 26.9 per cent). 2030 total internet traffic is 480.7 EX/year, enterprise internet traffic is 140 EX/year (480.7 x 29.2 per cent).

Estimates provided by Cisco concerning the proportion of internet traffic generated by enterprise were used to estimate the enterprise internet traffic from 2018 to 2030.

The final stage of analysis estimated internet traffic per employee by dividing internet traffic generated by enterprise in each country by the number of employees in countries.

These estimates provide the starting point for stage 4 analysis which examines the proportion of total enterprise internet traffic that might be flowing to cloud data centres.

5. Stage 4: Enterprise data flowing to cloud and edge data centres

5.1. Introduction

Stage 4 utilises information from the two preceding stages. Stage 2 provided standardised data about the number of employees in enterprises buying cloud computing services used over the internet. This information was provided for NACE sectors and different sizes of enterprise (employees). Enterprise internet traffic information calculated in Stage 3, particularly the estimates of average internet traffic (calculated as GB/month) per employee, was utilised in stage 4.

In stage 4, the number of employees in a sector or enterprises of a particular size (stage 2 analysis) was multiplied by average internet traffic per employee (stage 3 analysis). This analysis made it possible to estimate total enterprise internet traffic for the sector or enterprises of a particular size.

The final element of stage 4 analysis estimated the percentage of internet traffic for a sector or enterprise of a particular size that was flowing to cloud service providers.

Enterprise router information, for enterprises buying cloud computing services used over the internet, was analysed to find the average percentage of enterprise internet traffic flowing to cloud service providers.

5.2. Belgian example of the methodology

Stage 1 analysis estimated the number of people working in enterprises buying cloud services used over the internet; this was 2,896,000 people in Belgium in 2020.

In Belgium, in 2020, the average amount of internet traffic generated by Belgian employees was 45.5 GB/month. This figure was used as a multiplicand with the 2,896,000 people working in enterprises buying cloud services used over the internet to estimate the amount of internet traffic that would be used by these enterprises. Analysis estimated the figure 142,800 TB/month of data (1.63 EX/year found by multiplying 45.5 GB/month by 2,896,000).

Analysis described in the next section estimated that 17.2 per cent of this Belgian internet traffic will flow to cloud and edge data centres. In 2020, it was estimated that 24,560 TB/month (0.28 EX/year found by multiplying 142,800 TB/month by 17.2 per cent) of Belgian enterprise data is flowing to cloud and edge data centres from enterprises buying cloud computing services used over the internet.

5.3. Internet traffic insights from routers

As noted previously, a key objective of the study was to find 'an estimation of the proportion of enterprise internet traffic that flowed to cloud and edge data centres from enterprises buying cloud computing services used over the internet'.

Our study has developed an innovative new approach to estimating data flows to cloud and edge data centres by collecting information from internet routers.[305] No insights have been found in other studies; therefore our study brings new insight into discussion.

Internet traffic monitoring software is included in many enterprise routers. The software enables users to check connectivity, examine traffic levels and plan future capacity, examine network outages, and understand security issues. In this study, router information was used to find the volumes of data flowing to cloud services providers.

Figure 28 provides a screenshot of data that can be obtained from a Cisco router. Access to the IP addresses of all websites contacted from a router is valuable for examining internet traffic destinations and origins. Through IP look-up tables, it is generally possible to obtain the:

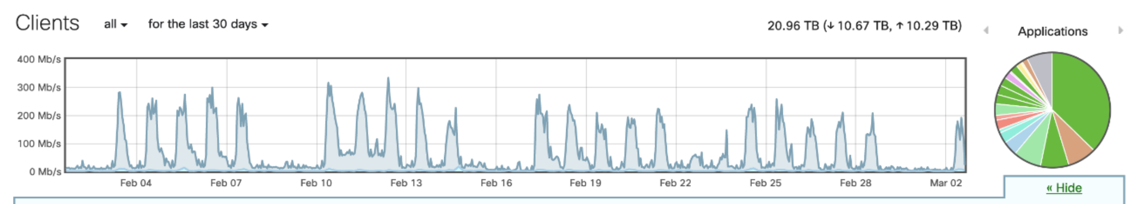
- latitude and longitude of the IP address;
- country of the IP address (generally with 99 per cent accuracy);[306]
- region/state of the IP address (90 per cent accuracy);
- city of the IP address (81 per cent accuracy).

Figure 28 shows the names and applications for many IP addresses that are provided by a router monitoring software. For example, items 8 (SharePoint, 885.9 GB), 13 (DropBox, 523.5 GB) and 17 (Windows Office 365, 885.9 GB) in Figure 28 are all cloud services. Router information therefore provides a useful method of monitoring the amount of internet traffic flowing through a router to a cloud data or edge centre. Thus, router management provides an innovative method for examining enterprise data flowing to cloud and edge data centres.

305 As its name implies, a router 'routes' traffic between the device in an enterprise or household and the internet.

306 https://whatismyipaddress.com/ip-lookup

Figure 28: An example of management information provided by an enterprise router



Applications details

#	Description	Group	Usage	% Usage	Group usage	Group % usage
1	Miscellaneous secure web	–	7.92 TB	37.3%	7.92 TB	37.3%
2	UDP	–	1.70 TB	8.0%	1.70 TB	8.0%
3	Miscellaneous secure web - 212.250.159.189	–	1.68 TB	7.9%	1.68 TB	7.9%
4	Miscellaneous web	–	1.67 TB	7.9%	1.67 TB	7.9%
5	Non-web TCP	–	887.61 GB	4.1%	887.61 GB	4.1%
6	CDNs	–	714.47 GB	3.3%	714.47 GB	3.3%
7	YouTube	Video	668.25 GB	3.1%	886.47 GB	4.1%
8	Sharepoint	Productivity	592.78 GB	2.7%	885.92 GB	4.1%
9	Miscellaneous secure web - 212.250.159.139	–	583.99 GB	2.7%	583.99 GB	2.7%
10	Miscellaneous secure web - 192.168.150.2	–	580.01 GB	2.7%	580.01 GB	2.7%
11	Miscellaneous secure web - 192.168.144.248	–	543.24 GB	2.5%	543.24 GB	2.5%
12	Miscellaneous secure web - github.com	–	490.38 GB	2.3%	490.38 GB	2.3%
13	Dropbox	File sharing	430.35 GB	2.0%	523.53 GB	2.4%
14	Google HTTPS	–	360.74 GB	1.7%	360.74 GB	1.7%
15	Host-based email (POP3/IMAP/SMTP)	Email	317.19 GB	1.5%	426.18 GB	2.0%
16	microsoft.com	–	290.94 GB	1.3%	290.94 GB	1.3%
17	Windows Office365	Productivity	272.95 GB	1.3%	885.92 GB	4.1%
18	SSH	–	263.04 GB	1.2%	263.04 GB	1.2%
19	Netflix	Video	120.42 GB	0.6%	886.47 GB	4.1%
20	Software updates	Software & anti-virus updates	96.68 GB	0.4%	113.29 GB	0.5%
21	Spotify	Music	95.80 GB	0.4%	178.30 GB	0.8%
22	Facebook	Social web	95.74 GB	0.4%	207.35 GB	1.0%
23	iTunes	Music	76.03 GB	0.3%	178.30 GB	0.8%

The information provided in Figure 28 was obtained in February 2020 from the Dock Innovation Centre in Leicester, UK; it is comprised of 55 small businesses.[307] Dock uses Cisco routers which provide details (in aggregate anonymously) of applications and websites used. The Centre recorded 7.90 TB of internet traffic in February 2020. 1.29 TB of this internet traffic was enterprise data flowing to cloud data and edge centres.[308] This means that 17.2 per cent of internet traffic flowed to cloud and edge data centres.

The study team also contacted innovation centres and cluster workplaces between September and December 2020 in the 31 countries being examined to seek similar router information. The response rate was poor. Many replied that due to lockdowns they and many of their tenants were working from home and results would not be representative of the pre-Covid-19 situation examined previously. Fourteen responses were received and seven had Cisco routers[309]. During discussions, respondents again stressed how working conditions were very different from normal. Despite assurances of anonymity and technical support to navigate the router dashboards, only three further responses were received. The average level of internet traffic flowing to cloud and edge data centres from innovation centres and cluster workplaces was 17.6 per cent.

In addition to router analysis, the micro-level analysis questionnaire also included questions about yearly data flow and the amount of such to cloud and edge data centres. 61 enterprises provided relevant information about these questions. The average level of internet traffic flowing to cloud and edge data centres was 16.6 per cent.

307 Dock Innovation Centre in Leicester, UK (https://www.dockleicester.co.uk/)

308 Relevant cloud apps listed in router output include SharePoint, DropBox, Windows Office 365, Microsoft OneDrive, iCloud, Google drive etc.)

309 The remaining seven used a variety of other routers.

The overall average across the three methods was 17.2 per cent; this figure was therefore used in analysis. We believe that in the future router analysis should be pursued. With time and investment, it would be possible to create remote access to sufficient routers in order to offer considerable insights into internet traffic and data flowing to cloud service providers on a regular basis.

Applying the figure of 17.2 per cent of data flowing to cloud among enterprises utilising cloud data services, it is possible to forecast data flows between 2018 and 2030. The methodology is relatively simple. The Belgian example was used in the previous section to provide an example of the methodology.

5.4. Outputs and conclusion

This penultimate stage of the analysis developed an innovative method of estimating the overall volume of enterprise data flowing to cloud infrastructures. Insights, from router management information and micro-level questionnaires, about the proportion of enterprise internet traffic flowing to cloud and edge data centres was used to estimate the proportion of enterprise internet traffic that was flowing to cloud and edge data centres from enterprises buying cloud computing services used over the internet. This average figure of 17.2 per cent was applied to the estimate average internet traffic per employee (from stage 3) for the number of employees in enterprises buying cloud services used over the internet (from stage 2). The average (17.2 per cent) was then applied to information about enterprise internet traffic from 2018 to 2030.

6. Stage 5. Cloud flow geography

6.1. Introduction

This final stage of the methodology examines the geographical flows of enterprise data flowing to cloud and edge data centres by enterprises buying cloud computing services used over the internet.

The enterprise cloud flow geography utilises information from stage 4 about the amount of enterprise data flowing to cloud and edge data centres by enterprises buying cloud computing services used over the internet in each of the 31 countries analysed.

Analysis assumes that the amount of enterprise data flowing to cloud and edge data centres will be proportional to the market share of the main cloud service providers. In each country estimates were made of the volume of data flowing to cloud and edge data centres of the ten main cloud services providers.

The locations of main cloud and edge data centres were identified in section 2.6. For each of the 31 countries, analysis identified the nearest main cloud and edge data centre for the ten main cloud services providers. Analysis assumed that 80 per cent of enterprise data flowed to main cloud data centres and 20 per cent flowed to edge data centres. On the basis of these assumptions, it was then possible to model geographical cloud data flows and examine the volume of enterprise data flowing to different countries. This analysis estimated enterprise data flowing to cloud and edge data centres in countries that possessed cloud infrastructure.

