Paper

Measuring the internet economy in the Netherlands 2016-2018

A big data analysis

Lotte Oostrom
Jaap Jansen
Raymond Kleingeld
Andries Kuipers
Magda Slootbeek-Van Laar
Joram Vuik
Adam N. Walker

May 2020

Content

1. Executive summary

This study primarily addresses the question; to what extent is the internet economy becoming an important part of the economy in the Netherlands? In 2016, Statistics Netherlands carried out a first study which defined the internet economy, developed a methodology to measure it, and produced statistics for the year 2015. The next step is try to understand how the internet economy is developing over time. Creating a time-series of data on the internet economy facilitates analyses of trends and provides a new angle for evaluating the big data approach. Therefore, the Ministry of Economic Affairs and Climate has asked Statistics Netherlands to describe the internet economy for the years 2016, 2017 and 2018.

In order to create the time series, Statistics Netherland has again worked in partnership with Dataprovider. Dataprovider has extensive experience in crawling the internet to collect data on websites on a regular basis, in particular companies' websites. When this data source is linked to the Statistics Netherlands data on businesses, the resulting dataset permits an analysis of diverse economic indicators (e.g. turnover, employment etc.) at both national and regional level, and over time.
The methodology employed classifies businesses into various categories depending on how a business makes use of the internet. These categories are:

- A: Businesses without a website
- B: Businesses with a passive (category B1) or active online presence (B2)
- C: Online stores
- D: Online services
- E: Internet-related ICT

Websites are allocated to these categories predominantly according to the information available from Dataprovider. Categories C, D and E as a group constitute the "core" of the internet economy. The core consists of online stores, online services such as dating sites, price comparison sites, or online entertainment, and of internet-related ICT such as app developers, web-hosting and internet marketing. Outside of the core we distinguish two further types of online presence for businesses: active and passive. Active online presence means that businesses provide a manner to interact with them directly, such as making a reservation or ordering a brochure. Passive online presence means that businesses purely use the internet to provide information about their activities and to publicise their organisation.

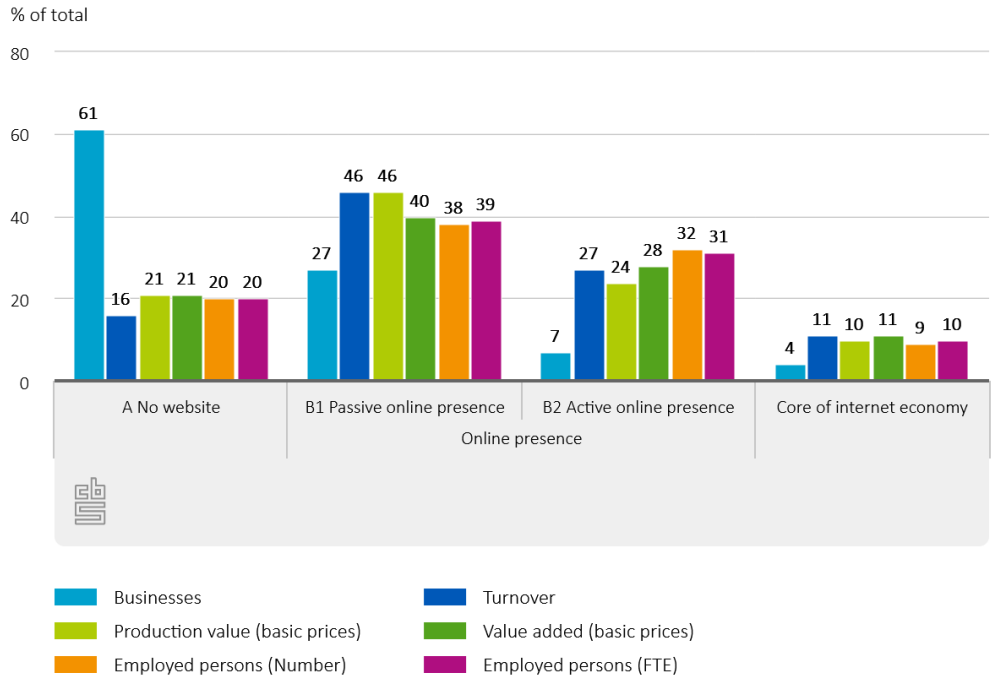
To analyse the internet economy in a coherent way, the characteristics of the websites need to be linked to data on the businesses behind the website. This methodological challenge is dealt with using various combinations of keys (Chamber of Commerce (CoC) number, website name, postcode, telephone and email address) into order to maximise the number of successful links which can be made. Some key combinations are more likely to result in a correct linkage than others. Accordingly, care is taken to ensure the greatest number of links while minimising the risks of incorrect links.

In 2018, 609,000 businesses were involved to some extent in the internet economy. This accounts for 36% of all businesses, and this share is rising over time: in 2016 the share was

34%. In 2018, for the economy as a whole 3% of all businesses belong to the core of the internet economy being online stores, other online services and internet-related ICT. These businesses contribute for 6.5% to the total number of jobs in the Dutch economy (6.2% in 2016). In 2018 of all jobs, 83% is associated with businesses with some kind of online presence (81% in 2016).

Figure 1.1 shows economic indicators per category of the internet economy. In 2018, businesses with only a passive online presence contribute the most to the business economy[1], varying from 46% of turnover to 38% of the number of employed persons. The contribution of the core of the internet economy to the business economy varies from 11% of turnover to 9% of the number of employed persons. In the period 2016-2018, the contribution of the core of the internet economy to the business economy was more or less stable and fluctuates round the percentages as shown for 2018.

1.1 Economic indicators for the internet economy as part of the business economy 2018



Within the core of the internet economy the category Internet-related ICT is dominant. Its contribution to the economic variables of the business economy varies from 8.3% of value added to 6.7% of employed persons. However, the number of businesses in this category is relatively small. Online stores constitute 2.4% of the business in the business economy but contribute relatively little in terms of the economic indicators. The contribution of online stores to the business economy varies from 1.1% of turnover to 0.7% of the production value. The labour productivity[2] for businesses in the category Online services is the highest of the categories of the internet economy.

[1] The "business economy" is a subset of economic activities which relates directly to commerce.
[2] Labour productivity is calculated as the quotient of value added and employment (FTE).

Geographically, the results show that online stores are more often located in the northern half of the Netherlands, roughly above the latitude of Amsterdam. The almost 15 thousands businesses in the internet-related ICT are relatively well represented in and around the cities Utrecht, Groningen, Eindhoven and Delft. Finally, one of the most interesting results relates to the role of the self-employed. We find that 84% of businesses which do not have a website are businesses of the self-employed. We also find that 70% of the self-employed do not have a website.

There are several important conclusions in this report relating to the methodological limitations of this study. Most important are the problems arising from the combination of variability in the big data (internet data) and the method used to delineate the internet economy and its categories. Analyses of our results for consistency at the micro-level show that there is substantial instability in how businesses are categorised over time. This means that care has to be taken when interpreting results, particularly the trends over time. Methodological improvements to alleviate this instability are desirable. This requires further research, but one possible way to deal with this is to adopt a “base-year with revision” approach, similar to that employed by the National Accounts

2. Introduction

Over the past few decades the internet, ICT and digital products and services have provided growth and start-up opportunities for many businesses. It is commonly understood that the economy and the internet are now inseparably linked. The recent proliferation of online stores for example has fundamentally altered the nature of retail and consumer spending habits. New key industries have sprung up around ICT and software creation which fundamentally depend on the existence of the internet to do business. Local community members support their neighbours by sharing goods and offering services on platform websites. It is in fact hard to think of any significant economic activity which does not make use of the internet in some way. Demand is growing to better understand the nature of the internet economy and effects. Further, traditional available statistics fall short of providing the desired insights.

Therefore, a few years ago, Statistics Netherlands in cooperation with Google and Dataprovider conducted a study into combining web-based (big data) sources and more classical statistical sources to get a more complete feeling of the internet economy and its impact in 2015[3]. This was an experimental step towards defining and understanding the internet economy, using the internet economy in the Netherlands as a case study. Dataprovider[4] is a Dutch company which crawls the web and structured and provided the data. This data source gave a unique insight into the structure and contents of the internet and provides many variables which are useful to study the link between the economy and the internet. To do so we linked the Dataprovider data to the Dutch General Business Register (GBR). The GBR contains information with which to determine economic indicators to describe the internet economy.

The strengths of this research lied in the innovative approach and the use of big data in combination with classical statistical sources. One of the limitations was that only the year 2015 was investigated creating a time-series of data on the internet economy may open up possibilities for analysis of trends and for gaining deeper insights in the consistency of the big data approach.
The principle aim of the current study is to investigate the robustness of the internet data and to evaluate the method in terms of its consistency over time.. To do so we consider the years 2016, 2017 and 2018. If the method is suitable for analyzing trends, there are many fundamental questions which we can answer such as, is the internet economy growing faster or slower than our economy as a whole? Are the number of jobs in the internet economy increasing? These insights can become more useful when we look at specific categories of the internet economy. Is the most growth in online services or in online stores? Are there any categories of the internet economy which are performing relatively poorly? The second important aim of this study is to provide such insights.

The method employed in the previous study was transposed as much as possible on to the current study. However, some methodological adjustments were necessary and several

[3] https://www.cbs.nl/-/media/_pdf/2016/40/measuring-the-internet-economy.pdf
[4] https://www.dataprovider.com/

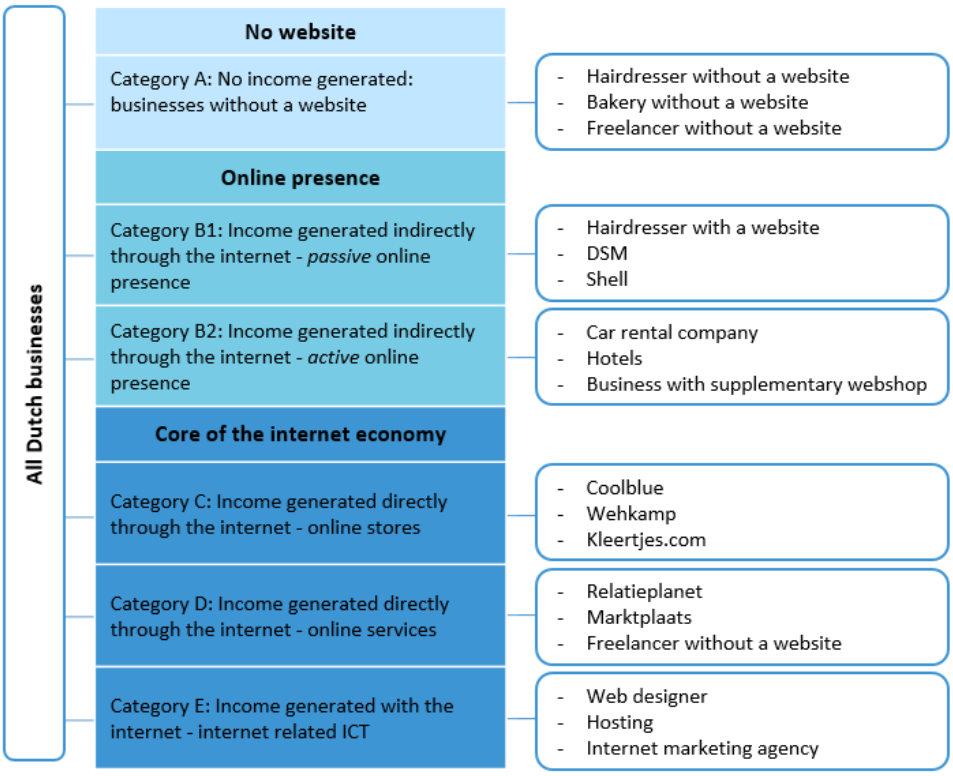
improvements were made. The results of this study for 2016 to 2018 are therefore not strictly comparable to those of the previous study for 2015. Accordingly, the emphasis in this paper is on the results for the period 2016 to 2018.

In this discussion paper, we begin in chapter 3 by explaining our definition of the internet economy and the different categories therein. We then consider the data to which the definition can be applied in chapter 4. In chapter 5, we explain how the data is processed and linked together in addition to explaining how the conceptual definition of the internet economy is operationalized. Chapter 6 presents the results in terms of the demographics of the businesses in the internet economy, the regional distribution of the internet economy, and other economic indicators such as employment and turnover. Chapter 7 discusses the strengths and limitations of the study and gives some recommendations for future research.

3. Definition of the internet economy

This discussion paper adopts a micro-data approach for determining the internet economy. Micro-data approaches in general provide the most detail and the greatest breadth of possibilities for analysis. Further, the Dataprovider data also exist at a micro-level (the website level). We can therefore build our study of the internet economy by combining data sources at the micro-level in order to maximise the analytical possibilities.
The method is based on a categorisation of businesses in terms of their relationship with the internet. This categorisation was developed in 2016 through analysis of Dataprovider data and by consultation with stakeholders (a steering group consisting of representatives of the Dutch government, business-world, academia, Google and Dataprovider) and is shown in Figure 3.1.

3.1. A categorisation of the businesses according to their use of the internet



The categorisation can be considered as a “micro-data” approach. We thus began with the smallest unit (the business) and considered how these businesses, in terms of how they use the internet, could best be grouped together into diverse categories. The easiest category to define is Category A, which consists of businesses without a website. These businesses are as such not considered part of the internet economy. In this way, our definition of the internet economy deviates from other definitions, such as that of the

"digital economy" as employed by NIESR and Growth Intelligence[5] (2013). In that study, a business can be considered part of the digital economy even if it does not have a website.

Categories B through E are all businesses with websites and can therefore, to varying extents, be considered part of the internet economy. The differences between these four categories revolve around how the business generates income in relation to the internet. Category E consists of businesses which make the internet possible. They are the web designers, the hosting companies and the internet marketers. Cloud services and app design among other services fall into this category. If the internet did not exist then these business could fundamentally not exist. This category is therefore referred to as "Internet-related ICT".

The distinction between Category D and E can be loosely understood by considering the nature of the services involved. The provision of cloud services is inextricably linked with the internet and thus cloud services belong in Category E. Dating services existed before the internet. Dating services now make extensive use of the internet: they have in fact become inextricably linked to the internet, but they would still exist without the internet. Another example is the housing market, which is now facilitated greatly by the internet, but would still very much exist without it. This category is therefore referred to as "online services".

Category C consists of online stores, which is defined as businesses with ecommerce activities. While this category seems simple, it is also the category for which the problem of internet dependence for revenue creation is most prominent. For example, bol.com generates all of its sales via internet: it has no physical/high-street shops. We can therefore attribute all of the revenue of bol.com to the internet economy. The Dutch department store Bijenkorf however sells both through its traditional shops and through the internet. Therefore, while many businesses have an online store, they can only be considered partially part of the internet economy. We will explain later how allowance is made for this.

Together, categories C, D and E are considered to be the "core" of the internet economy. Businesses which have a website but do not fall into the core of the internet economy are considered business with an "internet presence". Category B consists of businesses that only make indirect use of the internet to generate revenue. This category is therefore referred to as the "internet presence" category. Generally, these websites provide information about non-internet-related business activities. For example, a consultancy firm uses a website to provide information about its services, publicise its work, place job advertisements and to display information for potential and existing employees.

Given this categorisation of the internet economy, we can now be clear about what is not included in the definition of the internet economy. What is not included is predominantly determined by the available data. Dataprovider can only provide information on the publically available internet. This means that many business-to-business uses of the internet cannot be included in this study. Consumer-to-consumer economic activity can also not be measured. The best example of this is marktplaats.nl (the Dutch equivalent of

[5] National Institute of Economic and Social Research and Growth Intelligence (2013). Measuring the UK's digital economy with big data.

Craigslist or eBay). Marktplaats as a business is included, but it is not possible to consider the transactions between consumers that are facilitated by Marktplaats. As such, our definition excludes consumer-to-consumer transactions.

In the following section, we will introduce the data which allowed us to allocate businesses to a given category and to derive indicators to analyse the properties of the internet economy according to the above definition.

4. Data sources

Obviously, Statistics Netherlands possesses a great deal of data on the economy at the level of businesses. In order to study the internet economy, data on the internet is also necessary. This data is provided by the Dutch business Dataprovider. In short, Dataprovider uses web-crawling to index the internet. This data is then structured and made available to clients. Dataprovider shares this data with Statistics Netherlands. Vitally, this data often contains the Chamber of Commerce (CoC) number of the business which is the subject of the website. The CoC allows the internet data to be linked to data about the business connected to the website. In this way, a dataset is built which combines information on the economy and the internet, and as such provides insight into the internet economy.

4.1 Internet data: Dataprovider

The resulting Dataprovider database which Statistics Netherlands has been provided with is in principal a list of all Dutch websites. Each website is described according to a set of variables including business names, chamber of commerce numbers, shopping cart systems and site traffic estimation, among many others. The data is updated monthly. Dataprovider deals with the important question of whether or not a website is a Dutch website. Dataprovider treats all websites with a .nl Top Level Domain (TLD) as a Dutch website. If a website is hosted on a .com TLD then the following approach is taken. If a .com websites uses the Dutch language it is recorded as a Dutch website. If the websites uses a .com TLD and makes no use of the Dutch language then it is a Dutch website if it is hosted in the Netherlands and displays either a Dutch address or telephone number. According to these decision rules, the Dataprovider web-crawler finds approximately 4 to 10 million websites on the publically available internet in the Netherlands in 2016-2018. This is estimated to represent 95% of Dutch websites[6]. There are three principal reasons why a website of a business can be missed by Dataprovider. Firstly, a website with no links to it cannot be found. Secondly, a business which uses Facebook (or some such website) for its internet presence will not be identified[7]. Thirdly, Dutch websites which use a .com TLD, are hosted from abroad and do not use the Dutch language or have a Dutch address or phone number will not be identified as Dutch. For a detailed list of a the variables in this Database see Appendix A.

Dataprovider also provided us with an additional dataset referred to as the Call To Action (CTA) database, which we use to complement to the main Dataprovider database. The CTA database contains all Dutch websites for which there is at least one way in which a user can, loosely speaking, “interact” with the website. There are seven ways in which a user can interact with a website: order, buy, view the shopping cart, make a reservation/booking, subscribe, register or login. If such an interaction is facilitated by a

[6] This estimate is derived using the database of Stichting Internet Domeinregistratie Nederland, which registers all Dutch domain names (Dataprovider 2016).

[7] Dataprovider could not provide data from Facebook because Facebook does not allow third party web crawlers or indexing to take place on its site.

hyperlink or a button, then this website, and the methods of interaction, are recorded in the database. We use these variables in this research to refine our understanding of the activities of websites. CTA data is only available for the years 2016 and 2019. None the less we employ this data for all three years.

4.2 Economic data

Statistics Netherlands makes use of diverse available data sources to provide data on the economics of the internet economy. The General Business Register (GBR) provides the backbone and structure of the dataset. All the other datasets are used to complement and enrich this backbone. The last step in our methodology is to link all of the enriched datasets to the GBR in order to provide as much information as possible on the nature of the internet economy. This section first describes the GBR and then proceeds to describe the various datasets which are used to enrich the GBR.