Analysis thus identified two elements:

1. Enterprise cloud data flowing to cloud and edge data centres from a country.
2. Enterprise cloud data flowing to the cloud and edge data centres in countries which possess cloud infrastructure.

Having identified enterprise cloud data generated for the 31 countries included in the analysis and the volume of enterprise cloud data flowing to the cloud and edge data centres in the 19 countries that possess cloud infrastructure, it was possible to calculate net cloud data flows for each of the 31 countries.

6.2. Belgian example of the methodology

Previous analysis in stage 4 found that Belgium had 24,560 TB/month (0.28 EX/year) flowing to cloud servers. Table 26 below is replicated from section 3.1.6 of the main report. In the second column it shows the proportion of data flowing to each of the anonymised cloud service providers.

Table 26 Share of Belgium enterprise data flowing to different cloud service providers in 2020 and the locations of the nearest provider main cloud and edge facilities

Cloud service provider	Market share	% of total country cloud data flow (24,560 TB/month)	Nearest main cloud data centre for 80% of flow	Nearest edge facility for 20% of flow
Cloud Provider 2	33%	8,100	Germany	Netherlands
Cloud Provider 5	17%	4,170	Netherlands	Netherlands
Cloud Provider 3	8%	1,960	Belgium	Netherlands
Cloud Provider 1	5%	1,220	Germany	UK
Cloud Provider 4	5%	1,220	Germany	Germany
Cloud Provider 6	3%	737	Germany	Netherlands
Cloud Provider 7	1%	245	France	France
Cloud Provider 8	1%	245	Germany	Germany
Others	27%	6,630	-	-

Thus, 33 per cent of total Belgium enterprise cloud data (24,560 TB/month) is expected to flow to Cloud Provider 2 (CP2); this equates to 8,100 TB/month. This method is replicated for the proportion of enterprise data flowing to the other cloud service providers.

As noted previously, 80 per cent of the amount of data flowing to CP2 (6,480 TB/month, 0.07 EX/year) is allocated to the nearest CP2 main data server. As shown in the table, Germany is the nearest main CP2 data centre to Belgium. See column four for the nearest main data centres for each of the cloud service providers. The rightmost column lists the locations of the nearest edge server for each of the providers. The nearest edge provider to Belgium for CP2 is in the Netherlands. Thus, the main German cloud data centre of CP2 will receive approximately 6,480 TB/month (80 per cent of the Belgian enterprise data flowing to CP2 - 8,100 TB/month multiplied by 80 per cent). The remaining 20 per cent (1,620 TB/month, 0.02 EX/year) is allocated to CP2 edge servers in the Netherlands.

This process is replicated for the other seven cloud services providers for Belgium. Then, the overall approach was replicated across all 31 countries for the five years from 2016 to 2021.[310]

Having replicated the methodology explained in the Belgian example for all 31 countries included in the analysis, it was possible to examine the amount of data flowing from each country studied to the main cloud and edge data centres of the eight anonymised cloud service providers.

This data and an overview of results was provided earlier in the study. It is replicated here so that it sits alongside (and can be more easily compared with) information about enterprise data flowing into the cloud and edge data centres in the 31 countries studied.

The final stage of analysis in stage 5 utilised information about the locations of the 26 main cloud data centres and 73 edge data servers to examine the origin and destination of the 75,527 TB/day of data flowing to cloud and edge data centres will flow across the EU27 Member States, the UK and EFTA.

310 Due to the huge growth in edge data centres, analysis beyond 2021 would provide erroneous results.

For the eight main cloud service providers, the nearest cloud and edge data centres to each of the 31 countries investigated were identified. The cloud and edge data flows from each country were allocated to these[311]. This enabled insight about the destination and origin of where the enterprise cloud and edge data flows that were identified in Stage 4 were flowing to.

Having identified enterprise cloud data generated Belgium and the volume of enterprise cloud data flowing from Belgium to the cloud and edge data centres in the 19 countries that possess cloud infrastructure it was possible to calculate net cloud data flows.

Net data flow for 2020 was found by subtracting 23,486 TB/month data inflows to cloud and edge data centres located in Belgium from the 24,567 TB/month cloud data outflows from Belgium.

Net data flow in Belgium in 2020 is 1,080 TB/month (0.01 EX/year).

6.3. Outputs and conclusion

Stage 5 included innovative analysis for the destination of cloud data flows by allocating the volume of data flows to cloud and edge service providers' infrastructures according to their market shares across the EU27 Member States, EFTA countries, and the UK. Analysis then investigated the locations of the nearest main cloud and edge date centre for each cloud service provider in all 31 countries. Then, the appropriate market share cloud data flow volumes were allocated to the relevant countries.

After completing the analysis, it was possible to examine how much data was flowing to cloud and edge data centres in each country that possessed cloud and edge infrastructure in the 31 countries examined in the study.

7. Conclusions

This annex has provided a detailed overview of the methodology developed in research to meet the key objective for this study which was 'to develop and test a new, self-sustained and replicable methodology to estimate data flows'.

7.1. Updating the model

Analysis undertaken using the methodology described above is underpinned by Eurostat data with important data also provided by the ITU and Cisco. When these three sources update their statistics, the model can be updated.

Eurostat data about enterprises buying cloud computing services used over the internet was last provided in January 2021 with data provided up to 2020. Statistics for 2019 are missing. Eurostat statistics for enterprises buying cloud computing services used over the

311 The allocation of cloud data flows for each cloud service provider could include cloud data flows being served by a cloud or edge data centre located in the country where the data flow originated. For example, as Table 26 identifies the 20 per cent cloud data flowing to an edge data centre for cloud service provider 3 will flow to cloud service provider 3's edge data centre located in Belgium.

internet in odd years (2015, 2017 and 2019) is patchy. The next update might be in January 2022 with data for 2021, however the data for 2021 could be sparse. If 2021 data is not collected or it is too unreliable the next update is likely to be in January 2023.

ITU provides data (to subscribers who pay for their statistics) for fixed broadband traffic (exabytes (EX) per annum). The metadata states that the data relates to traffic generated by fixed-broadband subscribers measured at the end user access point. 2019 is the latest year for which ITU provide data, this data was released in January 2021. The next update is likely to be in January 2022.

Annex 4 Data tables

Annex 4 lists the Excel files summarising the model used for the mapping of data flows described in the previous Annex. They are sent as separate documents.

1. EC Ent_emps_bands aggregated
2. EC Data flow MASTER
3. EC Low Med High Aggregates Linear and S-shaped
4. EC cloud data flows

Annex 5 Detailed methodology for collecting primary data on the dimensions of data flows: Actors and data flows value chain

Annex 5 presents the detailed methodology used to collect primary data on the typology of data flows.

In order to get information on the dimensions that need to be further detailed, a micro level approach was designed. As shown by the figure below, the **interview programme** targeted cloud service providers and infrastructure network providers, by selecting the key players in terms of market share in Europe. The interview questionnaire was built to get qualitative insights on the dimensions and the data stock. It was distributed to key cloud service and network infrastructure providers. In the course of the study, cloud service providers were periodically contacted to provide additional insights in view of the study findings and development.

Due to the characteristic of the markets, it was estimated that interviewing 25 companies was possible to get above 75 per cent of coverage of the EU market from cloud service companies and infrastructure network providers' perspective. Considering the market share in Europe of cloud service companies and infrastructure network providers, the study team has selected the main players in the European market. According to the available statistics on cloud services providers and infrastructures network providers in Europe, 25 companies were identified that represent approximately 75 per cent of the market in terms of cloud services offered and mobile and fixed network provided.

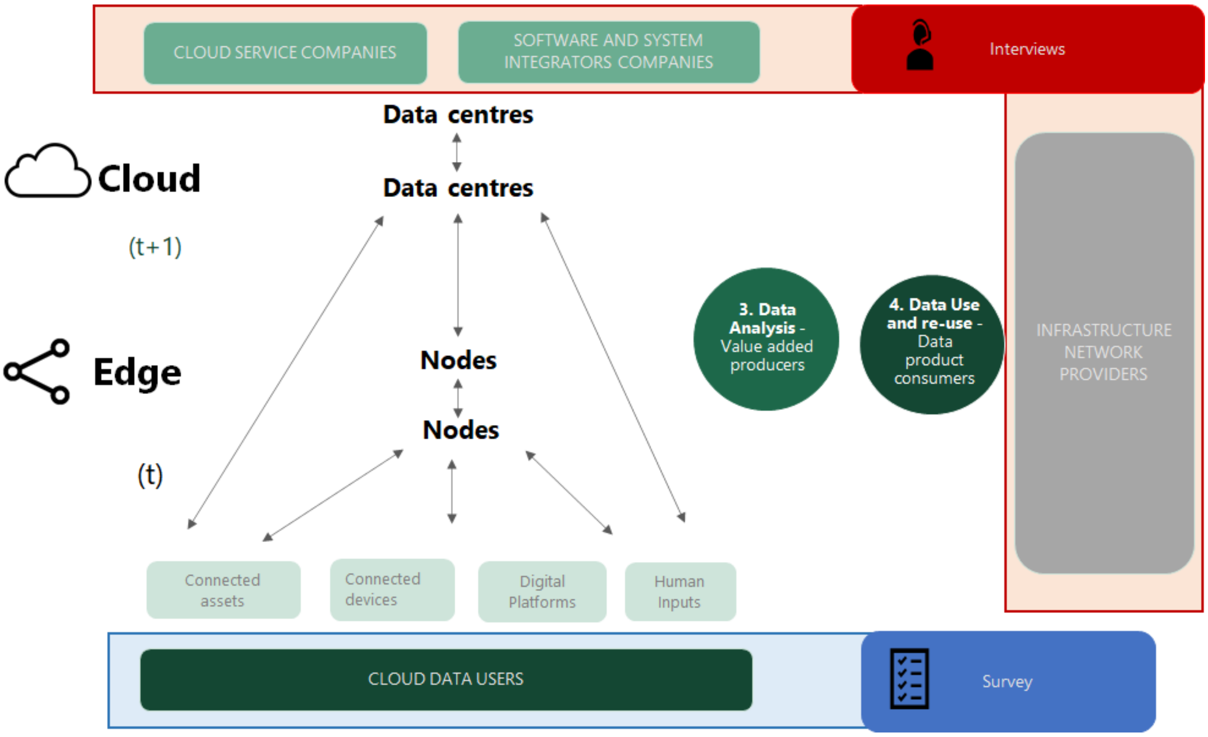
The study team reached out to 30 cloud service and network infrastructure providers and only 11 agreed to answer to some of the questions.[312] The discussion focused on the relevant dimensions from a service provider point of view: namely dimension 2 (enabling technology of the flows), dimension 5 (services offered), dimension 3 (actors) and dimensions 6 (sectors). Moreover, questions related to capacity of cloud data centers and data stock were also asked in a second round of interviews. The interview guide is available in Annex 7.