4.2.1 General Business Register

The GBR is a database which structures businesses in the Netherlands. A given "business" , roughly speaking, can consist of many smaller businesses, or be subsumed into a larger businesses. When ordinary people think of a business, they are generally thinking of, in Statistics Netherlands terminology, the Enterprise Group (EG). The EG is the top level of business aggregation. The best example is the Dutch business Philips, which consists of many smaller parts that operate independently on a day-to-day basis. These separate parts of the EG are "Business Units" (BUs)[8]. Alternatively, smaller EGs may only consist of one BU. Many of the statistics which Statistics Netherlands possesses are at the BU level. It is important to note that the relationship between the CoC number and the BU is not one-on-one. A BU may have more than one CoC number, and this relationship is made clear in the GBR. The business unit may also consist of multiple Local Business Units (LBUs). BUs in the retail industry often have multiple LBUs, which in that case are simply the multiple shops owned by the same BU. The GBR provides information on the 1) size, 2) sector and 3) age, of BUs.

4.2.2 Production Statistics (PS)

PS provide a picture of employment in, and the financial position of, businesses in the Netherlands. Statistics Netherlands compiles production statistics for the following sectors of industry: mining and quarrying, manufacturing and construction, production and distribution of energy and water, repair of consumer goods, wholesale and retail trade, hotels and restaurants, transport, storage and communication, business and personal services, environmental services and health and welfare. PS are derived from surveying a stratified random sample of businesses which are part of the "business economy" (see box 1). For small businesses (fewer than ten employees) data for the PS are taken from tax data as much as possible. Businesses with fewer than 50 employees receive a questionnaire on a sample basis and businesses with 50 employees or more are all

[8] A Business Unit is a statistical unit that groups all the parts of an enterprise contributing to the performance of an activity at class level (4-digits) of NACE Rev. 1 and corresponds to one or more operational subdivisions of the enterprise. We choose the term Business Unit because it is the most accurate translation of the Dutch term used. The appropriate term according to Eurostat is "Enterprise Unit", which we avoid because of the confusing abbreviation (EU).

included in the survey. The sample size varies strongly per sector of industry, as the number of businesses also varies per sector. Overall more than 80,000 enterprises are invited to participate in the survey, about 10% of businesses in the Netherlands.

4.2.3 Baseline

Baseline complements the GBR with data from the tax office regarding the Value Added Tax (VAT) and profit declarations received from the BU. In combination with the PS, this allows us to determine the following variables at BU level:

- Production value; the value of the goods and services produced, valued at basic prices[9].
- Value added; the value of all goods and services produced ('production value' or 'output'), minus the value of the goods used as inputs to production.
- Employed persons; these are all persons who are working in one or several jobs as employees or as self-employed for a resident institutional unit (company, institution or household). Employed persons include all persons who have a paid job for at least one hour a week.
- Full-Time Equivalents (FTE); the full-time equivalent is obtained by dividing the annual contractual hours of the job by the annual contractual hours considered full-time (in the same company). Two half-time jobs add up to one full-time equivalent.

Because Baseline extracts the data from VAT and profit declarations, only businesses who are obliged to provide this information are in the dataset. For example, companies that work in the supply of water are exempted from taxes on their profits, therefore they are not captured with Baseline. Due to this incompleteness, we only use the data to describe the business economy (see box 1). In addition, the VAT and profit declarations from large and complicated businesses are difficult to merge with the GBR. As a consequence they are occasionally not included in the Baseline dataset. Given the relative importance of some of these companies, we will use data from the PS to include their production, value added and employment.

4.2.4 Turnover Statistics (TS)

Because data in the PS only contains a sample of businesses, the TS is employed to improve coverage. Statistics Netherlands has developed a mixed-source production system that uses VAT data for the smaller units and sample survey data for the largest units to produce quarterly and yearly revenue statistics. This way, turnover is available for very nearly the entire population. The turnover statistics are excluding VAT and include returns from both primary and secondary business activities.

4.2.5 Regiobase

Regiobase is a database designed to provide insight into the geography of economic activity at the regional level. Regiobase contains all the LBUs in the GBR. Additionally, Regiobase contains variables which allow data at the LBU level to be derived from data at the BU level. Regiobase also provides insight into the business activities of the LBUs, as these can differ between the business activity of the BU as a whole. For each LBU, Regiobase contains a postcode, and this allows for data to be analysed and presented spatially. In this study we disaggregate the Netherlands into provinces and COROP-areas. It is important to note that the majority of the statistics in this study are presented at BU level whereas Regiobase employs LBUs.

[9] Basic prices are defined as the prices experienced by the producer. As such product related taxes have been subtracted from the original prices, and subsidies haven been added to them.

4.2.6 Policy Record Administration

The Policy Record Administration is used in this research to derive the number of employees for given BUs. The Policy Record Administration is a record of the employment history of all workers in the Netherlands. The data is collected via income tax returns, and is subsequently processed by the state unemployment insurance provider in order to determine rights to claim money from the state in the case of unemployment. As part of this record, the BUs at which a worker has been employed are also recorded. This allows, at any given time, to derive the total number of employees at a given BU. From this database we derive the number of employees and number of jobs of employees at a BU, whereby an employee is defined as a person who has a contract with an economic unit to carry out work in return for financial remuneration.

4.2.7 Retail Survey

The retail survey provides data on a sample of businesses with more than 10 employees in the retail sector. If the business operates an online store then the turnover from that online store is reported separately, next to the turnover for the whole business. However, turnover from online stores is not reported for businesses of a certain size classes and for certain SIC codes (retail and wholesale). This data in the retail survey is used in the categorisation of online stores.

4.2.8 ICT Use Survey

This survey contains annual data on automation and the use of information and communication technology (ICT) in companies in the Netherlands. The results describe, among other things, the use of computers, the internet, electronic buying and selling, software and ICT applications and show the trends in these phenomena for the period since 2003. The survey is carried out on a sample of roughly 11,000 businesses from the population of businesses with at least 10 employees (which consists of approximately 60,000 businesses). From 2017 the sample has been expanded to companies with a minimum of 2 employees .We used data from this survey to cross-validate our results and improve our methodology.

4.2.9 International trade in goods statistics

These figures provide an overview of the Netherlands' goods trade with other countries. They are of great importance for determining gross domestic product (GDP). In addition, the results provide insight into the trading position of the Netherlands in relation to other countries. The various trade flows and types of goods entering or leaving the Netherlands. A distinction is made between the origin and the destination of the goods. The population of the statistics consists of all enterprises (which have a VAT number and trade abroad) in the Netherlands. For the research data on intra- and extra-EU trade are used. Intra-EU trade refers to goods trade with member states of the European Union (EU). Declaration takes place via the INTRASTAT system. Imports of goods from EU Member States are referred to as intra-Community acquisition (ICA) and exports to EU Member States as intra-Community supply (ICS).

4.2.10 Other sources

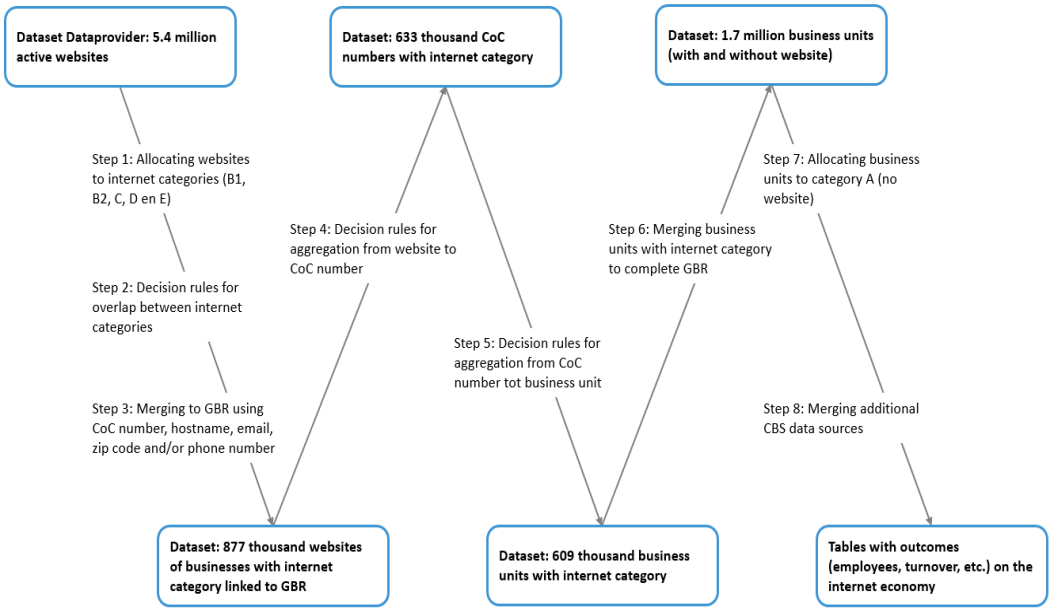
Finally, we make use of several other publicly available sources of data from the internet. For example, where lists of online stores are available online, we use this to check that our method to identify websites is capturing all the most important online stores (see section 5.1.1). We also use lists of ICT businesses for a similar purpose (see section 5.1.2). These

sources allow us to check the plausibility of our results and to complement the other data sources as necessary.

5. Methodology

This section describes the steps to create a database from which all the results are derived. Figure 5.1 shows a schematic summary of all the research steps.

5.1 Overview methodology in research steps, 2018



We begin with the Dataprovider data on Dutch websites and allocate each website to a category. We explain our method for this in section 5.1. After websites have been allocated to given categories, we link the websites to the GBR. The method for linking to the GBR is quite complex and employs diverse methods. For some linking methods more than others, we can be more confident that the link between the website and the BU is correct. Therefore, each linking method is described in detail in section 5.2. At this stage, we have a database in which a BU can have 1) no website, 2) one website or 3) multiple websites allocated to it. This means that the database is not unique at the BU level. In order to accurately represent the economy, this database needs to be unique at the BU level. Further, we need to translate the categorisation of the websites attached to each BU to the classification of the BU. Thus if, for example, one BU has two websites, one belonging to category B and the other belonging to category C, then the BU could be categorised as B or C. To deal with this, we develop a series of decision rules which allow us to create a database which is unique at the BU level and for which all BUs are allocated to only one category. These decision rules are explained in section 5.3. If no website can be allocated or linked to a given BU then that BU is classed as Category A: business without a website. Finally, we link the database to several additional Statistics Netherlands data sources. In this way, we enrich the backbone of BU's with as much information possible, with the aim of maximising the insight into the internet economy.

5.1 Allocating websites to categories

In this section, we describe the methods used to allocate websites to categories. These allocation methods can result in a website being allocated to more than one category. We refer to this as “overlap”. The last subsection explains the rules used to deal with overlap.

5.1.1 Category C: Online stores

Firstly, a point of terminology. Websites can have the functionality generally associated with the term ecommerce. A business with ecommerce functionality may be referred to as an “online store”. Of course, many businesses which engage in ecommerce also engage in other activities which would better describe their core business activities. Whether a business which engages in ecommerce is best classified as an online store is a question which is dealt with in sections 5.3 and 5.4.

Determining whether a website has ecommerce functionality is a relatively simple process because many variables in the Dataprovider database pertain to ecommerce. For commercial reasons, Dataprovider has collected data on, for example, the presence of shopping carts and the presence of different payment methods. Further, Dataprovider constructs a concrete ecommerce indicator. This indicator was derived using a machine-learning algorithm. The machine learning algorithm assigns a probability that a website is an ecommerce website. Thus, some websites have low probabilities (5% for example) and are thus most likely not ecommerce website, while some have high probabilities (95% for example) and are thus most likely ecommerce websites. Dataprovider chooses a cut-off point of 85%. During this research, we analysed this choice by looking at the websites either side of the cut-off point. On this basis, the performance of the machine-learning algorithm and the appropriateness of the 85% cut-off point appeared to be performing satisfactorily. Below the cut-off point most websites don’t seem to be an online store and above the cut-off point most websites were.

It is, however, possible that websites which are actually online stores can be missed using this method. We therefore searched for other data with which we could test whether the Dataprovider machine learning algorithm had identified at least all of the most important online stores. The Dutch website jouwaanbieding.nl contains an up to date list containing the most popular online stores. In addition, Twinkle annually publishes a list of the most important webshops in the Netherlands. Within these lists were several online stores which were better placed in other categories, so we manually removed these from the list. From the remaining websites a few hundred online stores were identified and added to category C.

Finally, we employed the Call To Action (CTA) dataset to further refine the category. The following three calls to action are most closely associated with online stores: order, buy and view the shopping cart. We analysed different combinations of these 3 calls to action in order to find a combination which added a significant number of online stores to the category without adding any websites which were not online stores. The best choice turned out to be 2 out of 3 of these specific calls to action. This choice of calls to action identified another few thousand online stores which were not yet categorised as such.

5.1.2 Category D and E: Online services and internet-related ICT

Category D (Online services) consist of websites which are used as a means to generate revenue without selling goods through what is generally understood to be an online store. Consider for example the case of marktplaats.nl. In this case, the service is a market facilitation: connecting buyers sellers and facilitating transactions. Dating websites are a similar example, as are price-comparison sites, or websites such as booking.com which facilitate the booking of holiday accommodation and hotels. News websites are included in this category because they provide an online service.

Online stores and online services naturally cannot exist without the internet and its related industries, the category E. Traditionally captured by a selection of SIC codes, the supporting industries of internet were often part of the 'digital economy'. With the big data approach in this publication, we aim for a more stringent definition of this type of website. Computer repair is for example not included in our definition of the internet economy. Examples of businesses in category E are host and cloud services, website and app developers and internet consultancy and marketing. Note that many of the services provided by category D could still be provided without the internet. However, category E is entirely dependent on the internet. Without the internet there can be no cloud services or app developers.

The method for allocating websites to categories D and E is fundamentally different from the method used for category C. This is because there are many variables in the Dataprovider dataset which relate to online stores. There are however no variables which can so directly be employed to allocate websites to categories D and E. This is especially the case because categories D and E include many different kinds of services and products. Online services range from dating services to online car auctions, while internet-related ICT ranges from hosting service to online consultancy. This makes a straightforward selection on the basis of one or more Dataprovider variables not applicable. Another approach was therefore needed. The method that has been developed applies a combination of steps which are described as follows:

1. Keyword selection: the indicator 'keyword' in the dataset from Dataprovider captures the words that appear most frequently on a given website. The keywords therefore provide insight into the type of website and the content. As such, if keywords can be identified which relate to a particular category, then the presence of these words in the keywords for a given website can be used to allocate the website to the appropriate category. We started with separating the categories D and E into subcategories with specific topics. For each of these subcategories lists of keywords were created, primarily on the basis of keywords from prominent websites in the subcategories. Combinations of keywords were also used, i.e. the presence of one out of two words would not count as sufficient evidence that a website belongs to a given category, but that keyword in combination with another keyword would count. In Appendix B , the full lists of subcategories and keywords are presented. Using these keyword lists, websites were allocated to the various subcategories with categories D and E. We used the same keyword lists for all three years. The lists from the previous study was evaluated and updated if necessary. A new feature in category D this time is the connection with the research into platforms that is ongoing within Statistics Netherlands. In the past year, Statistics Netherlands drew up a list of more than 1,000 websites that can be identified as a platform, based on various sources. Because

platforms are an online service, this list is included in category D. Many of these platforms are already included in one of the other subcategories, but new platforms will be added to category D. A new feature in category E is the use of a subset of keywords related to cybersecurity. We adopted this subset from another, ongoing (not yet published), research from Statistics Netherlands on cybersecurity businesses. That study used webscraping techniques to identify cyber security businesses in The Hague.

5.1.2 Subcategories for categories D and E

Internet category	Subcategory	Description
D: online services	Leisure	Hotels, Flights, Holidays, Food
	News and entertainment	News, Blogs, Vlogs, Games, Videos, Music, Books, E-learning, Gambling, Adult
	Business	Advertising, Finance, Consultancy, Jobs
	Retail	Housing, Price comparison, Tickets, Online trade
	General services	Dating, Visualisation, Transport, Online payment, Crowdfunding
	Platforms	Platforms n.e.c.
E: Internet related-ICT	Hosting and cloud	Webhosting, Cloud services, Datacenters
	Websites and apps	Web design and developing, App design and developing
	Software	Software products and services
	Marketing and consultancy	Internet marketing and consultancy
	Infrastructure and cybersecurity	Infrastructure, Firewalls, Cyber security
	Datamining and Big Data	Datamining, Big data, Webcrawling

2. Refinement of selection: while the keyword selection succeeded in capturing relevant websites, it also captured many websites which better fitted into categories other than D and E. This occurred because website keywords do not necessarily correctly describe the nature of the website. Therefore, other variables from the Dataprovider dataset and the GBR were employed to refine the selection. From the Dataprovider dataset, we mainly used variables regarding the topic (indicator 'Category') and type of website (indicator 'Websitetype'). From the GBR we employed information about the sector (the SIC codes) and the size of the business (size classes, based on number of employees). The result was a more precise categorisation of websites thanks to the exclusion of websites which were better placed in other categories.
3. Manual adjustments: this is the final stage in the process. Firstly, the websites in the different subcategories were ordered in terms of their "importance" (number of employees). The 100 most important websites for every category were then manually inspected to check that no websites were mistakenly included during the first two steps, and also to check for the presence of the larger well-known websites that should fall into given categories. This identified several areas for improvement which result from incompleteness of datasets and inaccurate links between websites and the GBR. In some cases, the categorisation of a website was reallocated manually. This often

involved visiting the specific websites in order to judge into which category they belonged. In other cases several external sources have been used to determine the completeness of our categorisation[10]. At this stage, we were satisfied with the categorisation for category D, but found evidence that websites belonging to category E were not being identified with sufficient completeness. We dealt with this by identifying SIC codes which fell under the subcategories (shown in table 5.1.3). Websites from businesses with 10 or more employees which fell under these codes according to the GBR were then allocated to category E.

5.1.3 SIC codes in category E

SIC code	Description
4651	Wholesale of computers, computer peripheral equipment and software
4652	Wholesale of electronic and telecommunications equipment and parts
5812	Publishing of directories and mailing lists
62011	Ready-made interactive leisure and entertainment software development
62012	Business and domestic software development
6202	Information technology consultancy activities
6203	Computer facilities management activities
6209	Other information technology service activities
6311	Data processing, hosting and related activities
6312	Web portals
70221	Financial management
7311	Advertising agencies
74901	Environmental consulting activities
74902	Quantity surveying activities
74900	Other professional, scientific and technical activities n.e.c.

5.1.3 Category B: online presence

Categories C, D and E constitute the "core" of the internet economy. The remaining websites are conceptualised as having an "online presence". This means that the website's principal function is to provide information on the business behind the website. The website may also provide services associated with categories C, D and E but these services are in principal additional services to the central purpose of the business. For example, Royal Dutch Shell has a website which provides information on the business but clearly the business should not fall into categories C, D and E.

Other websites do not fall into categories C, D or E but still provide somewhat more functionality than simply information provision. In the simplest case, one may be able to subscribe to a newsletter or to emails which provide information on new products. Another example is large companies which also manage an online store. Consider the case BMW, who have a small online store in the Netherlands selling BMW merchandise, but this

[10] The website appspecialisten.nl was used for its list of software developers who create apps. We also acquired a list of ICT businesses from the "MKB Innovatie Top 100", which list the top 100 best ICT business.

in no way means that BMW can be considered an online store. We therefore split category B into Categories B1 and B2, whereby B1 is termed “passive online presence” and B2 is termed “active online presence”.

To determine whether a website is categorised as active online presence (B2) we make use of the CTA dataset. If a website has one of the following calls to action, "order", "subscribe", "buy", "shopping cart", "register", "book", "login", then we can consider the website as having an active online presence if it has not yet been allocated to categories C,D or E. Category B1 is not dealt with at website level. Instead we determine Category B1 at the BU level. As such, we first link the categorised lists of websites to the GBR (as described in section 5.2). All websites which linked to businesses which had not yet been assigned a category, were then assigned the category B1. The definition of B1 is thus in practise: websites which are not allocated to any other category. These are thus websites outside of the core of the economy which provide purely an online presence to the businesses behind the websites.

Category B1 was also used to deal with problems associated with the businesses with the most complicated structures. These businesses are referred to as TopX and consist of businesses where the Enterprise Group (EG) consists of multiple BU’s. These complications result from the fact that websites are linked to businesses via the CoC number, and the CoC number is generally only linked to one BU. However, the website for that single BU, was often found to be more closely related to the whole EG rather than the specific BU. Conceptually then, all the BUs under the EG make use of the website even if the website is formally only linked to one BU. We therefore do not wish to assume that all the other BUs are not part of the internet economy. In order to include these other BUs in the internet economy, we allocate them to Category B1. The categorisation of the BU which links directly to the website is determined by the methods described above. Thus, for all TopX EGs, one BU is allocated to a category and all other BUs are automatically allocated to B1.

5.1.4 Dealing with overlap

Of all the websites, about 2% were allocated to more than one category within the core of the internet economy. This occurs because these websites have characteristics which are associated with multiple categories. Because we wish for a website to be allocated to only one category, we formulated a series of decision rules to remove the overlap.

In order to remove the overlap, the overlapped websites were analysed at the level of subcategory in order to understand the principal causes of the overlap. By analysing the principal causes, appropriate decision rules were formulated to allocate overlapped websites to a unique category in the most appropriate way. Often additional Dataprovider or GBR variables were employed in the specification of the decision rule. As an example, Dataprovider construct a variable named “category” that provides some general information on the type of website. Some websites have the category “IT-services and Telecom”. We found that websites which were allocated to both C and E could be best placed in category E if the website was categorised as “IT-services and Telecom” according to Dataprovider.

5.1.4 Percentage of websites by type of overlap, 2016-2018

	C, D and E	C and D	C and E	D and E
	%			
Year				
2016	0.0	0.7	0.8	1.0
2017	0.0	0.6	0.7	0.7
2018	0.0	0.6	0.8	0.7

5.2 Merging to General Business Register (GBR)

By this stage, every website is allocated to a category. The next step is to link websites to a BU. This allows the economic data available in Statistics Netherlands and the internet data from Dataprovider to be combined and thus analysed together at the BU level. In order to link two datasets, a key is required. A key is a variable which is present in both datasets to be linked. In this project we make use of various combinations of keys into order to maximise the number of successful links which can be made. The degree of confidence whether the correct link has been made varies between key combinations. We therefore describe in detail the different key combination which we have used.

Key combination 1
The first key combination consists of website names and CoC numbers. Many websites in the Dataprovider data have a corresponding CoC number which was obtained by web-crawling. Additionally, many BUs in the GBR record the website of the BUs. The fact that the link is made on the basis of both CoC numbers and website names, indicates a very high linking accuracy. This is referred to as Key Combination 1.

Key combination 2
Key combination 2 is used when the website names match in the GBR and in the Dataprovider data but the CoC number in the Dataprovider data does not concur with the CoC number in the GBR which is indicated by the website name. To solve this problem, we look at the telephone and email address data from Dataprovider and with the GBR. When the analysis of the telephone and email address indicates that the CoC in the GBR is the correct one, we assume that the website links to the business corresponding to the CoC number in the GBR.

Key combination 3
For some cases there are no logical links between the CoC number and hostname from Dataprovider and that from the GBR. Examples are when there is no hostname in the GBR available and/or there is no CoC present in the Dataprovider data. In this case, the best solution is to use the CoC number derived from the email and telephone details in the Dataprovider data. These email and telephone details are matched to those registered in the GBR to allocate the correct CoC to the website.

Key combination 4
Key combination 3 refers to a cases where there are no logical links. Key combination 4 refers to the more difficult case where a logical link is possible according to the CoC

number, but the hostname in the CoC is not the same as the website from which the CoC number was scraped. These cases are included only if email, telephone and postcode support the link made on the basis of the CoC number.

Key combination 5
This key combination contains two groups. The first group consists of cases where a link can only be made according to the CoC number and postcode. In other words, these are cases where neither the hostname, email or telephone number can be used to link the website to the GBR. The second group consists of cases where more than 10 websites have the same CoC number. Some businesses may indeed have more than one website, for example, there are several online stores which each have several websites selling different products. In many cases however, more than 10 websites link to the same CoC number because the CoC number on a given website does not correspond to the business behind the website. For example, a website may contain a list of businesses who supply a particular product, which includes the CoC number for each business. A given CoC number can show up multiple times over different websites. Another example is of hosting/web-design companies which display their CoC on the websites they host or designed.

All websites which could not be linked according to one of the 5 key combinations are from this point on excluded from the analysis. The websites which could not be linked will include all the websites of private individuals which fall outside of the definition of the internet economy, as well as websites where there was not sufficient information to make a link to the GBR. The results of the merging process per year are shown in table 5.2.1

5.2.1 Results of the linking process in terms of the number of websites which can be linked to the GBR

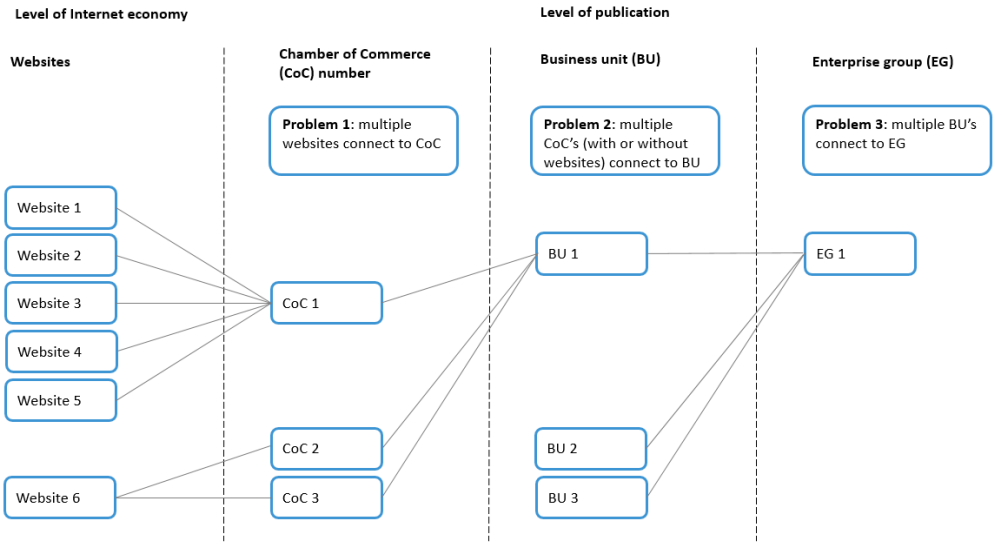
	2016		2017		2018	
	Count	%	Count	%	Count	%
Key combination						
1	249 000	31	227 000	27	211 000	24
2	75 000	9	115 000	14	182 000	21
3	18 000	2	15 000	2	19 000	2
4	196 000	24	173 000	20	168 000	19
5	276 000	34	320 000	38	297 000	34
Total	815 000		850 000		877 000	

The results of the merging process show some interesting trends over time. We see first that the percentage merged with key combination 1 drops from 31% (2016) to 27% (2017) and then to 24% (2018). This is probably due to businesses not updating their website information with the CoC. As a consequence of this key combination 2 is used more frequently over time. The use of key combination 3 remain relatively constant over time. The percentage of websites merged with key combination 4 decreases slightly over the years.

5.3 Decision rules

Decision rules are formulated to determine the role of BUs in the internet economy based on the information which we have gained about the website(s) of the BU. Decision rules would not be necessary if every business had only one website, only one CoC number and only one BU. The situation can be more complicated because businesses often have multiple websites. Consider for example a large business with separate websites for customers, businesses, careers and sales. It is likely that these websites will not appear in the same category of the internet economy. Additionally, due to the structure of GBR, multiple CoC numbers often link to a single BU. Subsequently, multiple BU's can link to one EG when large companies have diverse activities. The following figure presents these different problems.

5.3.1 From internet economy to level of publication; problems



5.3.1 From internet economy to level of publication; problems

Our aim was to create a list of all the businesses in the GBR and allocate every BU to a category of the internet economy according to the website(s) belonging to that BU. It was therefore necessary to develop a set of rules to counter the problems regarding the translation from website to BU level.

There are two rules which are applied to deal with these problems;

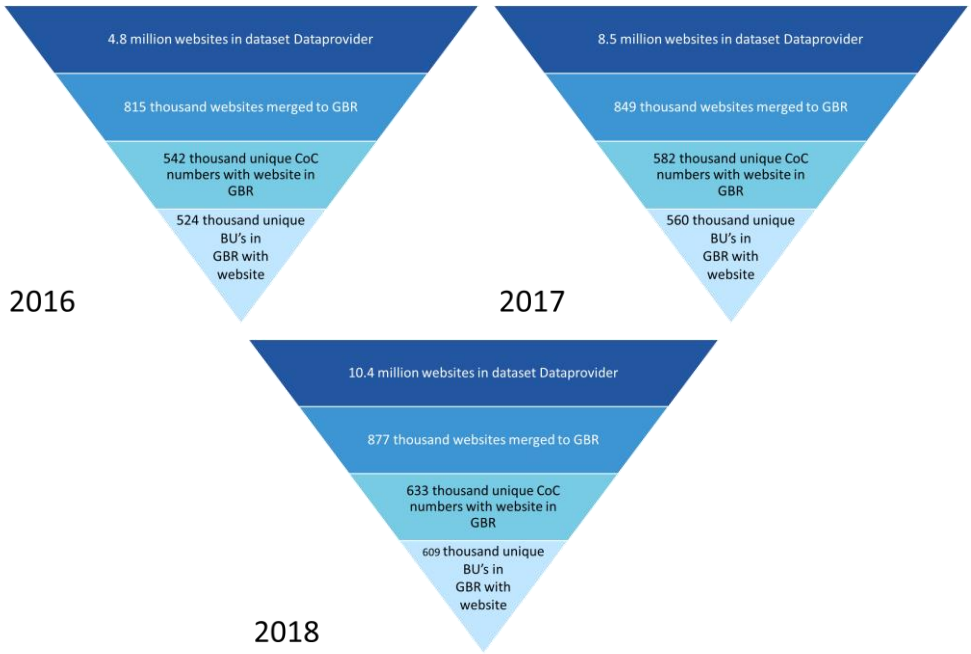
1. When a website is allocated to multiple categories, the categories within core of the internet economy (C, D and E) are always preferred to the category Online Presence (B). We choose this hierarchy in order to maximise the information provided by allocating a website to a category. Consider for example a website that could be allocated to D or B. If we allocate it to B then we know less about that BU because category B is more general.
2. When a BU has multiple websites and/or CoC numbers and therefore is allocated to multiple categories of the internet economy, decision rules are employed depending on the categories in question. Decision rules in general give preference to the categories with the BU whose website:

- has been merged to the GBR with the highest accuracy;
- links to a CoC number with the largest number of employees;
- has the highest average economic footprint (a Dataprovider variable).

An additional problem is that some BUs which fall under large and complex EGs may not be allocated to a category even though other BUs under the EG have been allocated to a category. This is a problem because the BUs which have not been allocated are likely to make use of the website of another BU within the EG. Consider a large company. A given BU may make use of the website of another BU to recruit staff. On this case, we therefore change the BU's with no website (category A) into online presence (category B1). The absence of a website for these BU's is the result of our link between websites and CoC's, i.e. it is not possible to link one website to multiple CoC's. A website linked to another BU in the EG is therefore most likely to represent the business activities of the remaining BUs.

Finally, the largest 100 business in each category (according to number of employees) were manually inspected to confirm that the decision rules were functioning appropriately. The process up to now is summarized in the following figure.

5.3.2. Summary of method



5.4 Turnover of online stores

At this stage, a list of online stores (at BU level) existed. This list however included businesses categorised as category C (online stores) which may only produce a small amount of their turnover from their online store. It was therefore necessary to ask whether such businesses can really best be classified as online stores. A logical approach, which we adopt in this study, is to classify online stores as such if more than 50% of their turnover comes from the online store. Applying this decision rule is however challenging

because the percentage of turnover generated by ecommerce is only available for the small sample of business in the Retail Survey, ICT Use Survey and Production statistics. Further, this is not a representative sample of businesses because it almost exclusively consists of businesses with 10 or more employees. We therefore wish to estimate the revenue from ecommerce for businesses which have an online store, and for which these data are not available from these surveys.

To do this, we adopt regression analysis. This method is employed to predict the turnover from ecommerce for businesses which are not in the Retail Survey, ICT Use Survey and Production statistics. The method attempts to explain variations in turnover from ecommerce with other variables about the business. The first variable which we consider is the total turnover of the business. In general, if total turnover increases, then supposing a fixed portion of turnover from ecommerce, the ecommerce turnover will increase also. We therefore include the variable total turnover to control for this effect. We then experimented with many other available variables (from Dataprovider and the GBR among others) to try to explain as much of the variation in turnover from websites as possible. We identified the size of the businesses behind the ecommerce website, the economic footprint, and an 'order button' on the website to all be significant determinants of turnover from ecommerce.

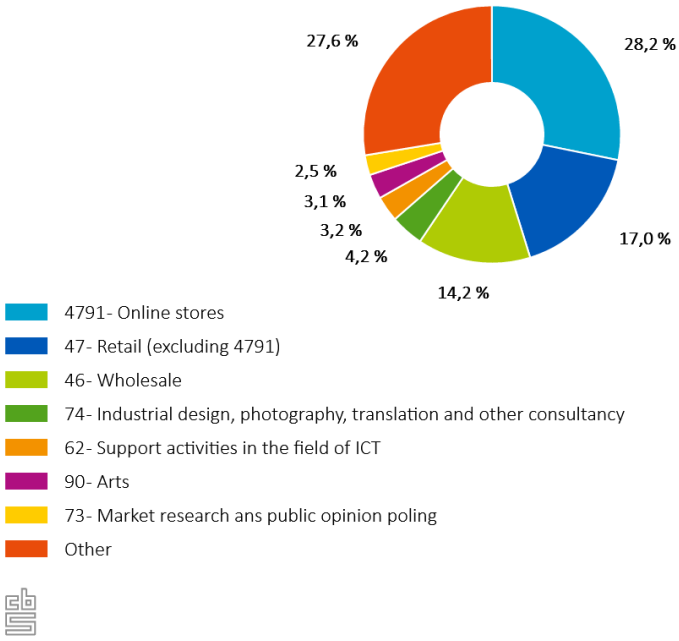
For each of these variables, the regression analyses produces coefficients. These coefficients show the direction and magnitude of the effect of the corresponding variables on ecommerce turnover. We can thus employ these coefficients and variables to estimate the turnover from ecommerce for businesses for which actual statistics are not available. However, we found that the model was not producing plausible results for the majority of small business. This is to be expected because the data upon which the regressions were run did not include many businesses with fewer than 10 employees. We therefore visited the websites of a sample of small businesses with websites to search for evidence that these businesses were generating revenue through other means than ecommerce. We found that the majority of the majority of these businesses had only ecommerce as means of generating turnover. We therefore simply assumed that all businesses with ecommerce websites with fewer than 10 employees generate at least 50% of their turnover from ecommerce. We therefore classify these businesses as online stores.

Online stores in the internet economy compared to other research

SIC codes

The use of big data within this project facilitates a different perspective on the nature of the Dutch businesses. It is particularly interesting to make comparisons between the nature of businesses according to the SIC codes and the categorisation of businesses in this study. The SIC is the most common way to publish about a company's economic activity. In this box we analyse the SIC codes of all the businesses which are classified as online stores according to our definition. The results are shown below for the year 2018.

5.4.1 SIC codes of businesses classified as online stores, 2018



The results show that almost one third of the online stores in het internet economy has the SIC code for online stores (28%). We also find that there are many businesses in SIC codes other than 4791 (online store). Unsurprisingly, a large number of retailers who are not online stores according to their SIC code, are online stores according to our definition. This can easily occur if a retailer decides to diversify into online sales. Several of the other SIC codes can also have retail activities (46, 62, 74, 90) alongside their other business activities. Overall, the data show that the SIC codes are not capturing the extent to which online stores are an important part of the economy. On the other hand, the SIC code 4791 also identifies businesses as online stores that are not part of our definition. For example our study misses businesses that do not have their own ecommerce website but sell their products through websites such as Amazon or another platform.

Research online retail trade

The main aim of this 'The Internet Economy'-project is to give a broad overview of the entire internet economy in the Netherlands, and to provide an estimate of its total economic value. In a related, but different project, entitled 'Online Retail Trade', Statistics Netherlands focuses on a small, but economically relevant, aspect of the internet economy, namely the online retail sales of goods to Dutch consumers. The aim of the latter project is to provide a more precise and fine-grained insight into the vastly growing online trade flow from business to consumer. As the economic scopes of the projects intersect, but the aims of the projects differ, the statistical results of the two projects might seems conflicting to users. Below, we clarify why this might seem so, and how to interpret the results of both projects in a coherent way.

First, the intersection of the economic scope of the two projects is as follows. 'The Internet Economy' describes the 'Online Stores' in the Netherlands and their total economic value. The project 'Online Retail Trade' describes the 'Total Online Retail Trade Turnover' of Dutch

companies. The different aims of the two projects naturally result in the following major methodological differences:
1. In the 'Online Retail Trade' project the estimated turnover solely refers to the turnover that is generated online. 'The Internet Economy' includes all of the turnover of a company if that company is classified as an online store.
2. In the 'Online Retail Trade' project the estimated turnover is limited to sales to consumers. 'The Internet Economy' also includes business-to-business sales.
3. The populations of both projects differ. The use of different (combinations of) data sources leads to different populations of businesses.

Consequently, these methodological differences result in a different estimate of the total economic value of 'Online Stores' in 'The Internet Economy' and that of 'Total Online Retail Trade Turnover' in 'Online Retail Trade'.

5.5 Consistency of the method

The previous sections gave more insight in the data sources and method used in this study. The method and data sources used in this update are, with some exceptions, the same as the original study in 2016. When we developed the method in 2016 we only had one year of data (2015) available, so there was no possibility to look into the consistency of the method and the results over multiple years. Now, we have a time series of three years and one of the objectives is to gain more insight in the consistency of the big data approach of the study. This section discusses the outcomes of these analyses.

Note: because we made some minor methodological adjustments and improvements it is not possible to make a good comparison between the results on 2015 of the previous study and the results on 2016-2018 of the current study. Therefore, we focus in this section on the consistency of the results for the years 2016-2018.

5.5.1 Dataprovider data

The algorithms used by Dataprovider to crawl the internet have not significantly changed over the years. Dataprovider has expanded its dataset with new variables but the set of variables necessary for our study was not subject to any changes. However the internet itself does change. In turn, the route which web crawlers take from website to website changes, and the resulting index of the internet is therefore also subject to variation over time. This is because every month Dataprovider starts the web crawling process all over again.