One of the key obstacles during the course of the project has been the responsiveness of cloud users and their willingness to share information. Stakeholder engagement activities have often not been able to obtain the depth of information required because cloud service provider representatives have suggested that detailed information about their cloud services and infrastructure is 'commercially confidential'. Moreover, with the few answers received, it was difficult to make robust assumptions due to very different response from the interviews. The lessons learnt of this exercise are detailed in the 'Lessons learnt' section below.

While, the objective of the **survey** was to target primarily **cloud data users**: enterprises from the different sectors targeted in the study and public entities (including public administrations and academia) departing from the EU survey developed by the Commission. It complemented the input from the interviews on the different dimensions of cloud-based data flows of the typology, focusing on the perspective of the users.

312 Amazon, Equinix, Google, IBM, Oracle, Salesforce, Swisscom, Telefonica, TIM, T-system, VMware.

Figure 29: Key cloud actors and data collection tools



The survey questionnaire designed by the EC was kept as a basis for the study survey, and questions were fine-tuned to reflect the data needs of the study. Once the new questionnaire approved by the EC, the survey was re-uploaded on the EU survey tool on 5 October 2020. Potential respondents were encouraged to participate before 6 November 2020. The survey was finally closed on 27 November 2020.

There were no resources to ensure a representative sample through one data collection exercise. The study team opted to design a second survey, to build on and to complement the questionnaire from the EC. An external survey provider was contacted to collect an additional 110 responses.

A sample of 10 countries was selected, comprising:

- 8 countries with highest share of companies using cloud computing according to Eurostat: Finland, Sweden, Denmark, Netherlands, Ireland, UK, Belgium, Estonia.[313]
- The two most populated EU countries: Germany and France.

The objective was to focus on the countries for which data could be extrapolated, should the number of responses received through the EU survey tool was insufficient. This was done as the study team was encountering difficulties to reach out to stakeholders in the interview programme. The survey questionnaire is accessible in Annex 8.

Survey published on the EU survey tool:

On 27 November, 68 responses had been received. This comprised the 44 responses received in the previous iteration of the survey in the summer 2019. The study team

313 Excluding Malta because of the small size of the country in terms of population. But they are still important in terms of cloud usage so I do not know if the population criteria is justified enough to exclude them on this ground.

identified two duplicate answers: one respondent replied before and after the survey re-upload; another respondent one replied twice after the survey re-upload. The latest answers were kept after checking their consistency. The total number of answers is therefore 66.

The following measures were taken to avoid duplicates:

- A disclaimer was added in the email invitations.
- Manual check for the respondents that had accepted to disclose their details.

Survey carried out by external provider:

On 27 November, 110 answers had been received. Two answers were removed from the sample as they were not identified as data intensive: a dog breeding professional and a hairstylist. The total number of answers is therefore 108.

The following measures were taken to avoid duplicates with the EU survey:

- A disclaimer was added in the external provider's survey introduction page.
- A screening was introduced to filter respondents of the survey. Companies that were mapped by the study team to fill the questionnaire on the EU survey tool were automatically screened out and could not answer to the external provider's survey, to avoid that the same organisation answers to the questionnaire on both channels.

Survey merging and data cleaning:

After the answers were received, the results of the two surveys were merged together to reach a total of 174 answers.

A first data cleaning was done as some questions were formulated differently in the two questionnaires. This was due to the difference between the survey tools used by the European Commission (EU survey) and the external providers.

The box below shows the questions concerned and the actions taken:

- **Question 9.b.:** The percentage between cloud and edge infrastructure storage. In the EU survey questionnaire, there was the possibility to only give one figure. The assumption was taken that the figure written in the EU survey was the percentage given for cloud, the rest for edge (done for respondent 20060, 20062, 20064, 20066, 20068). This question has been used to provide some estimations on the enabling technology dimension (see section 2.1).
- **Question 13:** The deployment mode used to store data in the cloud. The EU survey allowed to specify the 5 main countries in which the data are physically stored, but could give an estimate of TB stored only for one type of deployment mode, i.e. public or private. In the cases where the entry 'both' deployment modes was selected, the number of TB were added for the countries quoted twice (for public and private). This question has not been used in the analysis as the macro-level approach was providing a robust-enough information.

The data was then cleaned to maximise its reliability. This was mainly done by removing outlier answers. The data cleaning actions for each question is detailed in the survey analysis below.

Some reliability issues were identified in the survey. This could be attributed to the respondents' lack of knowledge about their cloud usage (generation), usage, and the details of their contracts. For instance, Estonia has no cloud infrastructure, but all five Estonian respondents responded that their cloud data is stored on main cloud or edge centres in Estonia (Q.8). Some of the answers in the total volumes of data generated (Q.9.a.) appear as unrealistically high when compared with the estimates produced in the context of this study.[314] The same could be said about all other estimations of volumes i.e. Q.9, Q.9.a, Q.10, Q.11, Q.12, Q.12.a., Q.13, Q.14. On cloud deployment (Q.13), the vast majority of respondents have answered to use a private cloud deployment mode, which is different to Eurostat findings (EU27 58 per cent public, 42 per cent private).

Another reason for the lack of reliability of the answers is the questionnaire itself. In particular, answers to open questions were particularly difficult to analyse as too diverse (see for instance Q.12 on the applications used - answers range from 'Google' which is a cloud service provider to 'CRM' which is a functional area).

Tech4i2 identified a smaller sample of 61 respondents as being more reliable for analysis of data generation/storage volumes by deleting the outlier answers in Question 9. This smaller sample of respondents was taken to analyse the questions linked to data generation and storage. i.e. Q.9, Q.10, Q.11, Q.13, Q.14.

In addition, some survey answers can be interpretated in a qualitative way (e.g. Q.17 - Reason to move the data from one cloud infrastructure to another) and be put into perspective with results from the literature.

Limitation and robustness of the findings

One of the key obstacles during the course of the project has been the responsiveness of cloud users and their willingness to share information. Some reliability issues were identified also in the survey. This could be attributed to the respondents' lack of knowledge about their cloud usage (generation), usage, and the details of their contracts. Another reason for the lack of reliability of the answers is the questionnaire itself. In particular, answers to open questions were particularly difficult to analyse as too diverse.

Lessons learnt

The usefulness of a micro-level approach based on stakeholder data collection to complement the findings of the macro-level approach predominantly using secondary sources remains valid. For example, in the study, the macro-level analysis found that 12,620 TB/day of enterprise data was flowing to cloud servers. The survey found that on average 27 per cent of survey respondents used Software as a Service (SaaS) (Dimension 5). If survey respondents were representative of all EU-27 Member State enterprises, this could suggest that 3,400 TB/day of enterprise data flowing to cloud servers is associated with SaaS (12,620 x 27 per cent). In this way, the more general macro-level insights can be further disaggregated by micro-level insights provided from the survey and provide data at country-level in future exercises.

Concerning **interviews**, in order to maximise the benefits, the following actions could be done in the future:

- Reach out to cloud service and infrastructure providers only to validate study assumptions, methodology, and getting feedback on proxies but not rely on them to

314 Average internet traffic is 25.3 GB/month (cloud estimated to be 4.3 GB/month) = 0.2964 TB/year (cloud 0.05039 TB/year).

give information on precise information to populate the model (e.g. storage capacity, storage usage, pricing, services used by their customers, etc.). The latter information is considered as too sensitive and respondents were reluctant to share it.

- Target high-level employers in companies (managing directors) or team leads working in the privacy, market or infrastructure units.
- Focus on the service provider perspective, hence avoid questions related to data type (personal or non-personal), type of services often used, functional areas, due to the fact that providers do not have overview on these dimensions. Also, for services, it is often the case that service providers offer bundle of services, hence it is difficult for them to get estimation on which service is mostly used by their customers.

In order to make the **survey** exercise more useful, the following points could be improved:

- Increase the number of respondents (at least 2,400) and build the sampling approach with a professional survey provider that can guarantee reliable responses through a more restrictive panel.
- The sample could be extended in order to cover more exhaustively public administrations and academia.
- Limit the number of questions to a minimum, focusing on the aspects that cannot be grasped by the macro-level approach. The macro-level approach has shown success to provide reliable estimations of Dimensions 3 - Actors (Enterprise data to cloud services), Dimension 6 - Sectors and Dimension 7 - Geography. The survey should therefore focus on the remaining dimensions, omitting those aspects that are difficult to grasp for cloud users (e.g. the location of their cloud data centres).
- Use only or mostly close-ended questions. For instance, in Q.12, give a list of 10 the most used applications and let the respondent choose between them. An open text can be added if the respondent answers ‘Other’.
- The survey target should still be cloud users, however the questions should focus gathering estimation on proportions rather than volumes of data used. It is easier for cloud users to work on proportions rather than giving estimations on the volumes of data.

Annex 6 List of workshops participants

Workshop on data typology - 6 February 2020

The expert workshop held on 6 February 2020 aimed at gathering feedback on the typology of data and flows. The workshop was attended by stakeholders from academia, think tanks, industry representatives and independent experts on data matters and it focused on different typology of cloud-based data and data flows.

Expert Webinar on measuring data flows - 17 March 2020

An expert webinar held on 17 March 2020 aimed at discussing possible indicators and proxies to measure cloud-based data stock and flows, given that existing row data are limited.

Annex 7: Interview guide to Cloud service providers and list of organisations interviewed

The objective of the **interviews** was to target primarily cloud service providers and infrastructure networks. The aim was to get feedback on the methodology and insights on the services provided by cloud providers in Europe as well as volume of data stored in cloud infrastructures and cloud capacity.

Interview guide:

A. Entity

1. The size of the branch of your entity working on cloud is:
   - ☐ Micro (1-9 employees)
   - ☐ Small (10-49 employees)
   - ☐ Medium (50-249 employees)
   - ☐ Large (250 or more employees)

2. Where your entity is headquartered in:

3. Could you specify which countries your entity operate in Europe and the size of your business in those countries?

Country	N. employee	Annual turnover

B. Enabling Technology

4. Where are your data centres located in Europe?
   Which is the maximum storage capacity of each of your cloud Infrastructure Data centres
   Which is the actual storage of each of your cloud Infrastructure Data centres?

Data centre location	Max. storage capacity	Actual storage

5. Where are your edge data centers located in Europe?:
   Which is the maximum storage capacity of each of your edge Data centres
   Which is the actual storage of each of your edge Data centres?

Edge data center location	Max. Storage capacity	Actual storage

C. Services and Costumers

6. What is the % of new customers you have every years?
7. What is the average time a customer remains a customer of yours?
8. What is the average % of client ending their cloud contract with your company per year?

9. Do you sell edge services to customers?
   - If yes how many of them leave you / how many new are coming?
10. Do you sell edge services outside the countries where you have physical data centers?
    - If yes, do you collaborate with other cloud providers, or telecoms in other countries to store the data of you customers in their edge data centres?
11. Do you foreseen to invest in edge nodes in the coming years?
    - if yes how many by when and where would they be physically located?

12. Could you specify the percentage of the service you provide the most?
    - SaaS :
    - PaaS:
    - IaaS:
    - Edge Services:
    - Added value cloud services: AI, blockchain, big data analytic

13. Do you have estimation on the percentage of data stored in your cloud data centres for back-up needs?
14. Do you have estimation on the percentage of data stored in your cloud data centres for replicability needs?

15. Could you specify which are the types of typical sectorial customer you have:
    - ☐ Manufacturing (NACE C)
    - ☐ Electricity, Gas (NACE D)
    - ☐ Water supply, waste management (NACE E)
    - ☐ Construction (NACE F)
    - ☐ Wholesale and retail trade (NACE G)
    - ☐ Transportation and storage (NACE H)
    - ☐ Accommodation and food service activities (NACE I)

- Information and communication (NACE J)
    - Motion picture, video and television programme production, sound recording and music publishing activities
    - Programming and broadcasting activities
    - Telecommunications
    - Computer programming, consultancy and related activities
        - Computer programming activities
        - Computer consultancy activities
        - Computer facilities management activities
        - Other information technology and computer service activities
    - Information service activities
        - Data processing, hosting and related activities; web portals
            - Data processing, hosting and related activities
            - Web portals
            - Other information service activities
- Financial and Insurance activities (NACE K)
- Real Estate Activities (NACE L)
- Professional, scientific and technical activities (NACE M)

- Administrative and support service activities (NACE N)
- Administration and Defence (O);
    - Public administration and defense
        - Administration of the State and the economic and social policy of the community
            - General public administration activities
            - Regulation of the activities of providing health care, education, cultural services and other social services, excluding social security
            - Regulation of and contribution to more efficient operation of businesses
        - Provision of services to the community as a whole
            - Foreign affair
            - Defense activities
            - Justice and judicial activities
        - Compulsory social security
- Human Health and Social Work Activities (Q),

Annex 8: Survey questionnaire

As mentioned in the report, the objective of the **survey** was to target primarily cloud data users: enterprises from the different sectors targeted in the study and public entities (including public administrations and academia) departing from a first survey developed by the European Commission as part of the European Data Flow Monitoring Initiative[315] and published on the EU survey tool in the summer 2019. The aim was to focus on the perspective of the users and provide insights on the dimensions that could not be captured by the macro-level approach (see full approach in section 3.3 below). The survey gathered 174 responses.

The main findings that were used for the study estimations are:

- **Data types:** the survey estimates that the total data stored in cloud infrastructure is 41% personal data and 59% non-personal data.
- **Enabling technology:** the survey findings highlight that in 2020, the percentage between cloud and edge infrastructure used to store data was approximately 73 per cent cloud vs 27 per cent edge.
- **Functional areas:** the survey findings show that the most data-intensive productive activity is research and development. For the more data-intensive, corporate activity is data storage and back-up.

A. Entity

*****1. Your entity is:

- ☐ Public
- ☐ Semi-public
- ☐ Private
- ☐ Other

- *If 'other' is the selected response, please specify:*

*****2. The size of your entity is:

- ☐ Micro (1-9 employees)
- ☐ Small (10-49 employees)
- ☐ Medium (50-249 employees)
- ☐ Large (250 or more employees)

*****3. Your entity is headquartered in:

- ☐ The European Union
- ☐ In a non-EU country

- *If 'the European Union' is the selected response, please specify the country in which you entity is headquartered in: Austria, Belgium, Bulgaria, Croatia, Republic of*

315 https://ec.europa.eu/digital-single-market/en/european-data-flow-monitoring-initiative

Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, the UK

- *If 'in a non-EU country' is the selected response, please specify the country in which your entity is headquartered in: the United States of America, China, Japan, South Korea, Canada, Brazil, New Zealand, Australia, Chile, Argentina, Andorra, Faroe Islands, Israel, Switzerland, Uruguay, Other*
  - *If 'other' is the selected response, please specify the country in which your entity is headquartered in:*

*****4. Your entity operates in:

- ☐ One European country
- ☐ Between 2 and 4 European countries
- ☐ Between 5 and 10 European countries
- ☐ More than 10 European countries
- ☐ Non-EU countries

- *If 'one European country' is the selected response, please specify the country in which your entity operates in:*

	Country
1	

- *If 'between 2 and 4 European countries' is the selected response, please specify the countries in which your entity operates in:*

	Country
1	
2	
3	
4	

- *If 'between 5 and 10 European counties' is the selected response, please specify the countries in which your entity operates in:*

	Country
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	

- *If 'more than 10 European countries' is the selected response, please specify the 10 most important countries in which your entity operates in:*

	Country
1	

2	
3	
4	
5	
6	
7	
8	
9	
10	

- *If 'non-EU countries' is the selected response, please specify up to the 5 most important countries in which your entity operates in:*

	Country
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	

***5**. Your entity operates in one of the following sectors:

☐ Academia
☐ Banking and financial services
☐ Cloud services
☐ Data analytics
☐ Health
☐ Manufacturing
☐ Professional services (consultancy)
☐ Public services
☐ Research
☐ Software
☐ Telecommunications
☐ Other categories

- *If 'other' is the selected response, please specify the sector in which your entity operates in following NACE Codes:* NaceRev2_2008.pdf

***6**. Do you agree to be contacted by the European Commission's services for possible follow-up information to your responses?

☐ Yes
☐ No

- *If 'yes' is the selected response, please provide the following information and agree to the personal data protection provisions. The personal information provided will not be published.*

First name:	
Surname:	
Email:	
Company you are working for:	
Job title:	

B. Cloud Computing: Data Stocks

*****7. Your entity stores data in the cloud:

- ☐ Yes
- ☐ No

*****8. Your entity's data are physically stored in:

- ☐ The European Union
- ☐ Non-EU countries
- ☐ I don't know

- *If 'the European Union' is the selected response, please specify the country in which you entity is headquartered in: Austria, Belgium, Bulgaria, Croatia, Republic of Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, the UK*
- *If 'in a non-EU country' is the selected response, please specify the country in which your entity is headquartered in: the United States of America, China, Japan, South Korea, Canada, Brazil, New Zealand, Australia, Chile, Argentina, Andorra, Faroe Islands, Israel, Switzerland, Uruguay, Other*
    - *If 'other' is the selected response, please specify the country in which your entity is headquartered in:*

*****9. The yearly data volume in terabytes (TB) your entity stores in the cloud for its own operations is: __

*9 bis. The yearly data volume in terabytes (TB) your entity generates is: ☐
*9.bis The percentage between cloud and edge infrastructure storage is: XX %Cloud vs XX %Edge

*****10. Please specify up to the 5 most important countries and the yearly associated volume of data stored in each of them by your entity for its own operations, in terabytes (TB)

	Country	**Volume (in TB)**
1		
2		
3		
4		

5		

***11.** Your entity stores data in the cloud on behalf of other entities:

- ☐ No
- ☐ Yes

- *If 'yes' is the selected response, please specify the yearly data volume stored in terabytes (TB) by your entity in the cloud on behalf of other entities:*
- *Please specify up to the 5 most important countries and the yearly associated volume of data stored in each one by your entity on behalf of other entities, in terabytes (TB)*

	Country	**Volume (in TB)**
1		
2		
3		
4		
5		

*12. Please specify up to the 5 most used cloud applications by your entity and the corresponding volume of data stored for each of them:

	Application	**Volume (in TB)**
1		
2		
3		
4		
5		

***12 bis.** Please specify the yearly volume of data in TB consumed by the selected services

Service	Volume (in TB)
Software as a Service (SaaS)[316]	
Platform as a Service (PaaS)[317]	
Infrastructure as a Service (IaaS)[318]	

316 Applications such as email, SharePoint, video conferencing, collaborative tools and program interfaces, ex. social networking sites, which the user does not manage or control the underlying cloud infrastructure, such as operating system, data centres, storage.

317 Such as databases, web data centres, runtimes, over which the user has control and configuration settings for the application-hosting environment.

318 Infrastructures such as storage, data centres, network, load balancers, where the consumer is able to deploy and run, by having the control over operating systems)

Edge services[319]		
Cloud enabled services	AI services	
	Business data services	
	Blockchain services	
	HCP as a service (HPaaS)	
	Digital twins services	
Others		

***12. bis** Please specify which of the following productive activities are data intensive from 1 to 10, (1 being not data intensive at all and 10 being the maximum that an activity can be data intensive)

Data Intensive Activities	Ranking
Research and Development activities	
Finance and accounting	
Purchasing and procurement	
Operation and assembly (product and service development, service design in case of public administrations)	
Marketing, sales and after sale services,	
Logistic	
Other (please specify)	

***12. bis** Please specify which of the following corporate activities your entity perform the most from 1 to 10, (1 being your entity does not perform the activity at all and 10 being your entity does perform the activity intensively)

Activities generating data	Ranking
Human Resource Management	
Client relationship management	
Enterprise resource management	
Enterprise collaboration activities	
Data storage	
Back-up	
Replicability	
Enterprise security and cybersecurity	
Content workflow and management activities	
Other	

*13. To store data in the cloud, your entity uses a cloud deployment mode that is:

- ☐ Private
- ☐ Public

- *If 'private' is the selected response, please indicate the type:*

319 Services enabled when using edge computing

- *On premise*
- *Off premise*

- *If 'private' is the selected response, please specify up to the 5 main countries in which your entity's data are physically stored and the associated yearly volume of data stored in each of them in terabytes (TB):*

	Country	Volume (in TB)
1		
2		
3		
4		
5		

- *If 'public' is the selected response, please specify up to the 5 main countries in which your entity's data are physically stored and the associated yearly volume of data stored in each of them in terabytes (TB):*

	Country	Volume (in TB)
1		
2		
3		
4		
5		

*****14. The types of data your entity stores in cloud infrastructures are

- ☐ Personal (e.g., data relating to identified or identifiable natural person)
  - If yes, specify the volume of personal data in a year (TB) __
    - Please specify the percentage of personal data generated by a user/individual
- ☐ Non-personal (e.g., data not relating to identified or identifiable natural person)
  - If yes, specify the volume of non-personal data in a year (TB) __
    - Please specify the percentage of non-personal data generated by a machine (connected devices, sensors, IoT, platform etc.)
    - Please specify the percentage of non-personal data generated by machine which is industrial data (data used as a direct input into the production process from a supply chain perspective)
- ☐ I don't know

*****15 The current cloud contract of your entity terminates in:

- ☐ 2019

- ☐ 2020
- ☐ 2021
- ☐ 2022
- ☐ Other
- ☐ I don't know

C. Cloud Data Flows

***16.** Your entity physically moved data stored in cloud infrastructures from one geographical location to another in the past 12 months

- ☐ No
- ☐ Yes, partially
- ☐ Yes, completely
- ☐ I don't know

- *If 'no' is the selected response, please specify the year when your entity moved the data to the current location:*
- *If 'yes, partially' is the selected response:*
  - *please provide the information below:*