Table 5.5.1. indicates what that means for the population of Dutch websites in the Dataprovider dataset. This population has grown substantially between 2016 and 2018, from 4.8 million websites in 2016 to 10.4 million in 2018. Apparently, the number of websites classified as belonging to the Netherlands has increased significantly. The reason is that Dataprovider does not know if a website (hostname) is still registered with an valid IP-address. These websites remain in their dataset but are not active, so the increase is mainly due to the non-active websites. In addition, these number are not unique websites, as there is some duplication in the datasets.

5.5.1 Number of websites in the Dataprovider dataset, 2016-2018

	2016	2017	2018
	Frequency		
Websites	4 784 000	8 535 000	10 396 000

To get an idea about how consistent the population websites is over time, we looked at the overlap between the unique (active) websites for these three years (table 5.5.2). In other words, how many websites in one year also occur in the other years? There are 2.6 million websites that existed in all of the three years. Further:

- around 65% of the websites in 2016 also occur in 2017,
- around 60 percent of the websites in 2016 also occur in 2018
- around 80 percent of the websites in 2017 also occur in 2018.

This indicates that although the majority of the websites can be found in multiple years there is a significant amount of new or changing websites. Of course, these are not only websites from companies. It is expected that company websites are more stable than the personal websites that are made as a hobby for example.

5.5.2 Overlap of unique active websites in the Dataprovider database, 2016-2018

	2016	2017	2018
	Frequency (%)		
2016	4 783 000 (100%)	3 073 000 (64%)	2 903 000 (61%)
2017	3 073 000 (70%)	4 405 000 (100%)	3 514 000 (80%)
2018	2 903 000 (53%)	3 514 000 (65%)	5 440 000 (100%)

In addition to the consistency of the population of Dutch websites, it is also important to know the consistency of the data within the important variables. The linking variables CoC number, postcode, email, and telephone number are important for the merging process to the GBR and the variables on keywords and ecommerce are necessary for the definition of the internet economy.

If a website has a different CoC number, postcode, email and/or telephone number in 2016 compared to 2017 it is likely that this website will be linked to a different business unit in de GBR in 2016 than in 2017. Or, at worst, can't even be linked with the GBR. This is not a problem if it is a real development (for example, a company has been taken over by another company). However, this can be a problem for the consistency of results over the years when it mainly reflects the changeability of the internet data. On the other hand, a lack of changeability of the information in the GBR (which is largely based on input from the CoC) can also be a problem. How many businesses report (immediately) to the CoC when they change their website, telephone number, or email? They probably update their website first.

Table 5.5.3 shows how often the same website has changes for the variables CoC number, phone number, email, postcode or keywords over the three years. For the majority of the websites the linking variables did not change. At first sight, these outcomes are not worrying and seem valid. However the effects of small changes can be magnified in this research because the linking process is based on a combination of these variables.
The content of the variable Keywords changes a lot. More than half of the records has a different set of keywords every year.

A note should be made that these numbers only include non-blank records for these variables. So when a website has a missing CoC number in one of the years, that website is not represented in these numbers, even though this can be considered a change.

5.5.3 Changes in Dataprovider variables for the same hostname, 2016-2018[11]

	CoC number	Postcode	Phone number	Email	Keywords
	%				
No changes	72	81	82	94	16
Different in 1 year	16	12	12	4	33
Different every year	11	7	6	2	51

5.5.2 Consistency of the merging process

The process of merging a website to a business unit in the GBR is complicated because there is no 1-to-1 relationship between a website and a business. In addition, the linking data that is available is not the same for every website or business and can change over the years. This introduces a level of uncertainty that is illustrated by the quality of the different key combinations of the merging process (sector 5.2). There is no guarantee that a website that can be linked to a specific business unit in the GBR one year, will be linked to the same business unit another year or will be linked to the GBR at all.

Table 5.5.4 shows how many unique websites can be linked to the GBR every year. This number increases from 815 thousand in 2016 to 877 thousand in 2018. Just over three-quarters of the linked websites in one year can be linked the GBR the next year and 502 thousand websites can be linked to the GBR in all three years. This means that around 60%of the websites in one year can also be linked to the GBR in the other two years.

[11] These analyses only included non-blank records for these variables.

5.5.4 Overlap of unique websites linked to the GBR, 2016-2018

	2016	2017	2018
	Frequency (%)		
2016	815 000 (100%)	618 000 (76%)	541 000 (66%)
2017	618 000 (73%)	850 000 (100%)	663 000 (78%)
2018	541 000 (62%)	663 000 (77%)	877 000 (100%)

Another question concerning consistency is whether the website is linked to the same CoC number in the GBR every year. This appears to be the case for the majority of the websites. When a website is linked to the GBR in multiple years, in about 95%of cases that website is linked to the same CoC number. The issue of the consistency of merging a website to the GBR is more an issue of whether the website is linked to a CoC than to what CoC number.

In addition to the process of linking a website to a CoC number in the GBR, there is the set of decision rules to allocate a category of the internet economy to a business unit when there is no 1-to-1 relationship between a website and a business unit or when a business unit is allocated to multiple categories (sector 5.3). These rules are the same every year, but that does not necessarily mean that this leads to the same classification every year. That depends on variables such as website category, economic footprint, the quality of the match or the number of employees.

5.5.3 Consistency of the definition of the internet economy

The effects of the consistency of the Dataprovider data combined with the effects of the consistency of the merging process to the GBR and the decision rules influences the consistency of the definition of the internet economy over time. One of the strengths of this research is that we classify companies based on recent information that can be found on their website. If the economic activities of a company change and the company changes their website we want that to be reflected in classification of that company to one of the internet economy categories. So we do not expect that every business unit belongs to the same internet category every year. And probably businesses in the core of the internet economy change faster than most other businesses due to the rapidly changing nature of the internet.

On the level of the business unit, we cross-tabulated the category of the internet economy of one year with the category of the next year. Table 5.5.5 and 5.5.6 show the outcomes. For the categories outside the core of the internet economy (A, B1 and B2) more than 80% have the same category the next year. For the categories within the core of the internet economy (C, D and E) this percentage is lower. This is especially the case for category D (online services) for which only about 50%of the business units are also category D the next year. One important reason for this is that the allocation of businesses for this category is almost exclusively based on the keywords variable of Dataprovider. And the content of this variable is very changeable as was shown in table 5.5.3.

5.5.5 Business units by internet category in 2016 and 2017

	2017						
	A	B1	B2	C	D	E	Not in 2017
	%						
2016							
A: No website	82	6	1	0	0	0	11
B1: Passive online presence	12	83	2	1	0	1	2
B2: Active online presence	8	4	82	3	1	1	2
C: Online stores	7	9	14	69	0	1	2
D: Online services	7	20	14	2	49	4	3
E: Internet-related ICT	6	16	7	1	1	66	3

5.5.6 Business units by internet category in 2017 and 2018

	2018						
	A	B1	B2	C	D	E	Not in 2018
	%						
2017							
A: No website	82	5	1	0	0	0	11
B1: Passive online presence	9	87	1	0	0	1	2
B2: Active online presence	5	4	85	3	1	1	2
C: Online stores	7	10	13	68	0	1	2
D: Online services	6	19	14	2	53	4	3
E: Internet-related ICT	4	16	6	1	1	69	3

5.5.4 Conclusions

The results of these analyses indicate that the consistency of the results over the years is questionable. The number of companies ultimately classified in the various categories of the internet economy has remained fairly stable over the years. However, relatively often a company changes category from one year to the next, especially within the core of the internet economy. The group of companies that belong to a certain category one year can be significantly different from the group of companies that belong to that category in the other year(s). We partly corrected for this by manually reviewing and where necessary reclassifying the 100 largest companies in each category of the internet economy every year. To prevent misrepresentation of the largest companies from distorting the results of indicators such as turnover, added value and employment in particular. However, this shows that we have to be careful when interpreting results about the trends in the internet economy.

It appears that the combination of the variability of the big data (internet data) and the chosen research method poses certain limitations on the production of time series and

analysis of trends. In chapter 7 we will make some recommendations for further research in response to these results.

6. Results

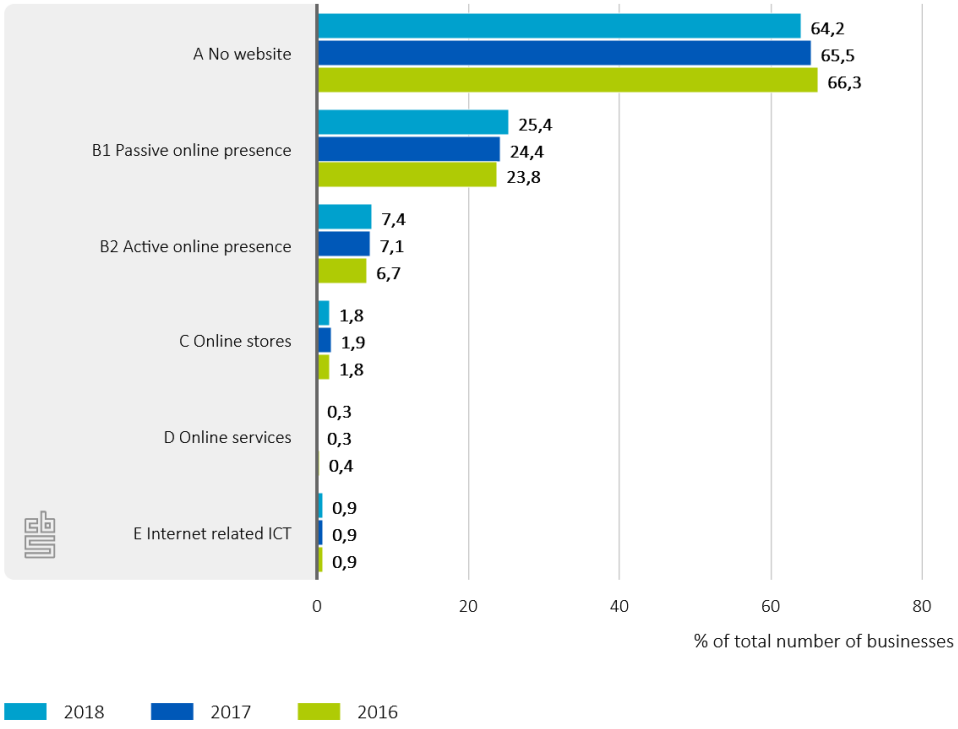
The application of the above methodology to the data facilitates the analysis of the internet economy. Compared to our first discussion paper on measuring the internet economy for the year 2015 only, we now have results for three years in a row which facilitates an analysis of trends. As explained in section 5.5 in some cases developments seem to be influenced by a lack of robustness of the data and methodology used. Nevertheless we will analyse the trends in the data in as far as they seem plausible.

6.1 General results

3.0% of Dutch businesses belong to the core of the internet economy

The starting picture is the number of Dutch businesses that have a website, and if they do, into which category that website falls. In 2018 we find that 35.8% of the Dutch businesses have a website, as shown in Figure 6.1.1. This percentage can be seen to increase every year. This is consistent with what one might expect: an increasing number of businesses with presence on the internet through their own website.

6.1.1 Number of businesses by internet category, 2016-2018



In 2018 3.0% of all Dutch businesses belong to the core of the internet economy (categories C, D and E). In 2016 this percentage was 3.1%. In absolute terms, the number of enterprises in the core of the internet economy increased from 48 thousand in 2016 to 50 thousand in 2018. Because the overall number of businesses increased as well, the

share held by the core of the internet economy did not increase. The number of businesses in the category Online presence, for instance, increased persistently over time; the businesses with a passive online presence as well as businesses with an active online presence. For the absolute figures and the breakdown by size-class, see table 1 in Appendix C.

Demographics of Dutch businesses

It is important for a proper interpretation of the results to have a complete picture of the Dutch business population. There are large differences in the number of businesses across different sectors and size-classes. For example there are over 1,3 million businesses with only one employee. There are also big differences between sectors, from several hundred companies in the sector Mining and quarrying to more than 300 thousand companies in the sector Other specialised business services. These differences influence our outcomes to a great extent e.g. analysis based on the number of businesses will be dominated by the smallest businesses because they greatly outnumber the larger businesses.

6.1.2 Demographics of Dutch businesses, 2018

	Size-class (number of employees)				
	1	2-49	50-249	250 or more	Total
Total	1.314.530	374.470	10.830	3.010	1.702.840
A Agriculture, forestry and fishing	36.790	37.150	150	10	74.100
B Mining and quarrying	320	110	30	0	460
C Manufacturing	41.300	21.330	1.960	400	64.990
D Energy supply	910	300	20	20	1.250
E Water supply and waste management	1.160	590	80	30	1.860
F Construction	147.130	28.460	730	100	176.410
G Wholesale and retail trade	139.170	87.970	2.060	350	229.550
H Transportation and storage	28.870	13.930	740	150	43.680
I Accommodation and food serving	24.380	32.290	370	60	57.090
J Information and communication	75.300	15.590	540	110	91.530
K Financial institutions	81.860	10.420	160	90	92.530
L Renting, buying, selling real estate	18.620	7.480	150	20	26.270
M Other specialised business services	297.810	43.920	790	160	342.680
N Renting and other business support	53.610	18.120	1.220	400	73.340
O Public administration and services	170	90	250	210	720
P Education	76.030	7.870	580	310	84.790
Q Health and social work activities	123.260	25.300	700	560	149.820
R Culture, sports and recreation	87.580	10.840	180	20	98.610
S Other service activities	80.250	12.730	130	20	93.130
T Activities of households	10	10	0	0	20
U Extraterritorial organisations	0	0	0	0	0
Business economy (total excluding NACE categories A, K and O-U)	828.570	270.080	8.680	1.780	1.109.120

The number of businesses in the business economy is added because a number of variables refer only to the business economy and not to the economy as a whole. The business economy can be characterised as those businesses or branches of industry which in general are market producers (see the other box in this paragraph). The table below presents the

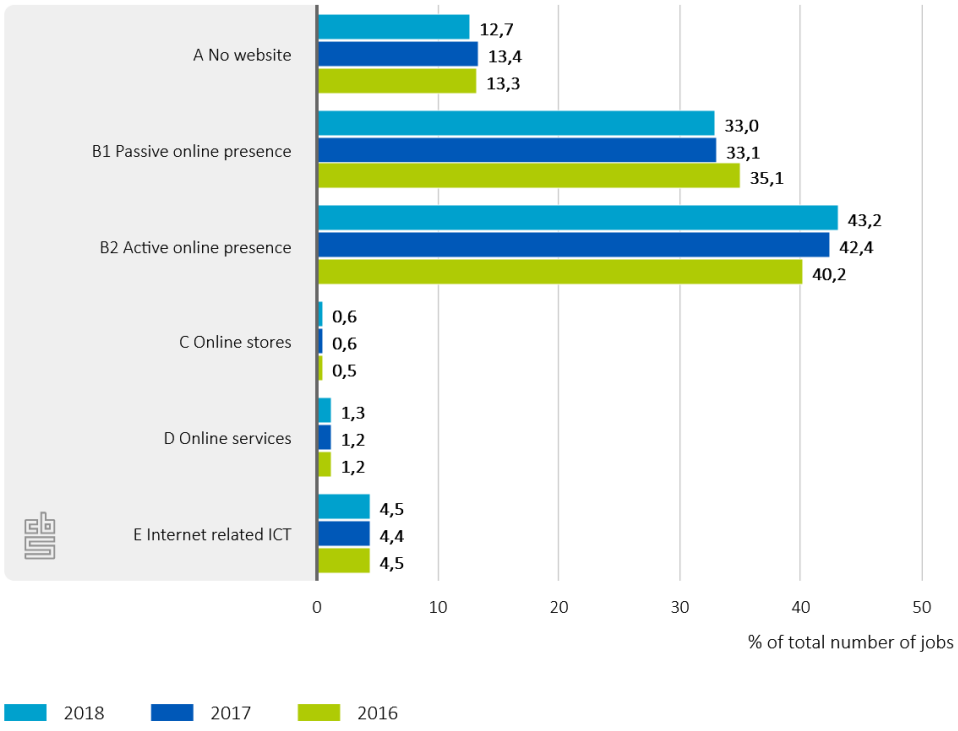
demographic distribution of all Dutch businesses across sectors and size-classes and can be helpful when interpreting the results.

6.5% of jobs related to the core of the internet economy

As illustrated in the box on the demographics of Dutch businesses there are many more smaller businesses than larger businesses. Figure 6.1.3 shows the number of jobs of the different categories of the internet economy. Compared to their share in the number of businesses, the businesses without a website are with a share of 12.7% in 2018 (13.3% in 2016) much smaller in number. The share of the core of the internet economy in the total number of jobs in 2018 is 6.5%, which is much more than its share in the number of businesses. Within the core of the internet economy in 2018 the category internet-related ICT has with 4.5% the largest share in the total number of jobs. Over the years, the share of the core of the internet economy in the total number of jobs increased from 6.2% in 2016 to 6.5% in 2018. At the same time, we see a small decrease in the share of the businesses without a website or only a passive online presence (comparable to the number of businesses). At the end of the day the number of jobs belonging to the businesses with some kind of web presence fluctuates around 87%.

For the absolute figures and a breakdown by size-class see table 6 (jobs) and table 5 (employees) in Appendix C.

6.1.3 Number of jobs by internet category, 2016-2018[12]

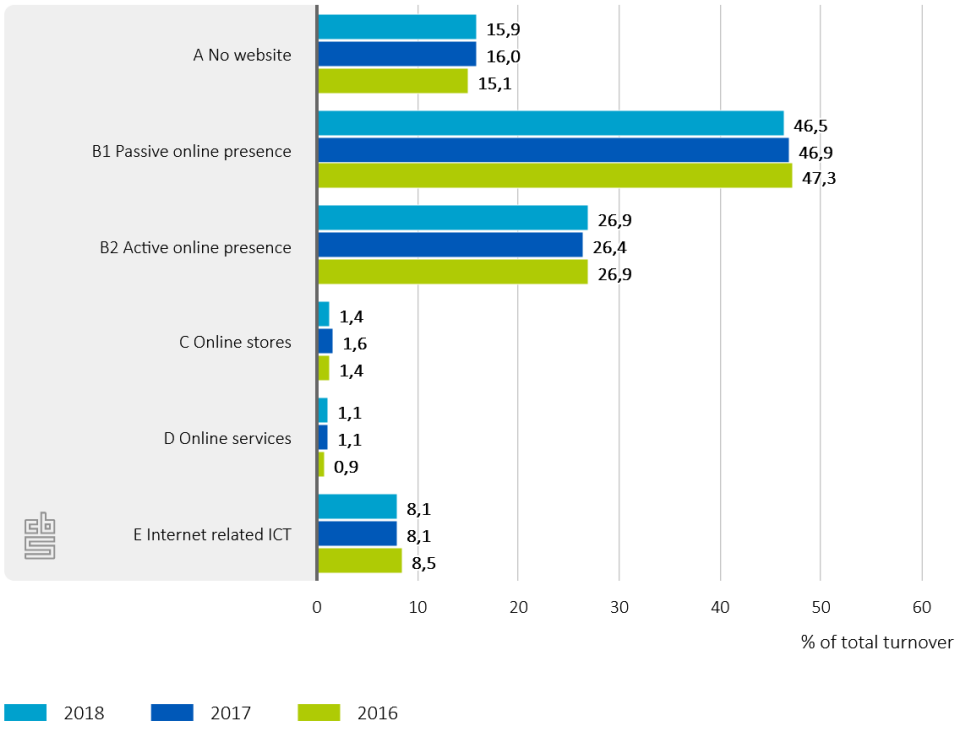


[12] Categories do not add up to 100 per cent because not every job could be matched to a company.

Contribution of the internet economy to main economic variables

So far, for the different categories of the internet economy the number of businesses and jobs were presented as part of the whole economy. Now, we will present the contribution of the different categories of the internet economy to the main economic variables of the business economy (see the box on the business economy in this section).
Figure 6.1.5 shows that in 2018 the core of the internet economy contributes for 10.6% to total turnover of the business economy (10.8% in 2016) with internet-related ICT being the largest category within the core of the internet economy contributing for 8.1% to total turnover of the business economy (8.5% in 2016). The turnover represented by businesses with some kind of web presence in 2018 was 84.1% (84.9% in 2016). The categories passive and active online presence constitute by far the largest categories in the business economy. Evidently, not all businesses can be online stores or internet-related ICT businesses. And not for all businesses it makes sense to communicate with their customers via a website. Some of the businesses with only an online presence are the large industrial businesses which do have a website for PR-reasons and other practical reasons, but not to sell their products or communicate with their customers on a large scale. Accordingly, the turnover of the business economy by internet category does not show a clear pattern of categories growing every year and others persistently declining. Only the share in total turnover of the businesses with a passive online presence declines systematically, but only with very small percentages. See table 4 in Appendix C for the absolute and detailed figures on turnover.

6.1.5 Turnover of business economy by internet category, 2016-2018



The business economy

The "business economy" is a subset of economic activities which relates directly to commerce. In this study, we present some results for the business economy and some results for the economy as a whole. Specifically, results in terms of turnover, value added and production[13] are presented for the business economy. One important reason for this is that only businesses can create value added (in the economic sense of the word). Government activities on the other hand, while being very valuable, do not create value added in a way that is observable in micro-data, and therefore we do not include turnover from outside the business economy. Other indicators such as the number of businesses and jobs are presented for the entire economy. Jobs are considered good for the economy regardless of whether they are in the business economy or not. Accordingly, we measure all businesses and jobs within the internet economy. An overview of which indicators are presented for the business economy and which are presented for the whole economy is shown below.

6.1.4 Overview of population per indicator

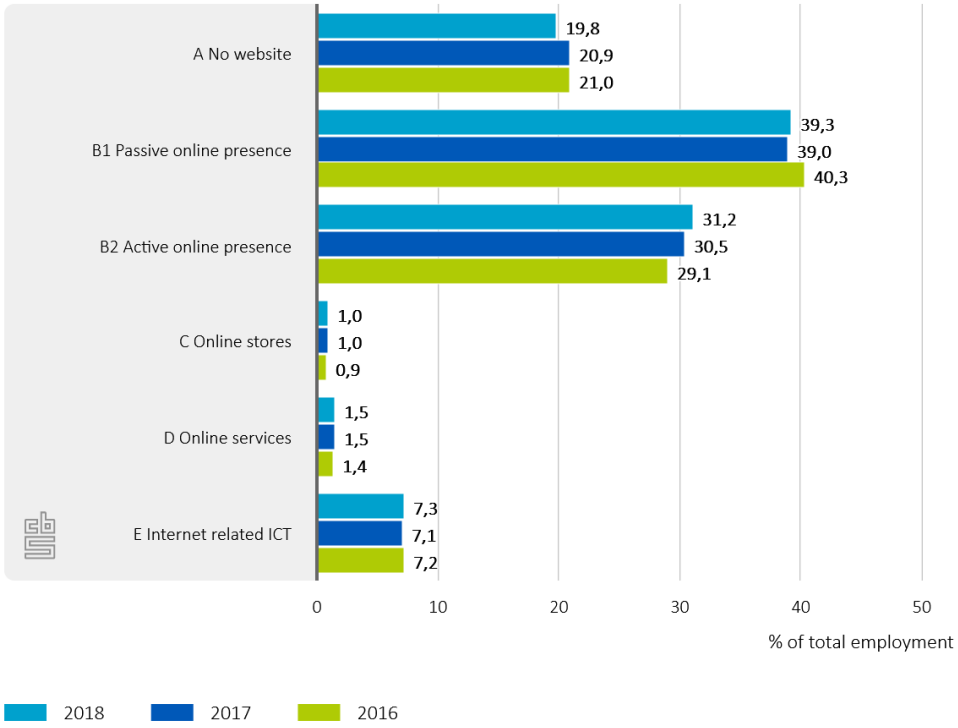
Indicator	Population
Number of businesses	Entire economy (NACE codes A-U)
Age	Entire economy (NACE codes A-U)
Size	Entire economy (NACE codes A-U)
Sector	Entire economy (NACE codes A-U)
Employees	Entire economy (NACE codes A-U)
Jobs of employees	Entire economy (NACE codes A-U)
Province and COROP	Entire economy (NACE codes A-U)
Turnover	Only business economy (NACE codes B-J, L-N)
Value added	Only business economy (NACE codes B-J, L-N)
Production value	Only business economy (NACE codes B-J, L-N)
International trade (goods)	Only business economy (NACE codes B-J, L-N)
Employed persons	Only business economy (NACE codes B-J, L-N)
Full-time equivalents	Only business economy (NACE codes B-J, L-N)

A second important economic variable is total employment expressed in FTE. The number of jobs as presented before referred only to the jobs of employees and not to self-employed persons. Figure 6.1.6 shows that the core of the internet economy contributes 9.7% to total employment of the business economy (9.5% in 2016). Employment associated with businesses represented online adds up to 80.2% of total employment of the business economy (79.0% in 2016). Compared to turnover, the shares of employment show a more clear picture, namely a slightly declining share of the businesses without a

[13] We present the indicators employed persons and full-time equivalents for the businesses economy only. These variables come from a data source which is only concerned with the business economy.

website, a persistent growth for the businesses with an active online presence and limited growth for the businesses belonging to the core of the internet economy. See table 7 in Appendix C for the absolute and more detailed data on employment.

6.1.6 Employment of businesses economy by internet category, 2016-2018

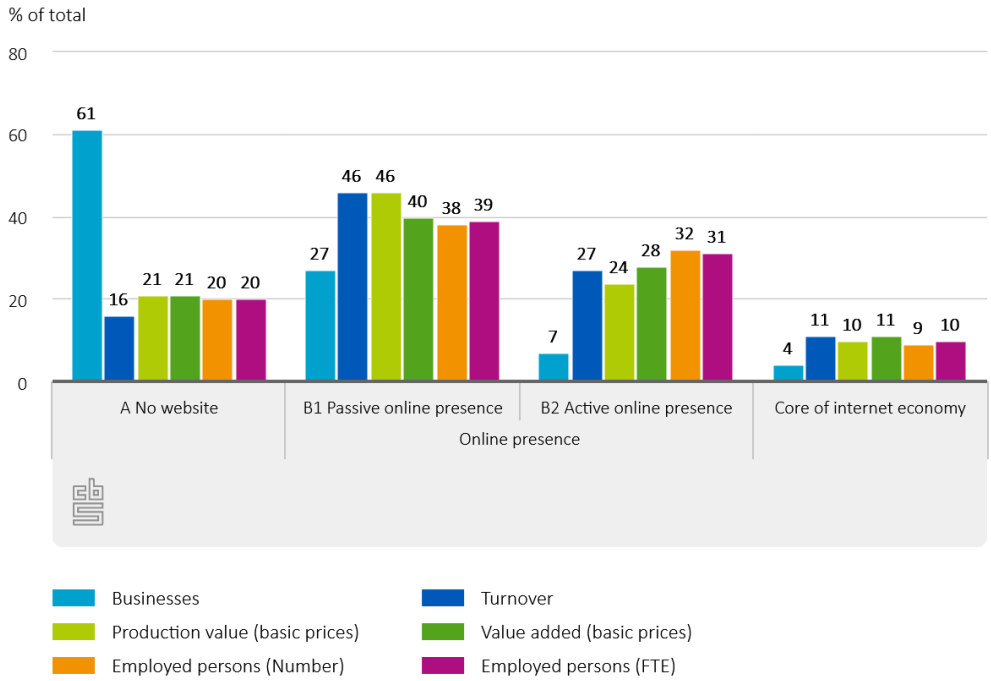


As shown in the box on variables and population, we have calculated six variables. We choose turnover and employment (FTE) to illustrate the development in the period 2016-2018. In Figure 6.1.7 we summarize five variables for 2018 to see how the economic importance of the different categories of the internet economy change depending on the view one takes.

Economic relevance core of the internet economy: 10% of business economy.

As illustrated before, the share in the number of businesses contrasts the most with the other – economic – variables. If we look at the core of the internet economy, its share in the number of businesses in the business economy is rather modest. Its share in the more economic variables as production value, value added and employment is larger and fluctuates round 10%. This suggests a number of larger businesses operating in the core of the internet economy. This will be discussed in more detail in section 6.4, as well as the differences between the categories within the core of the internet economy i.e. online stores, online services and internet-related ICT. Figure 6.1.7 shows (again) the importance of the businesses with a passive online presence. This category dominates the business economy with shares varying from 46% of turnover to 38% of the number of employed persons. See table 7 in Appendix C for an overview of these more macro-economic variables.

6.1.7 Economic indicators for the internet economy as part of the business economy 2018



In the following sections of this chapter, we discuss the most interesting results which give insights into the nature of the categories within the internet economy. Paragraph 6.2 focusses on businesses without a website and the differences between these businesses and the businesses which do have a website. The businesses with an internet presence (both passive and active) are discussed in more detail in section 6.3. Section 6.4 concerns the core of the internet economy according to this study: the online stores, the other online services and internet-related ICT. The complete set of tables with all the results which is the basis of this chapter, can be found in Appendix C.

6.2 Businesses without a website

Most businesses without a website are self-employed businesses

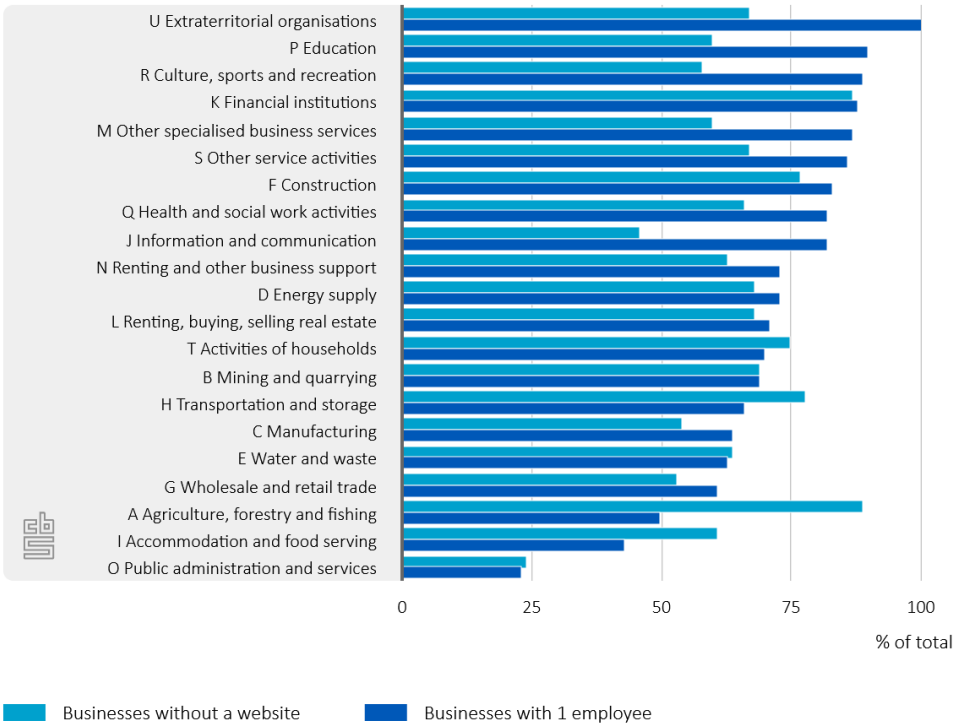
The results of this study show that in 2018 almost 64% (1,1 million businesses) do not have a website (category A of the internet economy). This was 66% in 2017 and 65% in 2017. Of all the categories, this category is by far the largest. However, 84% of the businesses without a website consist of only one employee. For example, self-employed contractors or financial advisors don't necessary need a website to do business, especially given the cost of setting up a website in comparison to using, for example, a Facebook page.[14] Also, there is an increasing number of (small) businesses that sell online via online

[14] In 2019 50% of the businesses with 2-10 employed persons use social media like Facebook; mostly for promoting their business and products (Source: CBS, ICT-survey on enterprises 2019).

intermediaries[15]. If we consider businesses with 2 or more employed workers we see in 2018 that almost 54% have a website.

Figure 6.2.1 shows the number of businesses without a website by NACE category and the share of businesses with only one employee in that same NACE category. This illustrates that there is perhaps a correlation between the size of the business and having a website or not. However, the size does not explain everything. The NACE category Information and communication for instance accommodates a large number of businesses with only one employee (82%) but nevertheless has the majority of the businesses in this branch a website (54%). So, there are other reasons for having a website than just size. Businesses in this branch themselves wish or need to have a website independent of size. Moreover, the barrier to create a website for these businesses seems small because it is part of their own business. On the other hand, Agriculture, forestry and fishing, is less dominated by small businesses (50%) but 89% of the businesses still do not have a website. In the Accommodation and food serving 43% of the businesses have only one employee, but 63% of the businesses in this branch do not have a website. This then is an example of a branch where businesses can sell via large intermediaries such as booking.com and thuisbezorgd.nl. Note that of all branches of industry the government institutes (Public administration) are most likely to have a website.

6.2.1 Businesses without a website broken down by NACE category, 2018



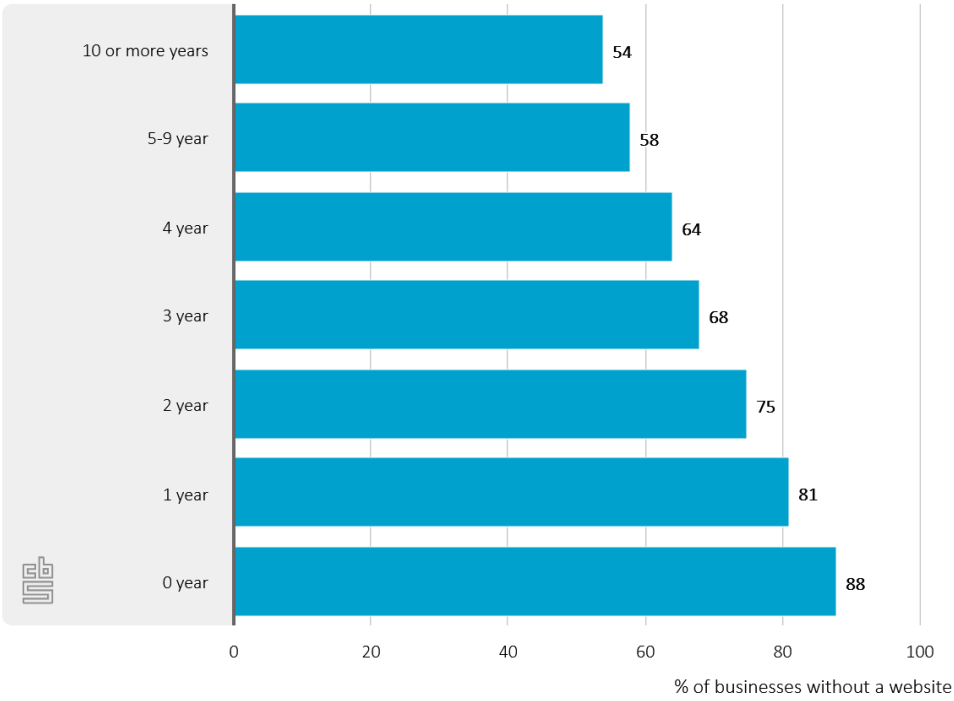
[15] In 2018 25% of the online sales of businesses with 2-10 employed persons was conducted via online intermediaries (Source: CBS, ICT-survey on enterprises 2019).

Older businesses most likely to have a website

There was a time when a website was associated with new and trendy businesses. Over the years this image has altered. And having a website is not 'the latest thing' anymore. Businesses that had existed for decades also adopted these new means of promoting their business, communicating with (potential) clients, selling online or just increasing their visibility on the internet.

Figure 6.2.2 shows that the oldest businesses in the entire population most frequently do have a website (46%) and of the youngest ones (88%) do not (yet).
The outcome of Figure 6.2.2 may be biased by the fact that older businesses are better recorded in the business register and can more often be linked to their website as provided by Dataprovider; but the differences are substantial. Besides this, larger business are more often also older and larger businesses have more often a website.

6.2.2 Businesses without a website by age, 2018



Large number of businesses but relative lower economic importance

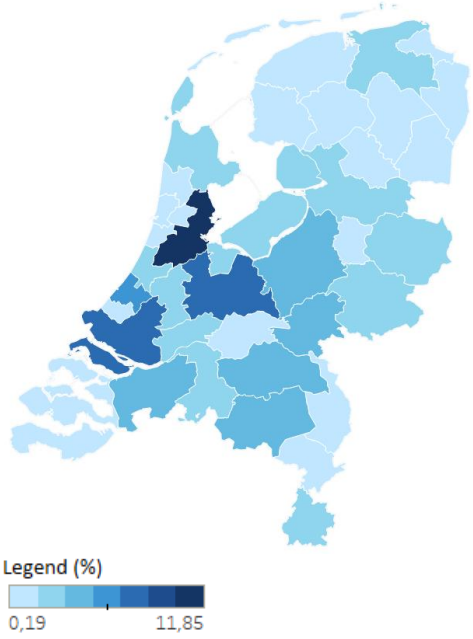
As shown before the majority (64%) of businesses do not have a website although this percentage decreased slightly in the period 2016-2018. The contribution of this large number of businesses to the macro-economic variables of the business economy like production, value added and employment fluctuates around 20% in the period 2016-2018 (see table 7 in Appendix C). The contribution to total turnover of the business economy increased from 15% in 2016 to 16% in 2018 (see table 4 in Appendix C).

Regional distribution of businesses without a website

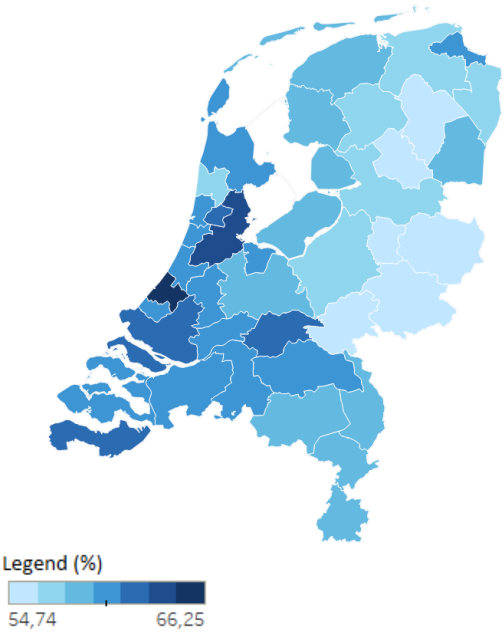
Where do we find the businesses without a website? On the left-hand map (6.2.3) the percentage of the total number of businesses without a website is plotted by region. This gives an impression of where these businesses are concentrated. We see the largest share of these businesses in and around the large cities Amsterdam, Utrecht, Rotterdam and The Hague. This more or less concurs with the regional distribution of all Dutch businesses. On the right-hand map the share of businesses without a website in the total number of businesses in that particular region is plotted. This provides an idea of the degree of specialisation in each region. In this case, Amsterdam and Utrecht are less important. Although these regions accommodate a large number of businesses without a website, their relative importance is more modest, because these regions accommodate large numbers of other businesses. Businesses without a website are also relatively important for The Hague, as well as more rural areas such as Zeeland. These are not the same kind of businesses without a website as observed in the urban regions, but mostly agricultural businesses.