	Volume moved (in TB)	Former location	New location	Data type
1				
2				
3				
4				
5				

  - *please provide an estimation of the cost (in €) associated to data migration born by your entity over the past 12 months:*
- *If 'yes, completely' is the selected response:*
  - *please provide the information below:*

	Volume moved (in TB)	Former location	New location	Data type
1				
2				
3				
4				
5				

  - *please provide an estimation of the cost (in €) associated to data migration born by your entity over the past 12 months:*

***16. bis.** You entity switched cloud service provider:

- ☐ 1 to 11 month ago
- ☐ 12 months to 24 months ago
- ☐ 25 months to 36 months ago
- ☐ 36 months to 48 months ago

- ☐ 49 months to 60 months ago
- ☐ Never
- ☐ I do not know

*17. If your entity moved data from one cloud infrastructure to another in the last 12 months across the European Union the reason is:

- ☐ A change of cloud service provider
- ☐ A change in cloud deployment mode (a) from public to private cloud
- ☐ A change in cloud deployment mode (b) from private to public cloud
- ☐ A change in cloud deployment mode (c) from central to edge cloud
- ☐ A change in cloud deployment mode (d)) from edge to central cloud
- ☐ A combination of central deployment mode with edge
- ☐ Security reasons
- ☐ Legal reasons
- ☐ The need to replicate data for the entity
- ☐ The need for entity internal backups
- ☐ Entity data management needs
- ☐ Entity internal efficiencies
- ☐ Cost efficiencies
- ☐ Intra-company needs to enable business activities among different entity locations
- ☐ Other
- ☐ I don't know

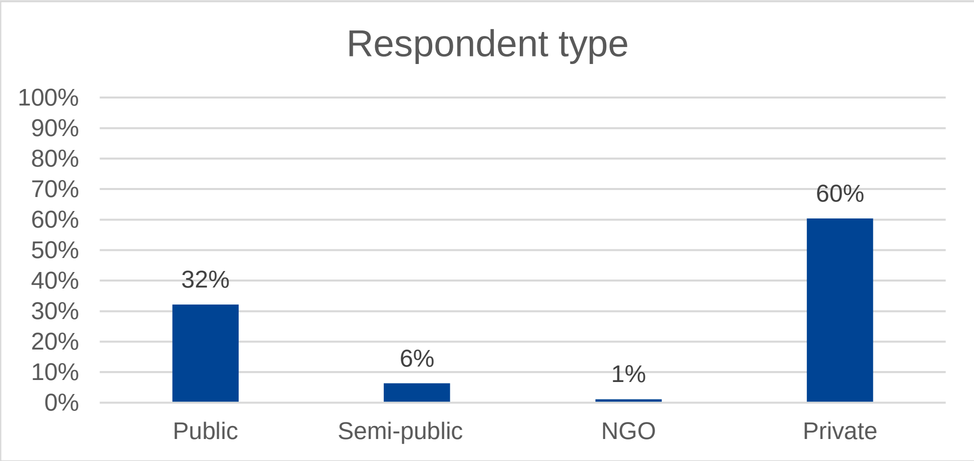
*18. Further information your entity would like to share as part of this survey please indicate:

__

Annex 9: Survey Results

A. Entity

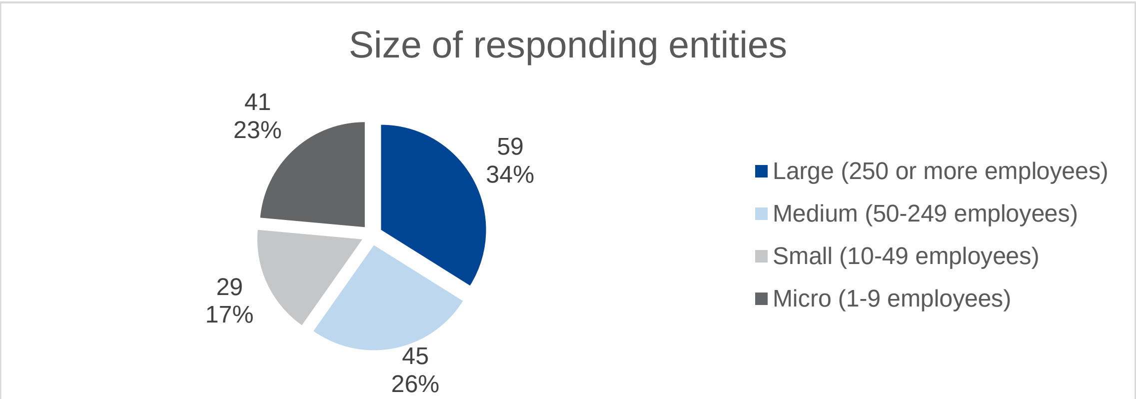
*1. Your entity is:



Analysis based on full sample

n = 174

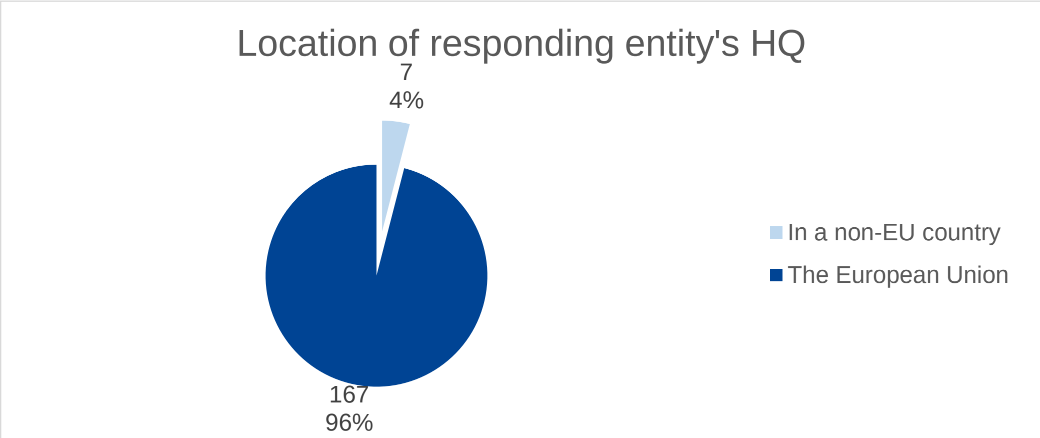
*2. The size of your entity is:



Analysis based on full sample

n = 174

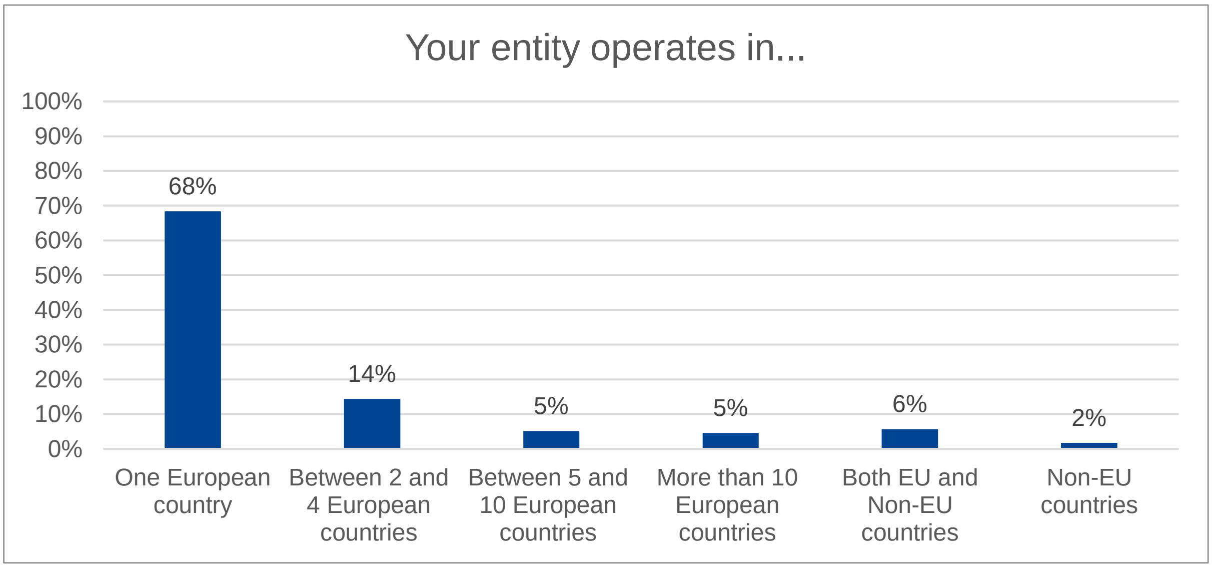
*3. Your entity is headquartered in:



Analysis based on full sample

n = 174

*4. Your entity operates in:



Analysis based on full sample

n = 174

Most frequently occurring countries in which respondents operate (top ten):

Country	Occurrence
Spain	21

Germany	19
France	19
United Kingdom	17
Italy	16
Netherlands	12
Poland	11
Greece	9
Denmark	8
Belgium	8

Analysis based on full sample Analysis based on full sample

n = 174

*5. Your entity operates in one of the following sectors:

Sector	Count	%
Professional services (consultancy)	23	13%
Manufacturing	15	9%
Public services	14	8%
Banking and financial services	14	8%
Software	9	5%
Health	9	5%
Cloud services	3	2%
Telecommunications	3	2%

Construction	3	2%
Academia	3	2%
Other	27	16%
Did not answer	51	29%
Total	**174**	**100%**

Analysis based on full sample

n = 174

'Other' encompasses the field 'Other categories' and all sectors with only one or two respondents

One respondent selected more than one sector (Data analysis, Data brokerage and Software). Individual research was done on this respondent to attribute it to Software

6* Do you agree to be contacted by the European Commission's services for possible follow-up information to your responses?