6.2.3 Regional distribution of businesses without a website, 2018

% of total number of business without a website



% of total number of businesses



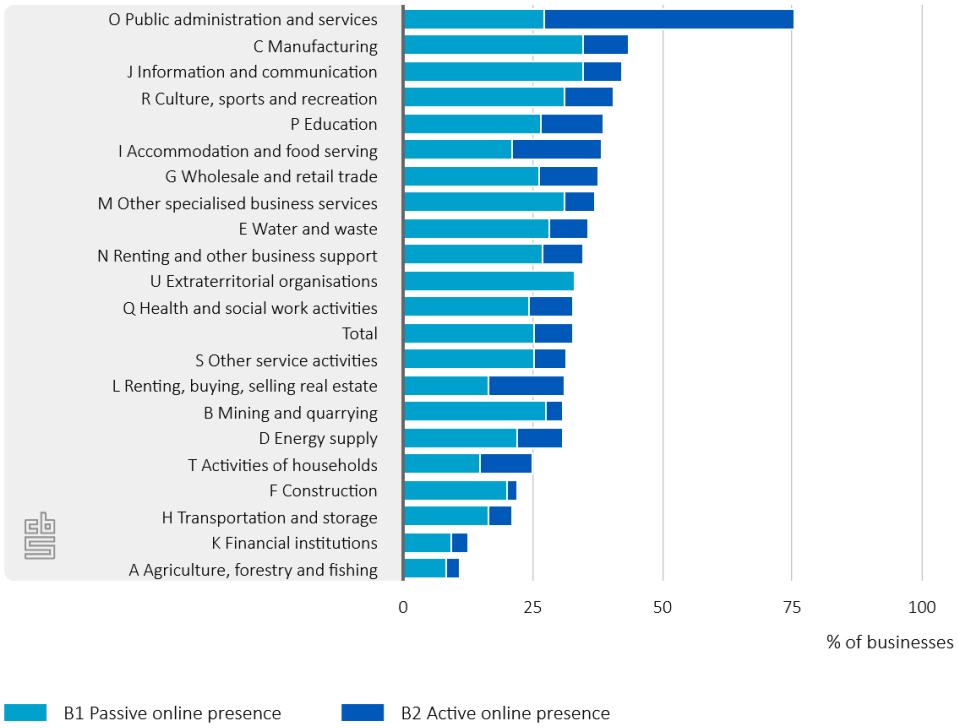
6.3 Businesses with internet presence

Businesses with an internet presence (category B1 and B2) are not allocated to the core of the internet economy (categories C, D and E). Business in category B1 and B2 do not generate income directly through or with the internet, but their website may indirectly help in their business activities. For example, websites serve a marketing function as well as providing information which encourages consumption of their goods or services.

Furthermore, a large number of businesses do not sell directly to consumers but to – a limited number of – other enterprises. So, they do not necessarily need a state-of-the-art-website to address their market. Internet presence businesses are divided into two categories: category B1 termed passive online presence and category B2 termed active online presence. Passive online presence businesses provide only information on, or marketing for, their business activities (one-way communication). Active online presence businesses provide some services on their website to support their core businesses activities e.g. making a reservation or fill in forms (two-way communication).

Figure 6.3.1 shows the share of businesses with an online presence in a particular NACE category. This shows that 76% of the institutes belonging to Public administration fall into this category. Another difference compared to other NACE categories is that the majority of the government institutes with online presence belong to the category Active online presence (B2). This is because most government institutes provide online services; citizens and businesses can make appointments, order official documents or fill in tax forms. Other NACE categories with relatively large shares of Active online presence are Accommodation and food serving, Education and Renting and other business support. Manufacturing and Construction on the other hand, are dominated by businesses with only a passive online presence.

6.3.1 Businesses with online presence by NACE category, 2018



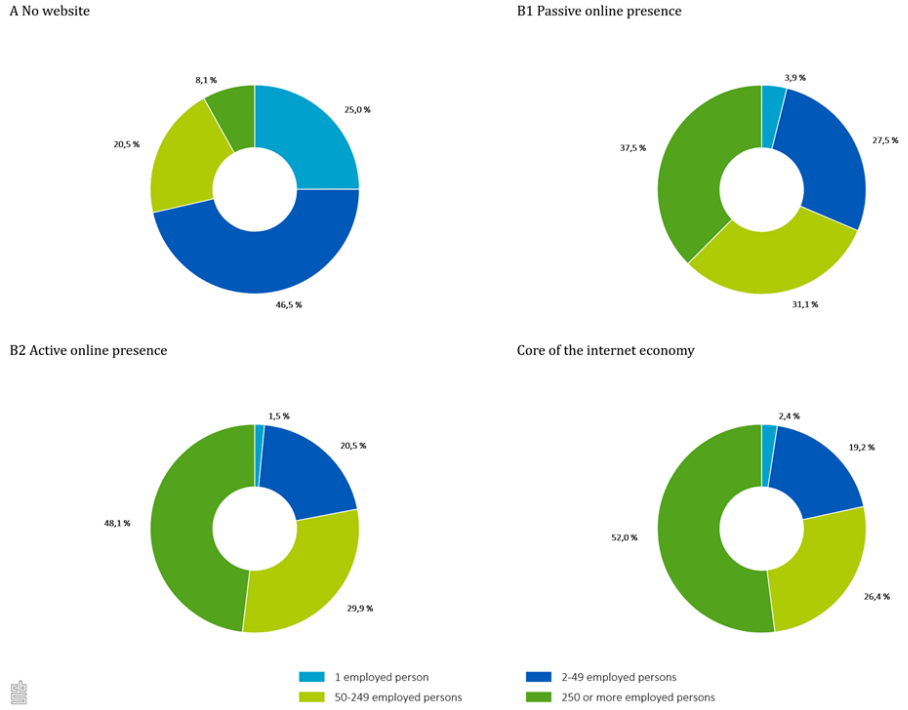
Category passive online presence economically most important

As already shown, the number of businesses in these categories of the internet economy increased in the period 2016-2018. The share of the number of businesses with a passive online presence in the total economy increased from 23.8% in 2016 to 25.4% in 2018; for the businesses with an active online presence these percentages are 6.7% and 7.4% respectively. The (macro-)economic contribution of the businesses with only a passive

online presence to the business economy in 2018 varies from 46% of total turnover to 38% of the number of employed persons. According to this metric, passive online presence is the largest of the internet economy. The contribution of the businesses with an active online presence in 2018 varies from 32% of the number of employed persons tot 24% of the production value.

Figure 6.3.2. shows how much turnover is produced by different size-classes per category. These are not market shares, especially not for the categories A, B1 and B2, because the similarity between the businesses in a category is not that they operate on the same market, but that they belong to the same internet category. Category A is dominated by the small and medium sized businesses. Also the turnover of the businesses with only a passive online presence is fairly evenly distributed over the three largest size-classes, with the smallest size class being unimportant in this case. In the other categories around 50% of total turnover is realized by only a limited number of large businesses.

6.3.2 Turnover internet categories by size-class, 2018

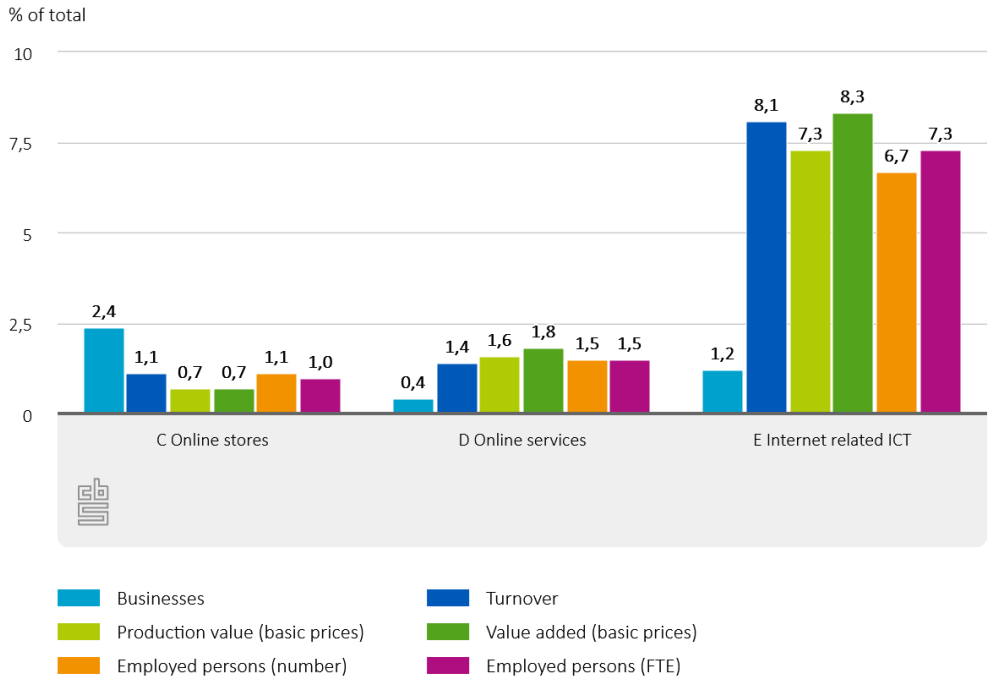


6.4 The core of the internet economy

The online stores, the other online services and the internet-related ICT (categories C, D and E) constitute the core of the Dutch internet economy. Their income is directly generated through the internet (online stores and other online services) or with the internet (internet-related ICT). In total, the core of the internet economy consists of more than 50,000 businesses representing 3.0% of all Dutch businesses. Together they account for a turnover of €146 billion euros, which is 10.6% of the turnover of the Dutch business economy. There are 550,000 jobs in the core of the internet economy. In absolute number of businesses the core of the internet economy grew in the period 2016-2018, but its

overall economic importance for the business economy fluctuates around 10% dependent on the indicator under consideration.

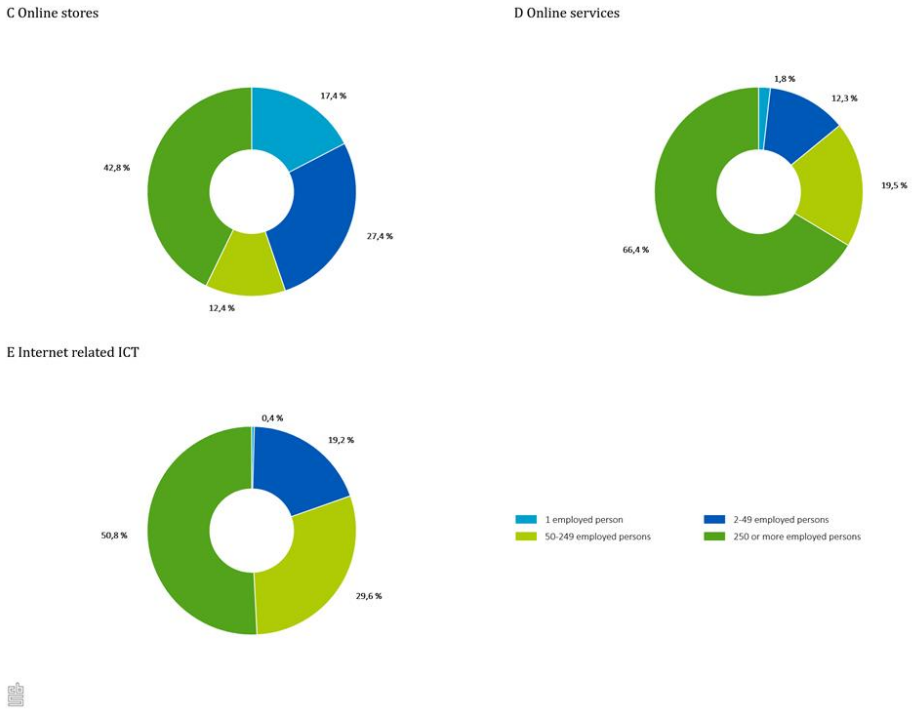
6.4.1 Economic indicators for the core of the internet economy as part of the business economy, 2018



Internet-related ICT dominant category within the core of the internet economy

Within the core of the internet economy the category internet-related ICT is the most dominant one. Its contribution to the economic variables of the business economy varies from 8.3% of the value added to 6.7% of the number of employed persons. The number of businesses in this category however, is relatively small. On the other hand we have online stores with 2.4% of all business of the business economy but a less impressive contribution to the economic variables. The contribution of these businesses to the business economy varies from 1.1% of turnover to 0.7% of the production value.
This picture is amplified by Figure 6.3.2 where one can see that the turnover of online stores is fairly distributed over the different size-classes. There is a small number of large online stores, but there is also a large number of small and medium sized online stores which serve e.g. a niche market and realise a certain turnover by doing this. The turnover of the other two categories is dominated by a limited number of large businesses. These businesses are not always necessarily competitors; running an online platform for booking hotels is something completely different than running a dating site. And being an internet provider is not the same as being a web-designer.

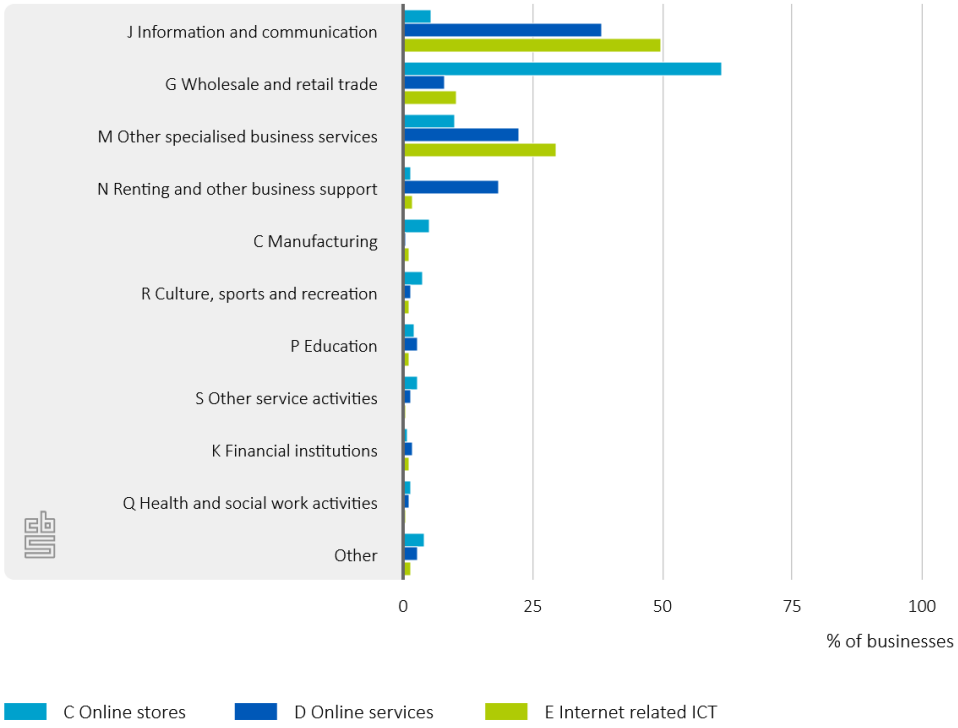
6.4.2 Turnover of the core of the internet economy, by category and size-class, 2018



Core of the internet businesses concentrated in a limited number of industries

In which NACE categories do we find the core of the internet businesses? Half of the businesses belonging to category E internet-related ICT fall in the NACE category Information and communication. This makes sense because this is the branch of industry which accommodates web hosting, data centres, cybersecurity, cloud services, software development, internet providers etc. One third of category E businesses are part of the NACE category Other specialised business services which includes the ICT-consultants, web-designers and other online advertisers.
Over 60% of all online stores are part of the NACE category concerning Wholesale and retail trade. The largest share of the businesses of category D Online services are part of NACE category Information and communication. The substantial share of category D businesses in the NACE category Renting and other business support services can be explained by the large number of businesses from whom one can rent and lease products via websites. In the NACE, classification renting and leasing is not seen as a “store” where one can buy things, but as a service which apparently can be conducted online. The extent to which businesses belonging to the core of the internet economy are found outside the anticipated four NACE categories mentioned above, is slightly less than expected. Online stores are so far the most widespread with 21% of the businesses falling outside the four NACE categories discussed. In 2018, for online services this percentage is 13% and for internet-related ICT 8%.

6.4.3 Distribution of businesses in the core of the internet categories by NACE category, 2018

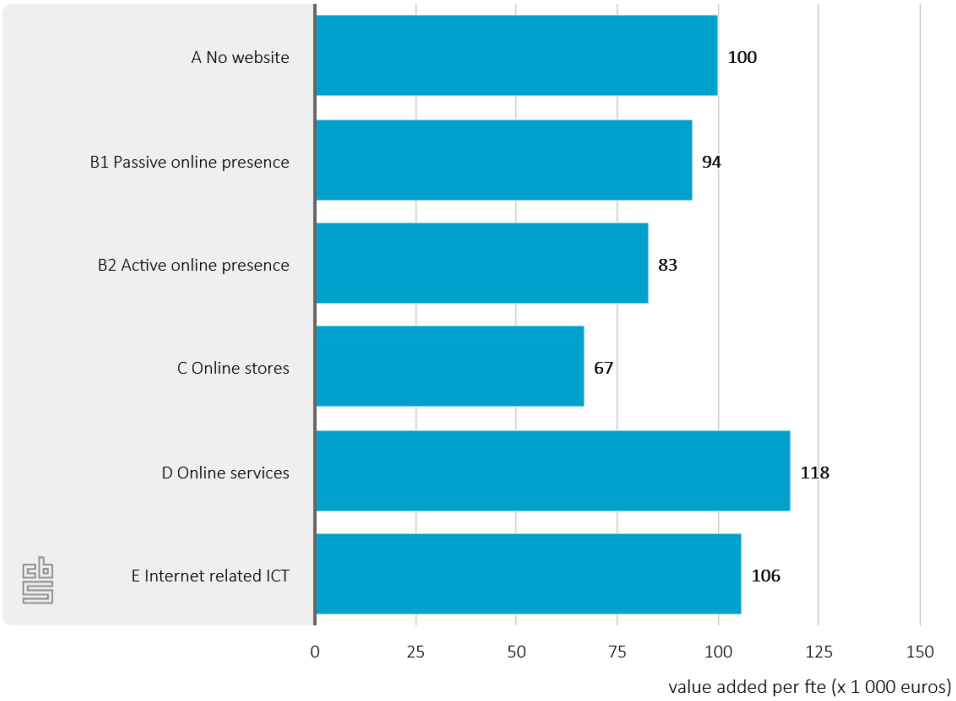


Online services highest labour productivity

Labour productivity[16] for businesses in the category Online services is the highest of the categories of the internet economy. This is due to online booking websites which have a high value added and relatively few employees. Furthermore, the labour productivity of the businesses in the category internet-related ICT is above average. This can be explained by a number of large businesses in this category who make an above average use of capital (as opposed to labour) in their production process. The productivity of the online stores is below average although a large part of their business processes is conducted online. One of the reasons may be the relatively low wages paid in this sector. The productivity of the other categories is hard to interpret because of the many different types of businesses which these categories contain. To provide some guidance, the productivity of online stores is comparable with the NACE category Construction (73,000 euros per FTE), the productivity of Manufacturing is 119,000 euros per FTE. Branches of industry with high productivity (upwards of 200 thousand euros per FTE) are among others the financial sector and the energy sector. The overall labour productivity of the business economy in 2018 was 97,000 euros per FTE.

[16] Labour productivity is calculated as the quotient of value added and employment (FTE).

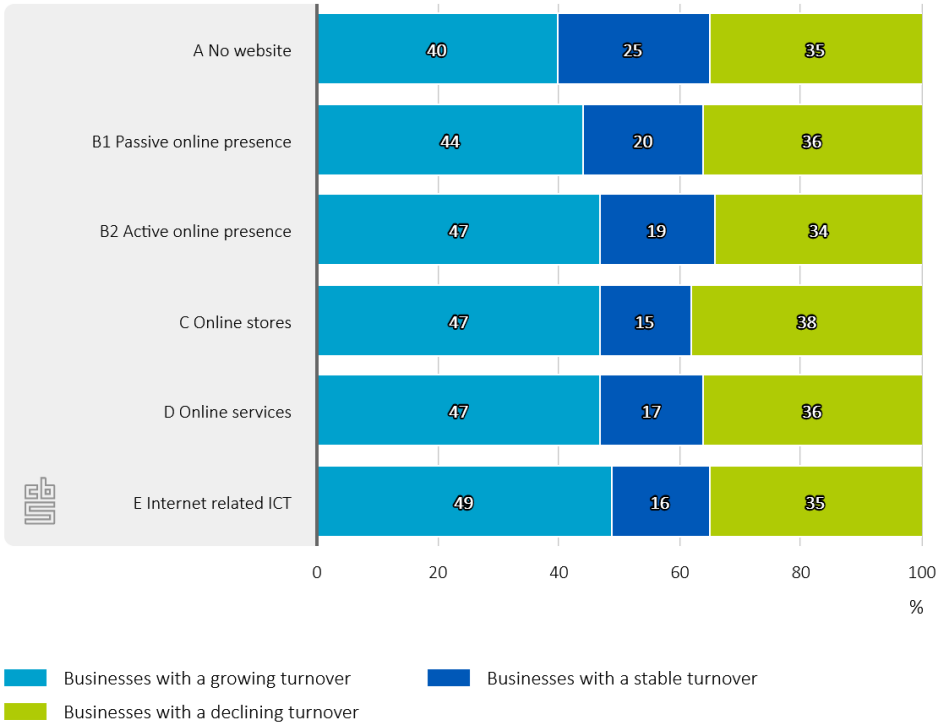
6.4.4 Labour productivity by internet category, 2018



Turnover instability and online stores

Do the businesses in the core of the internet economy grow faster compared to business in other categories? In Figure 6.4.5 the turnover of the 4th quarter of 2018 is compared with the turnover of the 4th quarter of 2017 for those businesses who existed in both periods. Of the online stores, only 15% of all businesses had a turnover which was more or less stable; the lowest percentage of all categories. A share of 47% of online stores saw turnover grow. However, 38% of the businesses in this very same category saw its turnover decline; the highest percentage of all categories. This variation in turnover is greatest for online stores. For the other categories of the core of the internet economy this picture may also be recognised, although somewhat less pronounced. Within the other categories of the internet economy, there are more businesses with a stable development of their turnover.

6.4.5 Turnover developments (growing and declining)[17] by internet category, 4th quarter 2018



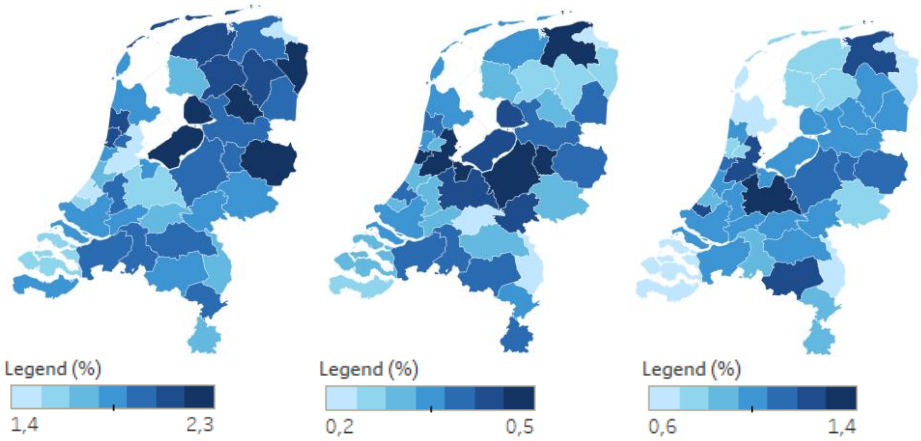
Regional distribution of the businesses of the core of the internet category

In general in the Netherlands, economic activity is centred around the largest cities of Amsterdam, Utrecht, Rotterdam and The Hague. The internet economy is also focused around these cities. However, there are many interesting exceptions to this rule, as well as variation between the categories of the internet economy. What is shown on the three maps below is the relative importance of a certain category of businesses for the particular region. For example, the online stores are often located in the northern half of the Netherlands, roughly above the latitude of Amsterdam. Even online stores need distribution centres and the choice of the location will be influenced by the price of land and logistical concerns. Interestingly, there is a relatively large concentration of online stores in border regions like Oost-Groningen and Twente.

[17] Growing turnover: an increase of 5% or more. Declining turnover: a decrease of 5% or more. Stable turnover: a decrease or increase of less than 5%.

6.4.6 Relative regional distribution of branches of the core of the internet economy (% of total number of businesses), 2018

Online stores *Online services* *Internet-related ICT*



Online services only exist because the internet offers them a platform to provide services. Among the online services for example, are price comparison sites, dating sites, online games and auction sites. Although the number of businesses in this category is still relatively low, they are more likely than average based in the regions in the centre of the Netherlands and Groningen. Groningen has a reputation for innovative economic activity, particularly related to digital and internet based industries. The local business units of the almost 15 thousands businesses in category E are relatively well represented in and around the cities Utrecht, Groningen, Eindhoven and Delft. The latter two being examples of cities with a University of Technology where a number of start-ups may have assembled.

7. Discussion and recommendations

7.1 General conclusion

The first important aim of this study was to investigate the robustness of the internet data and the adopted research method for analysing trends. The methodological weaknesses are compounded when considering trends over time. While we believe that the trends at a macro-level are valid indicators of the trends in the internet economy, the results at a micro-level suffer from the above methodological limitations to a greater extent. Analyses of our results for consistency at the micro data level show that there is substantial instability in how businesses are categorised over time. It is possible that a business is categorised in a different category in 2016, 2017 and 2018, or is not categorised at all in one year, for example because the website is not found by Dataprovider in that year or could not be matched to the GBR. It is therefore not recommended to follow specific business through time.

This means that we have to be careful when interpreting results about the trends in the internet economy. The group of companies that belong to a certain category one year can be significantly different from the group of companies that belong to that category the other year(s). Especially with outcome variables on volume such as turnover, employment, value added or trade. Because a few (large) companies can have a significant impact on the outcomes. Variables on more demographic subjects such as sector, size, age, regional distribution, and turnover development are less sensitive to the influence of a few companies and provide more reliable longitudinal results. Nevertheless we have focused on those developments which seem plausible and anticipated and tried to describe them as accurately as possible to provide some insights in the trends in the internet economy in the Netherlands (the second important aim of this study).

7.2 Strengths and limitations

The key strength in this study is the combination of big data from the internet with Statistics Netherlands' micro-data. The big data from the internet provided by Dataprovider is very up to date due to monthly updates and provides a wealth of information on the businesses behind the website. CBS micro-data provides reliable and detailed statistics on businesses. One particularly strong aspect of our methodology is the level of detail and flexibility in the categorisation of websites. One of the recommendations from the previous study was to revise the definition of the internet economy if needed due to new developments in the internet economy and / or the availability of new data sources. There were no new data sources available but we expanded the definition of online services to include crowdfunding websites and platforms. And the keywords which were used in the definition of internet-related ICT were also updated. Changes to the nature of the internet economy can therefore be incorporated in the methodology with relative ease. In the following sections we discuss

the main limitations of the study in more detail. These limitations provide insights in how to improve the method in future research.

7.2.1 Data

Regarding the data, the principal limitation is that the Dataprovider data does not include all Dutch websites. While the coverage is estimated to be high at 95%, the consequence is that not all businesses can be linked to a website. For example, some websites from foreign business that have a branch in the Netherlands are more likely to be missed. This means that it is likely that the size of the internet economy is somewhat underestimated in this study. Another weakness in the data also leads us to suspect that the size of the internet economy is somewhat underestimated. This is because many business do not have an "individual" website, but make use of Facebook and other social media outlets to facilitate an online presence. These companies cannot be identified as part of the internet economy within this study. It is likely to be the smaller businesses that make use of social media for an online presence because this is cheaper than constructing and maintaining one's own website. This suggests a degree of bias in our results towards larger businesses.

7.2.2 Definition

Regarding the definition of the internet economy, it is important to note that the definition is, to a certain extent, subjective. A good example of this subjectivity are banks. Banks provide many online services which have significant economic value (online banking). In this regard, banks could be categorised as an online service (category D). However, our definition categorises them as having an active online presence (category B2). In practice, banks belong to both category B1 (consider online information about mortgages), category B2 (active online presence, one can apply for a credit card online) and also category D (online banking). A pragmatic choice was made to allocate banks to category B2. Similar issues exist for many kinds of businesses. In all cases, decisions were made which as far as possible reconcile pragmatic considerations with the beliefs about what the internet economy is. It is important to regularly evaluate these choices. For example, both the banking industry is changing, as is the way that banks provide services via the internet. It may be desirable to add certain banks into category D if the banks do not have high street branches.

The subjective nature of the definition can also be seen in some of the assumptions made as part of the definition. It is important to make these assumptions, and their likely effects on the results, explicit. An important assumption is that the definition is conservative. Businesses are only included in the core categories if it is reasonably certain that the majority of their business activities fall into one of the core categories. In this way, businesses which are defined as having internet presence may also perform some business activities associated with the core of the internet economy.

7.2.3 Method

Regarding the method, there are three key limitations. The first concerns the link to the GBR. Although we made some improvements in the linking method, it has not been possible to link every website to the correct business. There is nothing to stop a website from putting the CoC number and contact information (telephone number, postcode or email address) of another business, for whatever reason, on their own website. It is also not possible to ascertain with certainty the extent to which incorrect links have been

made. There is also no guarantee that the contact details or CoC numbers on the website have been correctly (without typos) placed on the website. Further, large, complex business with multiple business units per enterprise group are difficult to link with the GBR because an enterprise group often only has one website for all of its business units. And finally, the linking variables from the GBR may also be obsolete.

The second limitation regarding the method concerns the issues of multiple websites per CoC, multiple CoCs per enterprise group and business unit belonging to multiple categories (overlap). We constructed decision rules to deal with these problems and we are confident that the use of decision rules is the best approach. As such, the categories are accurate in broad brush terms, but we cannot guarantee that all businesses fall into the correct category.

The third limitation concerns the use of keywords (the most important words used on the website) to classify companies to category D (online services) and E (internet-related ICT). Although we updated the lists of keywords, we used the same set of keywords for all three years. Because of the way that Dataprovider collects data from websites and speed and extent which the internet changes, there can be variation from month to month in the keywords which are recorded about a websites. Analyses of the consistency of the main Dataprovider variables showed the changeability of the keywords variable which can explain why so many businesses in these two categories change category from one year to another. For example a website may be classified in D in 2016, then not classified in D in 2017, and then classified in D again in 2018 even though little has changed regarding the fundamentals of the business or the website.

7.3 Recommendations

One of the goals of new innovative research is that it provides insights for improvements of the method in the future. In this section a few recommendations are given to deal with the most important limitations of this study.

A possible way to address the weaknesses that the Dataprovider data does not include all Dutch websites and does not include online presence through platforms, is to acquire data from Facebook or other platforms on the number of pages which classify themselves as businesses. While it may not be possible to establish any links to the GBR on the basis of these data, there may be some opportunities to better understand to what extent our results are biased.

A way to deal with the problem of the changeability of keywords variable is to reset the list of keywords every year. The reason why we did not do so for this study is that this is a time consuming process (a lot of manual work) and that we wanted to test the consistency of our method. Another solution can be the use of machine learning techniques to classify the businesses into a category of the internet economy. Last time we did a number of tests with machine learning algorithms without too many encouraging results. Probably more data preparation is needed to get better results.

Methodological improvements to alleviate the problem of the robustness of the method for analysing trends are desirable. This requires further research, but one possible way to

deal with this is to adopt a “base-year with revision” approach. This involves first establishing a base year, for example 2019. Dataprovider data for every month in the base year can be combined analyzed and merged with the GBR to produce the most comprehensive possible delineation of the internet economy. The next year in the series (2020) would then build directly on the base year. If a company is classified in category D in the base year, it would therefore per se be classified in category D in 2020, if the business and the website still exist in 2020. Any new websites in 2020 which are not in the base-year would also be added. The base-year approach improves the consistency of the time series. However, the accuracy resulting statistics declines over time if the base-year is not updated. Therefore, if for example, a fourth year in the series is requested (2022), it may no longer be appropriate to retain the same base year. It is therefore necessary to carry out a revision in which the base-year is recalculated in a more recent year. Every time the base year is revised there is a break in the series. A base-year with revision approach is the preferred approach if this study is repeated in the future and the objective is to analyse trends.

Appendix A - Dataprovider dataset

Group	Field	Example data	Description
Geolocation	Country	*The Netherlands*	Where the business is actually operation
	Region	*Noord-Holland*	The state or province where the business operates
	Zip Code	*9723 HS*	The zip code of the business
	Zip Code Quality	*75%, (from 0% - 100%)*	How certain we are that this is the right zip code
	City	*Amsterdam*	The city where the business operates
	Address	*Grote Markt 23*	The address of the business
	lat / long	*32.70179, -97.62637*	Geographic coordinates of the business
Business	Company Name	*R. Goulooze Holding B.V.*	The name of the company
	Legal Entity	*B.V.*	The legal entity of this business
	Chamber of Commerce	*30125826*	The chamber of commerce number of the business
	Bank Account Number	*325324603*	The bank account number of the business
	IBAN Number	*NL36RABO0325324603*	The IBAN bank account number of the business
	BIC Number	*RABONL2U*	The BIC number of the bank that the company uses
	Bank	*Rabobank*	What bank the company uses
	Tax Number	*NL115441359B01*	The tax number of the company
	Phone Number	*31(0)306579252*	The most important phone number of the company
	Phone Number Quality	*83%, (scale from 0% - 100%)*	How certain we are if this is the right phone number
	Phone Numbers	*31(0)325324603*	The phone numbers we have found
Content	Hostname	*www.shirtplanet.nl*	The URL of the company
	Title	*ShirtPlanet.nl De Leukste T-shirts Online*	The title that is used on the website
	Description	*ShirtPlanet leuke grappige t-shirts kopen van baby rompertje tot extra grote maten.*	The meta description that is used on the website
	Keywords	*shirtplanet, shirt, grappige, rompertje, verjaardag, kado, leuke, humor, funny*	The most important words that are used on the website
	Category	*Fashion*	Dataprovider has xx default categories and based on the content of the websites assigns it accordingly
	Authors	*Frans de la Haije*	The author that is found in the html code, often the designer of the website
	Copyright	*MediaCT*	The copyright notice that is found in the html code, often the designer of the website
	Multi Language	*No*	The website contains copy in one or more languages
	Language	*NL, EN*	What languages are used on the website
eCommerce	Online Store	*Yes*	The website contains an online store
	eCommerce Probability	*37%, (scale from 0% - 100%)*	How certain we are that the website contains an online store
	Shopping Cart Software	*Magento*	What kind of shopping cart software does the website use
	Trustmarks	*Thuiswinkel waarborg,*	The trustmarks the website uses
	Delivery Services	*TNT, DPD*	What delivery services are offered
	Payment Methods	*Mastcard, VISA, PayPal*	What payment methods are offered
	Payment Services Provider	*PayPal, Docdata*	What online payment services are used
	Currency	*EUR*	In what currency the priced are advertised
	Average Price	*29*	The average price of the products offered
	Products	*3,924*	Estimation of how many products are offered on the website

Group	Field	Example data	Description
Marketing	Alexa Rank	*17,601,874*	How does the website rank on visitors (lower score is more visitors) according to Alexa
	Tweets	*15*	How many tweets are sent by the Twitter account that is published on the website
	Facebook Shares	*23*	How many posts are shared on the Facebook account that is published on the website
	Incoming Links	*115*	How many incoming links the website has
	Refering websites	*12*	How many other websites are linking to the website
	Anchor texts	*www.shirtplanet.nl,* t-shirts kopen, *tshirts, online t-shirtsbestellen*	What Anchor text is used in the incoming links
	Site traffic	*4,078*	How many unique visitors the website has, this is an estimate by Dataprovider
	Analytics Id	*UA-9625634-15*	The Google Analytics ID that is found on the website
	Adsense Id	*PUB-5709822177168445*	The Adsense ID that is found on the website
	Analytics Id	*Google Analytics*	What kind of statistics software the website uses
	Ad Network	*Google Adsense*	Which Adnetworks the website works with
	Affiliates	*Daisycon*	Which affiliate networks the website works with
	Social	*Twitter, LinkedIn*	Does the website use Facebook, Twitter, LinkedIN, Pintrest, Google +
	Social Profiles	*www.linkedin.com/company/feederlines, www.twitter.com/feederlines*	The social profiles the website publishes
	Live Chat Software	*CoBrowser*	Visitors of the website can chat with the people behind the website
	Jobs	*Yes*	The website has jobs advertised
Technical	CMS	*Plone*	What Content Management System is used on the website
	Scripting Language	*PHP*	What scripting language is used to build the website
	Technical evaluation	*6,7 (scale from 1 - 10)*	How is the website coded, the W3C is the benchmark
	SEO Score	*84% (scale from 0% - 100%)*	How well does the website use all the HTML elements to tell search engines what the website is about
	Flash	*Yes*	The website uses Flash software
	RSS	*Yes*	The website offers a RSS feed
	Login	*Yes*	The website has content that's protected by a login and password
	HTML version	*XHTML 1.0 Strict*	What version of HTML is used to code the website
	Generator	*Plone - http://plone.org*	What HTML generator is used to code the website
	Mobile version	*No*	Is there a mobile version of the website or is the website responsive
	Mobile App	*App Store*	Does the website refer to apps in the Apple App store
	Maps	*Bing Maps*	The website makes use of maps
	Libraries	*MooTools, Slimbox, SWFObject*	What scripting libraries are used
Hosting	Top Level Domain	*Nl*	The top level domain of the website
	Subdomain	*www*	Subdomain of the website
	Domain	shirtplanet.nl	The domain of the website
	Hosting Country	*NL*	In what country the website is hosted
	Domain age	*120*	The number of months ago the domain was registered
	IP Address	*141.255.181.112*	The IP address of the website
	AS number	*51,686*	ID of the owner of the IP block
	AS Company	*Antagonist B.V.*	Name of the company that belongs to the AS number
	Reverse DNS lookup	www.antagonist.nl	Domain associated with IP address
	Operating System	*Ubuntu*	The operating system of the server where the website is hosted
	Webserver	*Apache/2.2*	The software used to deliver the website
	Server Signature	*Apache/2.2.14 (Ubuntu)*	Information about the hosting server that is included in the header response
	SSL Certificate	*No*	The website uses Secure Sockets Layer for secure connections
	Status codes	*200,404*	The status codes of the pages found during indexation
	Average load time	*109 Kb/s*	The average loadtime of the website
	CDN	*Akamai, Google API*	Does the website use a content delivery network
	Video	*Vimeo, Youtube*	What third party streaming video suppliers the website uses
	Parking	*GoDaddy*	There's no website, it is parked with a default page from the webhoster
	Email provider	*Gmail*	Hosting provider of the email mentioned in the MX records of the DNS