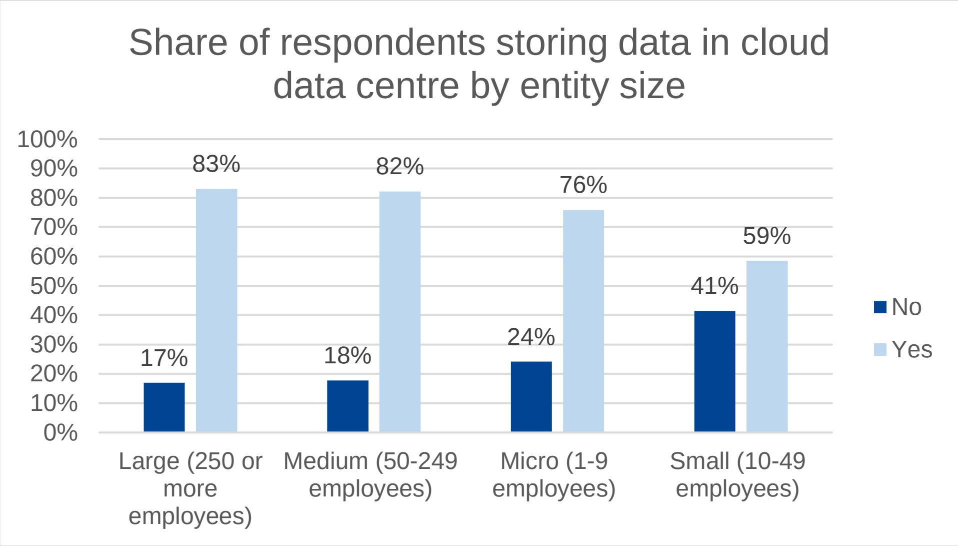
	Count	%
No	73	42%
Yes	101	58%
Grand Total	174	100%

Analysis based on full sample

n = 174

B. Cloud Computing: Data Stocks

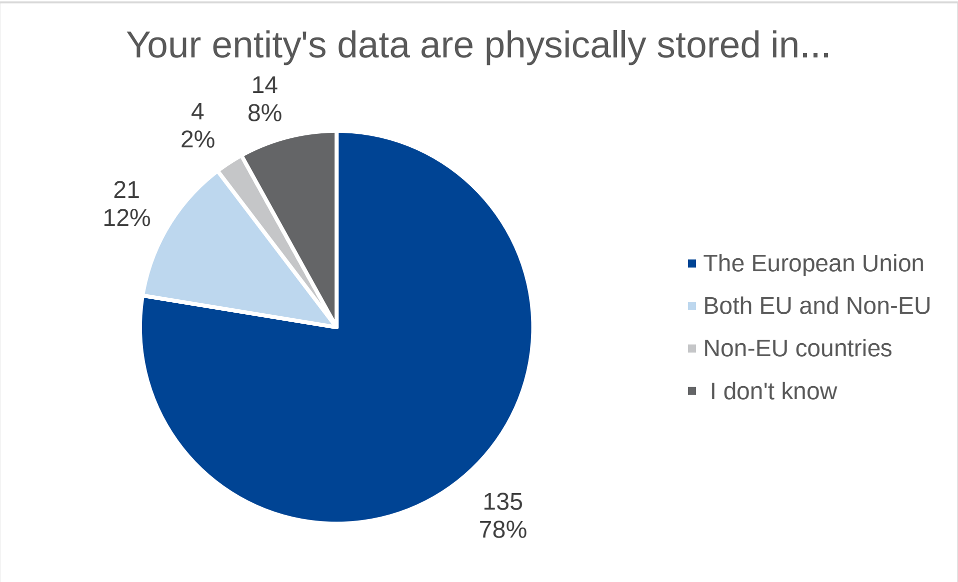
*7. Your entity stores data in the cloud:



Analysis based on full sample

n = 59 (large entities), 45 (medium), 29 (small), and 41 (micro)

*8. Your entity's data are physically stored in:



Analysis based on full sample

n = 174

Your entity's data are physically stored in (country):

Country	Count
Ireland	37
Germany	31
Netherlands	27
France	25
Finland	18
Sweden	17
United States	17
United Kingdom	17
Belgium	15
Denmark	14
Estonia	14
Spain	9
Italy	6
Austria	5
China	5
Poland	4
Brazil	4
Luxembourg	3
Japan	3
Canada	3
Greece	2

Latvia	2
Portugal	2
Romania	2
Slovenia	2
Australia	2
Switzerland	2
Singapore	2
UAE	2
Bulgaria	1
Croatia	1
Czech Republic	1
Hungary	1
Lithuania	1
South Korea	1
Chile	1
Israel	1
New Zealand	1
Other	*3*

Full sample

NB: some entities have indicated more than one country in which they store data

*Other excludes those responses where a country was subsequently specified i.e. Russia (1), Singapore (2), South Africa (1) and UAE (2). These countries have been added to the table

*9. The yearly data volume in terabytes (TB) your entity stores in the cloud for its own operations is:

By entity size:

Row Labels	Average of 9. The yearly data volume in terabytes (TB) your entity stores in the cloud for its own operations is:
Large (250 or more employees)	30,0
Medium (50-249 employees)	3,5
Small (10-49 employees)	1,2
Micro (1-9 employees)	4,0
Grand Total	**10,4**

Analysis based on smaller sample of respondents (61 entries) identified by Tech4i2 as being more reliable for analysis of data generation/storage volumes (see Excel with key parameters for the selection)

n = 16 (large entities), 17 (medium), 7 (small), and 21 (micro)

By entity type:

Row Labels	Average of 9. The yearly data volume in terabytes (TB) your entity stores in the cloud for its own operations is:
Private	4,2
Public	25,2
Semi-public	4,8
NGO	0,0
Grand Total	**10,4**

Analysis based on smaller sample of respondents identified by Tech4i2 as being more reliable for analysis of data generation/storage volumes (see Excel with key parameters for the selection) n = 36 (private entities), 18 (public), 6 (semi-public), and 1 (NGO)

*9.a. The yearly data volume in terabytes (TB) your entity generates is:

By entity size:

Row Labels	Average of 9.a The yearly data volume in terabytes (TB) your entity generates is:
Large (250 or more employees)	478,2
Medium (50-249 employees)	11,7
Small (10-49 employees)	0,8
Micro (1-9 employees)	5,0
Grand Total	**101,2**

Analysis based on smaller sample of respondents identified by Tech4i2 as being more reliable for analysis of data generation/storage volumes (see Excel with key parameters for the selection)

n = 16 (large entities), 17 (medium), 7 (small), and 21 (micro)

By entity type:

Row Labels	Average of 9.a The yearly data volume in terabytes (TB) your entity generates is:
Private	123,5
Public	97,8
Semi-public	19,9
NGO	0,0
Grand Total	**101,2**

Analysis based on smaller sample of respondents identified by Tech4i2 as being more reliable for analysis of data generation/storage volumes (see Excel with key parameters for the selection)

n = 36 (private entities), 18 (public), 6 (semi-public), and 1 (NGO)

By country:

Country	Number of entities:	Average yearly data volume in terabytes (TB) generated:
France	3	693.7
Sweden	5	205.2
Finland	5	201.1
Spain	2	80.0
Germany	4	57.1
Netherlands	7	9.9
United Kingdom	6	7.0
Ireland	3	4.3
Estonia	6	3.2
Belgium	7	1.0
Luxembourg	2	0.3
Denmark	4	0.3
Sample	**54**	**101.2**

Analysis based on smaller sample of respondents identified by Tech4i2 as being more reliable for analysis of data generation/storage volumes (see Excel with key parameters for the selection)

n = 54 (seven respondents in the smaller sample did not answer this question)

*9.b. The percentage between cloud and edge infrastructure storage is: XX %Cloud vs XX %Edge

Entity size/ type	Sample size	Average Cloud storage (%)	Average Edge storage (%)
Large (250 or more employees)	16	70.6	29.4
Medium (50-249 employees)	16	58.6	41.4
Small (10-49 employees)	7	74.1	25.9
Micro (1-9 employees)	18	93.6	6.4
Private	34	74.3	25.7
Public	17	74.2	25.8
Semi-public	6	66.0	34.0
NGO	0		
All entities	**57**	**73.3**	**26.7**

Analysis based on smaller sample of respondents identified by Tech4i2 as being more reliable for analysis of data generation/storage volumes (excel with key parameters for the selection)

Entries for the respondents with the following ID numbers were removed because they were either very small numbers (98, 110, 111) or did not sum to 100% (631)

n = sample size shown for each entity size/type

*10. Please specify up to the 5 most important countries and the yearly associated volume of data stored in each of them by your entity for its own operations, in terabytes (TB)

Top ten countries ranked by average data stored:

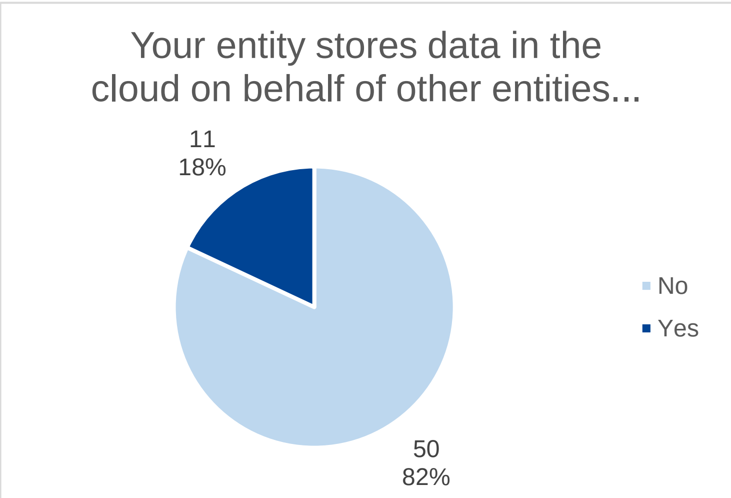
Country	Count	Sum of TB	Average TB
Finland	8	1114.8	139.4
Germany	8	251.5	31.4

Italy	3	83.0	27.7
France	8	193.5	24.2
Sweden	7	80.3	11.5
Spain	2	16.0	8.0
Netherlands	9	60.0	6.7
United Kingdom	10	54.2	5.4
Croatia	1	5.0	5.0
Belgium	8	35.3	4.4

Analysis based on smaller sample of respondents identified by Tech4i2 as being more reliable for analysis of data generation/storage volumes (see Excel with key parameters for the selection)

One respondent answered ‘Less than 1TB’, which was rounded up to 1

*11. Your entity stores data in the cloud on behalf of other entities:



Analysis based on smaller sample of respondents identified by the identified by Tech4i2 as being more reliable for analysis of data generation/storage volumes (see Excel with key parameters for the selection)

n = 61

If 'yes' is the selected response, please specify the yearly data volume stored in terabytes (TB) by your entity in the cloud on behalf of other entities:

Only one answer: ‘20’, among the smaller sample of 61 respondents identified by the Tech4i2 as being more reliable.

Please specify up to the 5 most important countries and the yearly associated volume of data stored in each one by your entity on behalf of other entities, in terabytes (TB)

Countries ranked by average data stored on behalf of other entities:

Country	Count	Sum (TB)	Average (TB)
Germany	3	230.0	76.7
France	1	70.0	70.0
Sweden	2	104.2	52.1
Finland	4	108.0	27.0
Denmark	1	10.0	10.0
Belgium	1	8.0	8.0
Spain	1	6.0	6.0
Italy	3	15.5	5.2
Netherlands	1	5.0	5.0
Greece	1	5.0	5.0
Luxembourg	1	4.7	4.7
Ireland	1	4.0	4.0
United Kingdom	1	3.0	3.0
Total responses*	**21**	**573.4**	**27.3**

Analysis based on smaller sample of respondents identified by Tech4i2 as being more reliable for analysis of data generation/storage volumes ((see Excel with key parameters for the selection)

Nine respondents answered this question

*Some respondents indicated up to five countries in which they store data on behalf of other entities

*12. Please specify up to the 5 most used cloud applications by your entity and the corresponding volume of data stored for each of them:

Top ten applications ranked by frequency of use among survey respondents:

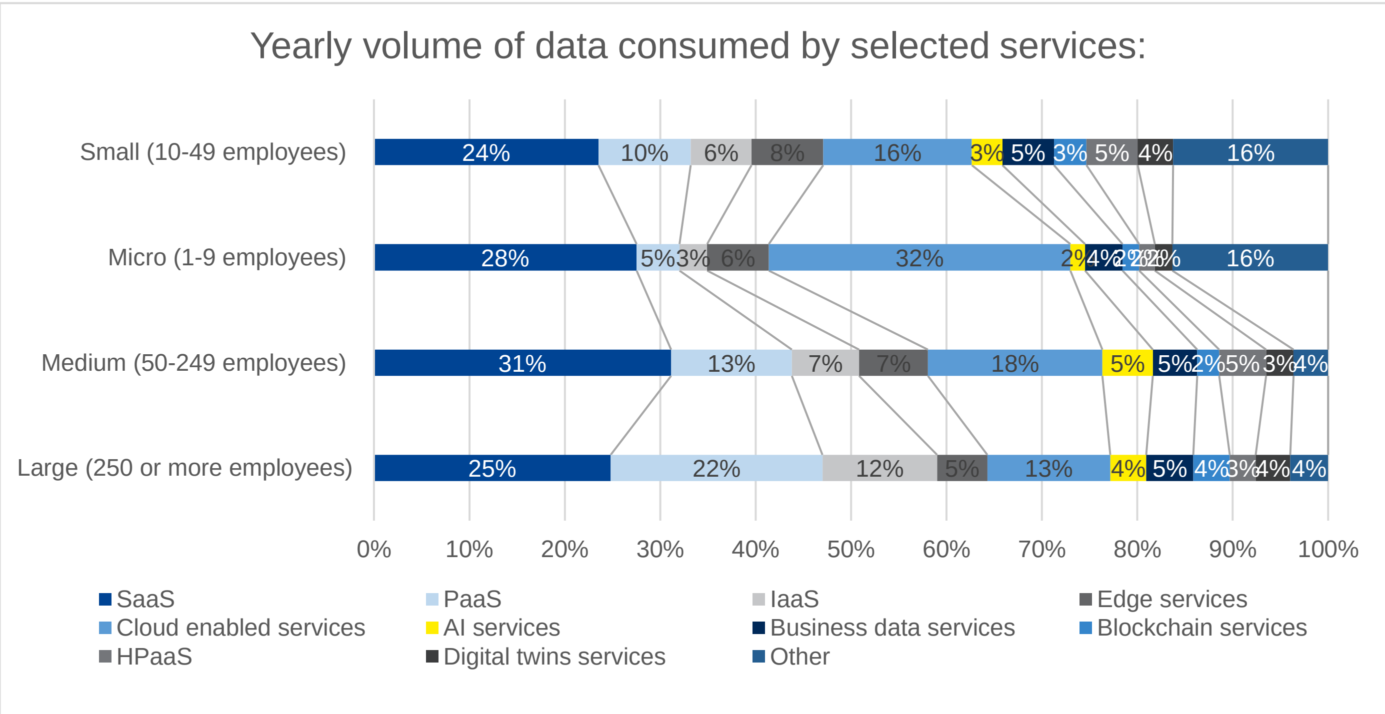
Row Labels	Applications	Average volume of data stored (TB)
Google	19	4769,1
Microsoft	18	5584,4
Data storage	12	1027,6
Amazon	8	7732,0
Dropbox	8	88,9
Security software	6	1022,0
Email	5	71,4
CRM	5	4,0
Backup	4	250006,0
ERP	4	1257,8
Workflow services	3	1005,0
ICloud	3	34,3
IBM	3	308,7
Multimedia	3	19,7

SAP	2	678,0
Cloud	2	262,0
Bureautic applications	2	0,8
Oracle	2	5005,0
Apple	2	40001,0
GIS	2	350,5
Other/not specified	135	95763,3
Grand Total	**248**	**57670,7**

Analysis based on full sample

n=174

Responses coded to identify applications intended by the respondents even where their answers differed e.g. references to ‘Office 365’, ‘Sharepoint’, ‘OneDrive’ and ‘Excel’ were all coded as ‘Microsoft’.



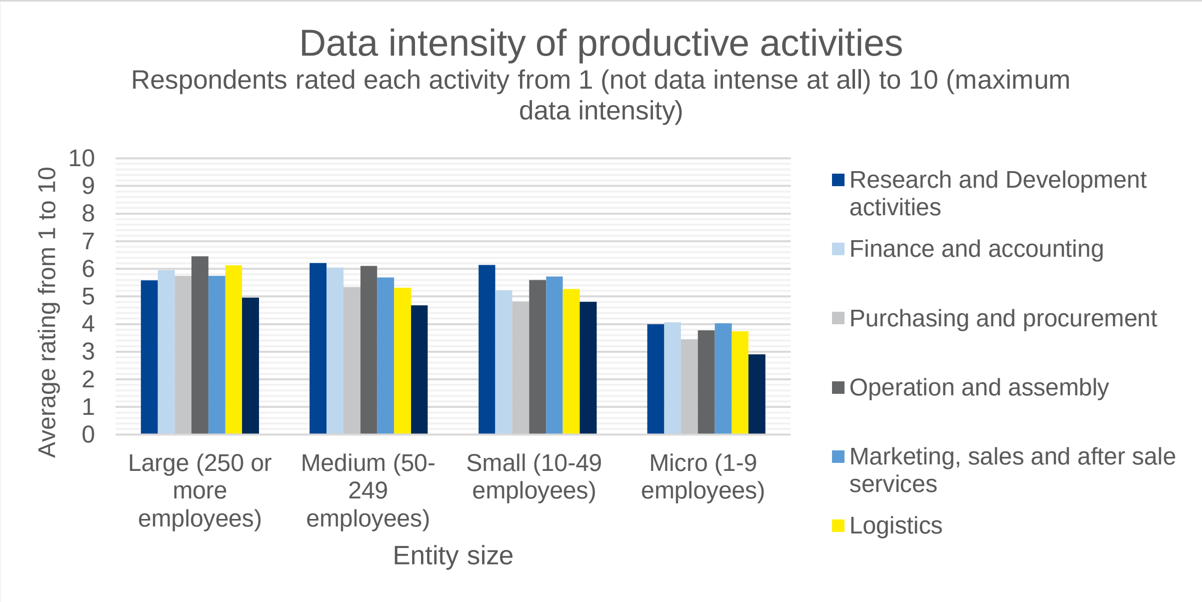
*12.a. Please specify the yearly volume of data in TB consumed by the selected services

Analysis based on full sample, respondents who answered to the question

n = 21 (large entities), 34 (medium), 20 (small), and 23 (micro) = 98 (total)

To find the total size of all answers in TB and then work out the percentage of TB used for each service (e.g. if total is 10TB and AI services BIS_6 is 2TB we know that would be 20% of cloud usage)

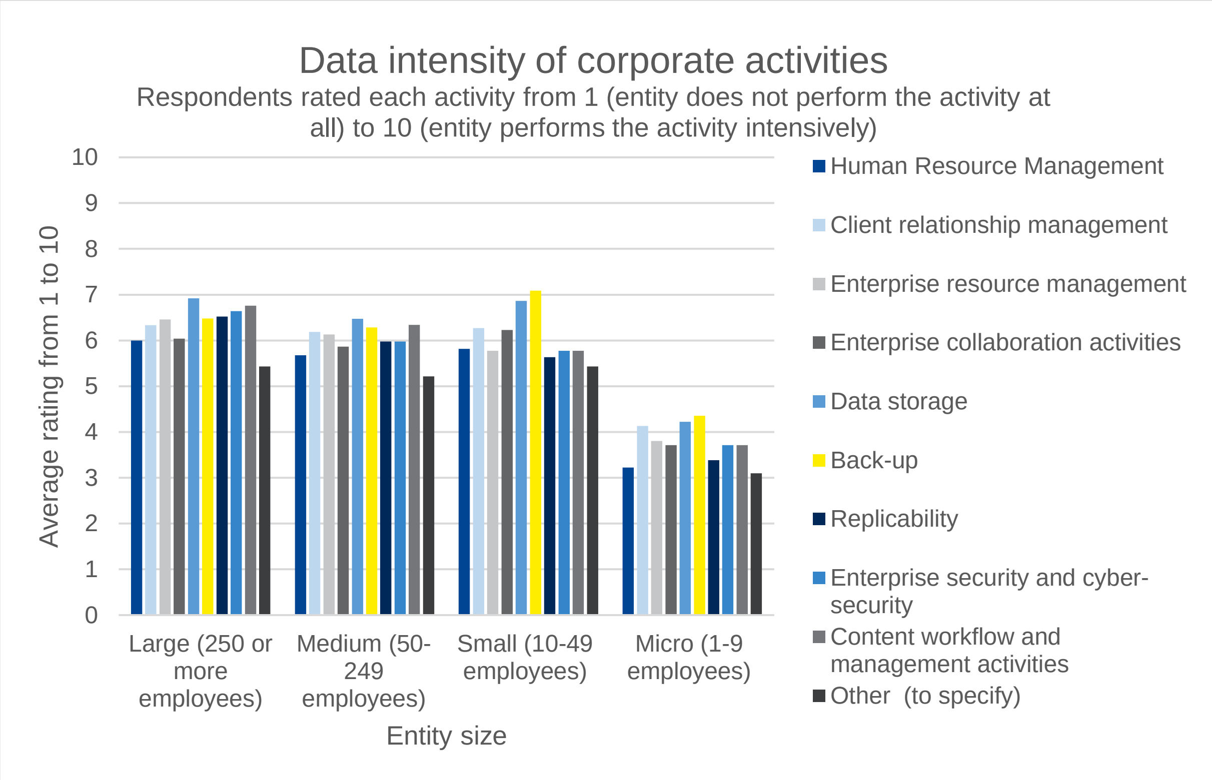
*12.b. Please specify which of the following productive activities are data intensive from 1 to 10, (1 being not data intensive at all and 10 being the maximum that an activity can be data intensive)



Analysis based on full sample minus responses which we understand were ‘early’ submissions made before this question had been added to the survey. These were excluded from the analysis (data was cleaned to remove these ‘0’ answers in relation to the following 56 submissions by ID number: 20001 to 20053, 20055, 20057, and 20058)

n = 29 (large), 38 (medium), 22 (small), and 23 (micro)

*12.c. Please specify which of the following corporate activities your entity perform the most from 1 to 10, (1 being your entity does not perform the activity at all and 10 being your entity does perform the activity intensively)



Analysis based on full sample minus responses which we understand were 'early' submissions made before this question had been added to the survey. These were excluded from the analysis (data was cleaned to remove these '0' answers in relation to the following 56 submissions by ID number: 20001 to 20053, 20055, 20057, and 20058)

n = 29 (large), 38 (medium), 22 (small), and 23 (micro)

*13. To store data in the cloud, your entity uses a cloud deployment mode that is:

Row Labels	Both	Private	Public	Grand Total
Large (250 or more employees)	0%	23%	3%	26%
Medium (50-249 employees)	0%	21%	7%	28%
Small (10-49 employees)	2%	10%	0%	11%
Micro (1-9 employees)	0%	20%	15%	34%
Grand Total	**2%**	**74%**	**25%**	**100%**

Analysis based on smaller sample of respondents (61 entries) identified by Tech4i2 as being more reliable for analysis of data generation/storage volumes (see Excel with key parameters for the selection)

n = 16 (large entities), 17 (medium), 7 (small), and 21 (micro)

If 'private' is the selected response, please indicate the type:

Row Labels	On premises	Off premises	I don't know
Large (250 or more employees)	10	8	4
Medium (50-249 employees)	9	6	5
Small (10-49 employees)	4	2	1
Micro (1-9 employees)	7	4	11
Grand Total	**30**	**20**	**21**

Analysis based on smaller sample of respondents identified by the contractor as being more reliable

n = 16 (large entities), 17 (medium), 7 (small), and 21 (micro)

NB: on and off premises are not mutually exclusive

*14. The types of data your entity stores in cloud infrastructure are

Entities by size:

	Personal	Non-personal	I don't know
Large (250 or more employees)	51%	75%	14%
Medium (50-249 employees)	49%	71%	11%
Small (10-49 employees)	55%	48%	21%
Micro (1-9 employees)	51%	56%	22%
All entities	**51%**	**65%**	**16%**

Analysis based on full sample

n = 59 (large entities), 45 (medium), 29 (small), and 41 (micro)

Entities by type:

	Personal	Non-personal	I don't know
Private	55%	59%	18%
Public	36%	73%	18%
Semi-public	50%	50%	0%
NGO	50%	68%	15%
All entities	**51%**	**65%**	**16%**

Analysis based on full sample

n = 105 private entities, 56 public, 11 semi-public, 2 NGOs

Personal (e.g., data relating to identified or identifiable natural person)

Specify the volume of personal data in a year (TB) __

6 entries

Average (mean) = 11.36 TB (if outlier '60' is disregarded, average = 1.6 TB)

Please specify the percentage of personal data generated by a user/individual

Disregarding one answer of '200', the mean average is 11.4% (count of 19 responses).

There is a doubt whether all respondents understood this question was about % or whether some respondents may have provided a TB figure.

Non-personal (e.g., data not relating to identified or identifiable natural person)

Specify the volume of non-personal data in a year (TB) __

9 entries

Average (mean) = 4.4 TB

Please specify the percentage of non-personal data generated by a machine (connected devices, sensors, IoT, platform etc.)

The mean average is 22.4% (count of 29 responses).

As above, there is a doubt on whether all respondents understood this question was about % or whether some respondents may have provided a TB figure e.g. ‘1’.

Please specify the percentage of non-personal data generated by machine which is industrial data (data used as a direct input into the production process from a supply chain perspective)

The mean average is 18.9% (count of 29 responses).

As above, there is a doubt on whether all respondents understood this question was about % or whether some respondents may have provided a TB figure e.g. ‘1’.

*15 The current cloud contract of your entity terminates in:

Year	Count
2019	3
2020	27
2021	35
2022	24
I don't know	61
Other (specify)	21
Grand Total	**171**

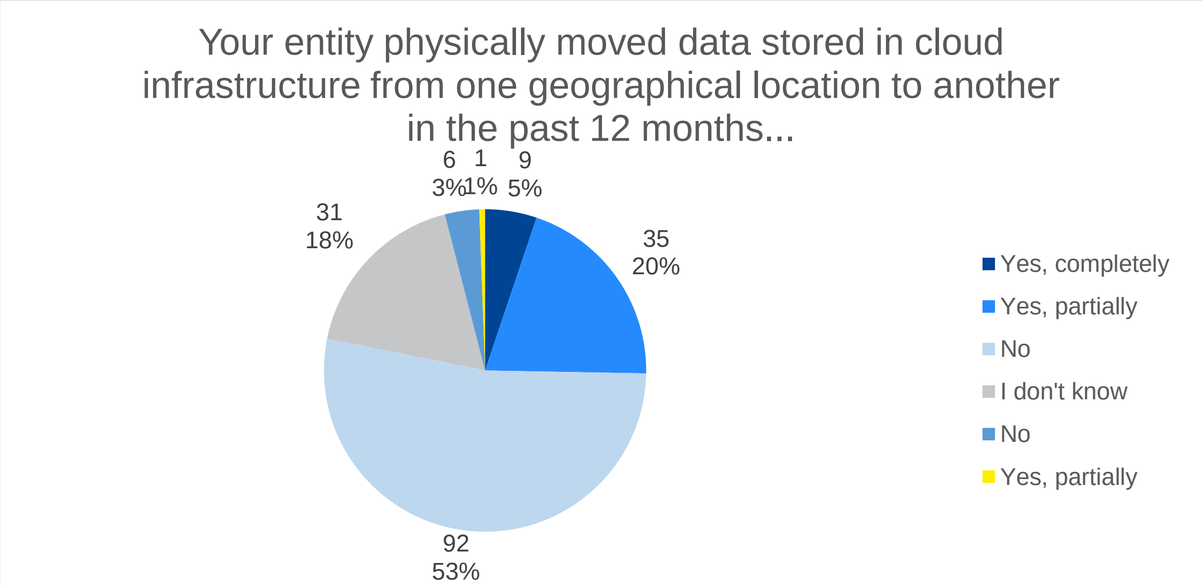
Analysis based on full sample minus 3 respondents who answered ‘2019’ but were respondents of the survey carried out in 2020.

‘Other’ answers: ‘2023’, ‘2024’, ‘2025’, ‘2050’, ‘2099’, the rest left blank.

n = 171

C. Cloud Data Flows

*16. Your entity physically moved data stored in cloud infrastructures from one geographical location to another in the past 12 months



Analysis based on full sample

n = 174

If 'no' is the selected response, please specify the year when your entity moved the data to the current location:

Average year: 2015 (78 responses, excluding one outlier response of 1989 (Respondent 557) and one ‘0’ (Respondent 20021)

If ‘yes, partially’ is the selected response:

Average volume moved: 6.4 TB (23 movements across 11 respondents)

‘0’ answers from Respondents 80 and 546 disregarded

please provide an estimation of the cost (in €) associated to data migration born by your entity over the past 12 months:

€ 137,000 is the average cost although the data range from €1 to €2,000,000 (15 responses)

Excluding outliers of €1, €5 and €2,000,000, the average cost is €4,660

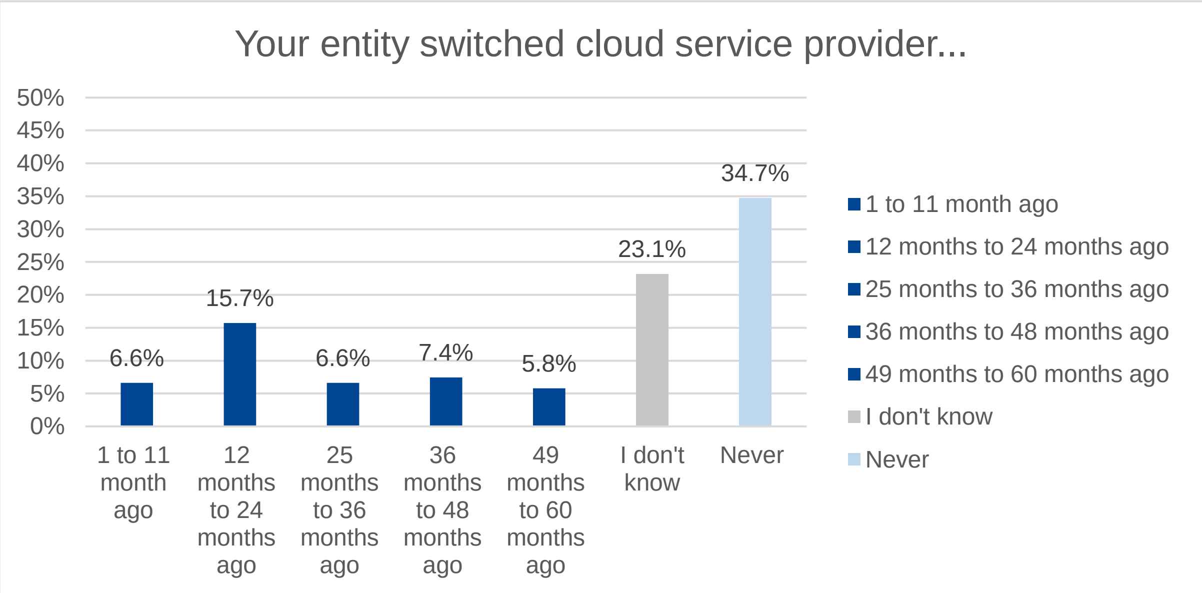
If ‘yes, completely’ is the selected response:

Average volume moved: 14.3 TB (4 movements across 4 respondents)

please provide an estimation of the cost (in €) associated to data migration born by your entity over the past 12 months:

No data on this in the smaller sample.

*16. bis. Your entity switched cloud service provider:



Analysis based on full sample, respondents who answered to the question

n = 121

*17. If your entity moved data from one cloud infrastructure to another in the last 12 months across the European Union the reason is:

Reason for moving data from one cloud infrastructure to another	Count	%
I don't know	34	33%
Legal reasons	9	9%
Security reasons	8	8%
The need to replicate data for the entity	8	8%
Cost efficiencies	8	8%

A change of cloud service provider	6	6%
Entity internal efficiencies	6	6%
A change in cloud deployment mode (b) from private to public cloud	5	5%
The need for entity internal backups	7	7%
Entity data management needs	5	5%
Intra-company needs to enable business activities among different entity locations	6	6%
A combination of central deployment mode with edge	1	1%
A change in cloud deployment mode (a) from public to private cloud	0	0%
A change in cloud deployment mode (c) from central to edge cloud	0	0%
A change in cloud deployment mode (d)) from edge to central cloud	0	0%
Other	1	1%
Total	**104**	**100%**

Analysis based on full sample, respondents who answered to the question

n = 59

*18. Further information your entity would like to share as part of this survey please indicate:

—

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

– by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

– at the following standard number: +32 22999696, or

– by email via: https://europa.eu/european-union/contact_en

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU publications

You can download or order free and priced EU publications from: https://op.europa.eu/en/publications. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

EU law and related documents

For access to legal information from the EU, including all EU law since 1952 in all the official language versions, go to EUR-Lex at: http://eur-lex.europa.eu

Open data from the EU

The EU Open Data Portal (http://data.europa.eu/euodp/en) provides access to datasets from the EU. Data can be downloaded and reused for free, for both commercial and non-commercial purposes.

KK-03-21-388-EN-N

Publications Office
of the European Union

doi:10.2759/7355
ISBN 978-92-76-41937-2