Appendix B - Keywords

Internet category	Subcategories	Topic	Keywords: base	Keywords: combination
Category D: Online services	Leisure	Hotels	hotel, hotels, resort, resorts, hostel, bedenbreakfast, hostels, booking	hotels
		Flights	vliegtickets, vliegticket, vliegwinkel, tickets, ticket	vluchten, luchtvaart, airlines, luchtvaartmaatschappijen
		Holidays	reizen, rondreizen, reisspecialist, reis, vakantie, vakanties, travel, travels	online, website, site, vergelijk, vergelijken
		Food	thuisbezorgd, iens, eten, recepten, restaurant, restaurants, koken	eten, online, vergelijk, vergelijken
	News and entertainment	News	nieuws, weer, news, media, magazine	online
		Blogs	blog, blogger	online
		Vlogs	vlog, vlogger	online
		Games	spelletjes, spelletje, game, games, spellen, spelen, speel, spel	online, website, site, gratis
		Videos, music	tv, film, films, filmpjes, filmpje, movie, movies, serie, series,	online, streaming, stream, kijk, bekijk
		Books	boeken, books, boeken	online
		e-learning	learning, learn, cursus, cursussen, leer, leren	online
		Gambling	gokken, casino, casinos, gok, gokkasten	online
		Adult	escort, porno, porn, seks, seks, naakt, kinky, gay, erotic	
	Business	Advertising	adverteren, adverteer, reclame, reclamebureau, reclamebureaus	online, optimalisatie, adwords
		Finance	bankieren, bank, beleggen, onlinebeleggen, beleggers, belegger, beleggingen	online, digitaal, bitcoin, digitale
		Consultancy	advies, consultancy	online
		Jobs	vacaturesite, vacaturesites, vacatures, vacature, uitzendbureau, uitzendbureaus, uitzendwerk, uitzendkrachten, vakantiewerk, vakantiebaan, vakantiebanen, bijbaan, bijbanen, baan, banen, recruitment, recruiter, job, jobs, carriere, career	online, zoek, zoeken, vind, vinden
	Retail	Housing	woning, woningaanbod, huis, huizen, koopwoningen, makelaars, makelaar, huurwoningen, hypotheken, hypotheek, funda	verkopen, aanbod, koop, huren, huur, verkoop, kopen
		Price comparison	energie, verzekering, abonnementen, prijzen, aanbieders, goedkoopste, bespaar, autoverzekering, prijsvergelijk, prijsvergelijking, prijsvergelijker, abonnement, mobiel	vergelijker, vergelijker, vergelijk, vergelijken
		Tickets	tickets, ticketing, ticket	online
		Auctions	veiling, onlineveiling, veilingen, onlineveilingen, auction, auctions, veilingmeester, internetveiling, internetveilingen	online
		Online trade	marktplaats, tweedehands, speurders	online
	General services	Dating	datingsite, dating, date, sexdate, sexdating, daten	online
		Visualisations	visualisatie, visualisaties, visualiseer, animatie, animaties, ontwerp, ontwerpen	online, 3d
		Transport	car, meeliften, lift, liften, meerijden, 9292, reisplanner, routeplanner, parkeren, auto	online, app
		Online payment	diensten, payment, betalen, betaaloplossingen	online
		Crowdfunding	crowdfunding	
Category E: Internet related ICT	Hosting and cloud	Webhosting	hosting, domeinnaan, webhosting, server, service, cloud, development, ontwikkeling, cloudaanbieder,websitedevelopment,	websites, opslag, cloud, online, clouddiensten
		Cloud services, Datacentres	clouddiensten, datacenter, datacentrum, cloudcomputing, massaopslag, dataopslag, databanken, datawarehouse, gegevensverwerking, ontdek, levert, gebruik, omzet, zakelijk, leverancier, aanbieders, handelaar, database, cloudwebsites, cloudgebruik, cloudsourcing	websites, opslag, cloud, online, clouddiensten
	Websites and apps	Website design, developing	webdesign, internetbureau, websitebouw, webportal, webportals, developer, websites, design, ontwerp, software, develop, mobiele, ontwikkelaar, developer, ontwikkelen, developer, ontwikkeling, development, grafisch, origineel, bekeken, bediening, ontwikkeling, beeld, beelden, geluid, marketing, interactief, privacy, interactieve, interactie, bewegende, verhaal, animaties, multimedia, bouwers, bouw, bouwen, gebouwd, producent, handelaar, ontwerp, websolutions	software, pakketten, virtual, online, websites, web
		App design, developing	mobiele, iphone, android, tablet, mobile, opdracht, uitgeverij	app, apps, digital, digitale, applicaties
	Software	Software products and services	software, develop, ontwikkelen, produceren, uitgeven, adviseren, databanken, webdesignbureau	
	Marketing and consultancy	Internet marketing	ecommerce, emarketing, adwords, internetmarketing, zoekmachine, zoekmachines, b2b, b2c, verkoop, inkoop, online, marketing, seo, optimalisatie, socialemedia, sociale, analytics, adverteerders, gedreven, intelligence, privacy	online, websites, elektronisch, internet, netwerk, gegevens, search, business, digital
		Internet consultancy	internet, online, ontwikkeling, consultancy, research, onderzoek, advies, analyse, analyseren, techbedrijf, technologie, boekhoudleverancier, communicatie, wiki, community, hangout, skype, berichtendienst, communicatie, youtube, tweets, twitter, snapchat, facebook,	internet, consultancy, time, online, digital, digitale, internet, webdiensten, netwerk, it, ict,
	Infrastructure and cybersecurity	Infrastructure	itinfrastructuur, itnetwerk, itinfrastructuren, itbeheer, itdienstverlening, ictinfrastructuur, ictdienstverlening, ictbeheer, ictnetwerk, ictsystemen, gegevensuitwisseling, bedrijfsnetwerk, hotspots, wifi, 3G, 4G, 5G, SDN, it, ict, beheer, aanbieders, leverancier, handelaar, snelheid, bouwen, kanaal, kabel, hardware, import, netwerkbeheer, serverbeheer, ictoplossingen	software, digital, digitale, opslag, cloud, internet, webdiensten, netwerk, veiligheid, infrastructuur, it, ict
		Firewalls	firewall, firewalls, cyber, vpn, spyware, antispam, netwerkbeveiliging, internetbedreigingen, hacking, hackers, security, beveiliging, cybercrime, spam, phishing, tracking, pharming, risk, risico, incident, virus, managed, oplossingen, oplossing, beschermt, specialist, instellen, inbraak, verminking, wachtwoorden, creditcardgegevens, beschermen, beveiligd, blokkade, cybersecurity, Ddos, malware, ransomware	applicaties, geavanceerde, cyber, web, internet, webdiensten, netwerk, gegevens, eigendom, publieke, veiligheid
	Datamining & Big Data	Datamining & Big Data	robots, automatiseren, crawler, dataverzameling, datamining, textmining), webdesign, data, leverancier, aanbieder, handelaar, tekst, specialist, toepassen, google, overweegt, ontdek, patronen, intelligentie, ontwikkel, ingeprogrameerd, platform, recognition, machine, verzamelen, gebied, bedrijfsgegevens, verzamelt, science, big, dataanalytics, machinelearning,	data, kunstmatige, telecommunicatie, software, pakketten, learning, virtual, online, websites, app, apps, digital, digitale, elektronisch, applicaties, geavanceerde, opslag, cloud, internet, consultancy, webdiensten, netwerk, gegevens, veiligheid, it, ict, artificial, dataming

Appendix C - Complete set of tables

1. Number of businesses in the internet economy broken down by size, 2016-2018

		Size				
		Total	1 employed person	2-49 employed persons	5-249 employed persons	250 or more employed persons
2016	**Total**	1 555 740	1 175 550	367 180	10 240	2 770
	Internet category					
	Category A: no website	1 031 760	840 440	189 840	1 370	110
	Category B1: passive online presence	370 790	252 590	112 640	4 480	1 090
	Category B2: active online presence	104 920	52 570	47 570	3 390	1 380
	Category C: online stores	28 470	19 930	8 490	40	10
	Category D: online services	5 510	3 160	2 080	230	40
	Category E: internet related ICT	14 280	6 850	6 560	730	140
2017	**Total**	1 623 630	1 236 990	373 280	10 500	2 860
	Internet category					
	Category A: no website	1 062 740	874 420	186 780	1 420	120
	Category B1: passive online presence	396 410	272 330	118 410	4 530	1 130
	Category B2: active online presence	115 000	59 290	50 790	3 520	1 410
	Category C: online stores	30 250	21 270	8 910	60	10
	Category D: online services	5 320	3 080	1 980	220	50
	Category E: internet related ICT	13 910	6 600	6 420	760	140
2018	**Total**	1 702 840	1 314 530	374 470	10 830	3 010
	Internet category					
	Category A: no website	1 093 780	915 070	177 140	1 450	120
	Category B1: passive online presence	432 290	301 650	124 940	4 590	1 120
	Category B2: active online presence	126 470	66 630	54 580	3 710	1 550
	Category C: online stores	30 300	21 110	9 120	60	10
	Category D: online services	5 370	3 150	1 960	220	40
	Category E: internet related ICT	14 620	6 930	6 740	800	160

Source: CBS

2a. Number of businesses in the internet economy broken down by sector, 2016

		Internet category					
	Total	Category A: no website	Category B1: passive online presence	Category B2: active online presence	Category C: online stores	Category D: online services	Category E: internet related ICT
Total	1 555 740	1 031 760	370 790	104 920	28 470	5 510	14 280
Sector							
A Agriculture, forestry and fishing	72 320	65 470	5 250	1 380	200	10	10
B Mining and quarrying	420	290	120	20	0	0	0
C Manufacturing	60 890	33 600	20 560	5 140	1 390	40	160
D Energy supply	1 050	720	240	80	0	10	0
E Water supply and waste management	1 600	970	490	130	10	0	0
F Construction	156 420	121 710	31 080	2 930	540	40	120
G Wholesale and retail trade	224 790	130 580	52 650	21 970	17 630	440	1 540
H Transportation and storage	37 880	29 740	6 300	1 700	100	20	30
I Accommodation and food serving	52 930	34 100	10 490	7 980	320	40	20
J Information and communication	82 980	40 210	26 310	5 440	1 620	2 210	7 200
K Financial institutions	85 640	74 470	8 100	2 540	270	100	150
L Renting, buying, selling real estate	24 280	17 050	3 930	3 150	90	30	30
M Other specialised business services	313 600	196 160	92 720	16 400	2 850	1 270	4 210
N Renting and other business support	65 040	41 540	17 040	4 700	520	940	300
O Public administration and services	690	180	200	300	0	0	0
P Education	69 900	42 910	18 010	8 110	600	130	140
Q Health and social work activities	132 180	89 800	30 870	10 850	480	90	90
R Culture, sports and recreation	87 370	52 390	26 180	7 640	920	60	180
S Other service activities	85 750	59 870	20 260	4 470	940	80	120
T Activities of households	20	20	0	0	0	0	0
U Extraterritorial organisations	0	0	0	0	0	0	0

Source: CBS

2b. Number of businesses in the internet economy broken down by sector, 2017

		Internet category					
	Total	Category A: no website	Category B1: passive online presence	Category B2: active online presence	Category C: online stores	Category D: online services	Category E: internet related ICT
Total	1 623 630	1 062 740	396 410	115 000	30 250	5 320	13 910
Sector							
A Agriculture, forestry and fishing	73 680	65 780	5 970	1 700	220	10	10
B Mining and quarrying	430	290	120	20	0	0	0
C Manufacturing	62 820	33 960	21 620	5 530	1 520	50	140
D Energy supply	1 090	740	240	100	10	10	0
E Water supply and waste management	1 690	1 040	510	130	10	0	0
F Construction	163 890	125 600	34 250	3 320	570	50	110
G Wholesale and retail trade	227 410	126 870	56 140	23 840	18 700	420	1 440
H Transportation and storage	40 710	31 880	6 780	1 880	120	30	20
I Accommodation and food serving	54 960	34 520	11 270	8 740	360	20	40
J Information and communication	87 290	41 960	28 220	6 230	1 620	2 110	7 160
K Financial institutions	89 490	77 840	8 500	2 640	270	90	150
L Renting, buying, selling real estate	25 230	17 490	4 110	3 500	80	30	30
M Other specialised business services	325 320	202 450	96 890	17 720	3 010	1 210	4 060
N Renting and other business support	68 450	43 230	18 180	5 260	570	960	270
O Public administration and services	690	160	210	330	0	0	0
P Education	77 020	47 240	19 900	9 000	620	120	130
Q Health and social work activities	140 690	95 430	33 000	11 600	500	80	80
R Culture, sports and recreation	93 190	54 890	28 460	8 490	1 110	70	170
S Other service activities	89 540	61 340	22 050	4 980	970	80	120
T Activities of households	20	20	0	0	0	0	0
U Extraterritorial organisations	0	0	0	0	0	0	0

Source: CBS

2c. Number of businesses in the internet economy broken down by sector, 2018

		Internet category					
	Total	Category A: no website	Category B1: passive online presence	Category B2: active online presence	Category C: online stores	Category D: online services	Category E: internet related ICT
Total	1 702 840	1 093 780	432 290	126 470	30 300	5 370	14 620
Sector							
A Agriculture, forestry and fishing	74 100	65 630	6 400	1 840	210	10	10
B Mining and quarrying	460	320	130	20	0	0	0
C Manufacturing	64 990	34 800	22 560	5 890	1 530	40	170
D Energy supply	1 250	860	280	110	10	0	0
E Water supply and waste management	1 860	1 180	530	140	10	0	0
F Construction	176 410	136 350	35 780	3 570	530	60	130
G Wholesale and retail trade	229 550	122 330	60 900	25 740	18 640	430	1 500
H Transportation and storage	43 680	34 280	7 210	2 040	110	20	30
I Accommodation and food serving	57 090	34 760	12 110	9 780	370	30	40
J Information and communication	91 530	41 820	31 870	6 760	1 710	2 060	7 300
K Financial institutions	92 530	80 240	8 910	2 820	280	100	180
L Renting, buying, selling real estate	26 270	17 920	4 350	3 840	100	30	30
M Other specialised business services	342 680	206 730	107 770	19 540	3 100	1 200	4 340
N Renting and other business support	73 340	46 060	19 720	5 740	520	990	310
O Public administration and services	720	170	200	350	0	0	0
P Education	84 790	51 090	22 520	10 220	650	150	170
Q Health and social work activities	149 820	99 630	36 480	13 040	490	70	110
R Culture, sports and recreation	98 610	56 890	30 910	9 390	1 160	80	180
S Other service activities	93 130	62 710	23 670	5 650	890	90	120
T Activities of households	20	20	0	0	0	0	0
U Extraterritorial organisations	0	0	0	0	0	0	0

Source: CBS

3. Number of businesses in the internet economy broken down by age and size, 2016-2018

		Age							
		Total	0 year	1 year	2 year	3 year	4 year	5-9 year	10 or more years
2016	**Total**	1 555 740	166 800	143 380	124 320	100 210	100 470	404 870	515 690
	Internet category								
	Category A: no website	1 031 760	142 670	108 600	89 210	67 950	65 840	252 060	305 420
	Category B1: passive online presence	370 790	16 960	24 640	24 790	22 960	24 920	111 030	145 490
	Category B2: active online presence	104 920	4 530	6 420	6 490	5 870	6 180	27 810	47 630
	Category C: online stores	28 470	1 760	2 540	2 610	2 230	2 270	8 280	8 790
	Category D: online services	5 510	290	400	420	390	370	1 640	2 000
	Category E: internet related ICT	14 280	590	770	800	810	900	4 050	6 360
2017	**Total**	1 623 630	169 390	146 840	129 710	112 770	92 480	409 680	562 770
	Internet category								
	Category A: no website	1 062 740	146 490	117 650	94 900	77 590	60 340	248 140	317 630
	Category B1: passive online presence	396 410	16 150	20 360	24 310	24 590	22 670	117 100	171 220
	Category B2: active online presence	115 000	4 140	5 550	6 830	6 900	6 180	30 490	54 900
	Category C: online stores	30 250	1 870	2 300	2 580	2 570	2 130	8 670	10 140
	Category D: online services	5 320	250	310	350	350	360	1 570	2 130
	Category E: internet related ICT	13 910	480	670	740	770	810	3 710	6 740
2018	**Total**	1 702 840	184 740	148 940	132 940	118 120	103 880	405 460	608 760
	Internet category								
	Category A: no website	1 093 780	162 030	121 230	99 830	80 120	66 480	236 040	328 040
	Category B1: passive online presence	432 290	16 500	19 600	23 140	26 920	26 520	122 570	197 050
	Category B2: active online presence	126 470	3 980	5 220	6 530	7 470	7 410	32 930	62 920
	Category C: online stores	30 300	1 510	2 050	2 350	2 450	2 300	8 610	11 030
	Category D: online services	5 370	210	270	320	370	360	1 510	2 330
	Category E: internet related ICT	14 620	500	570	760	800	810	3 790	7 400

Source: CBS

4. Turnover in the internet economy broken down by size, 2016-2018

		Size				
		Total[1)]	1 employed person	2-49 employed persons	50-249 employed persons	250 or more employed persons
		Billion euro				
2016	**Total[1)]**	1 358	87	388	390	493
	Internet category					
	Category A: no website	205	55	93	42	14
	Category B1: passive online presence	642	24	186	208	224
	Category B2: active online presence	365	5	76	100	183
	Category C: online stores	12	3	4	2	4
	Category D: online services	19	0	3	3	12
	Category E: internet related ICT	115	1	25	35	55
2017	**Total[1)]**	1 457	93	420	412	532
	Internet category					
	Category A: no website	233	58	112	48	16
	Category B1: passive online presence	684	26	195	213	249
	Category B2: active online presence	385	6	82	112	185
	Category C: online stores	16	3	5	2	6
	Category D: online services	23	0	3	4	16
	Category E: internet related ICT	117	1	23	34	60
2018	**Total[1)]**	1 576	102	438	450	586
	Internet category					
	Category A: no website	251	63	117	51	20
	Category B1: passive online presence	733	29	201	228	275
	Category B2: active online presence	424	6	87	127	204
	Category C: online stores	18	3	5	2	8
	Category D: online services	23	0	3	4	15
	Category E: internet related ICT	127	1	25	38	65

Source: CBS

[1)] Total of the business economy. This does not include NACE categories A, K, O-U.

5. Employees in the internet economy broken down by size, 2016-2018

		Size				
		Total[1)]	1 employed person	2-49 employed persons	50-249 employed persons	250 or more employed persons
		x 1 000				
2016	Total[1)]	7 562	254	1 996	1 311	3 626
	Internet category					
	Category A: no website	996	176	587	168	64
	Category B1: passive online presence	2 675	52	792	540	1 291
	Category B2: active online presence	3 034	18	468	481	2 067
	Category C: online stores	40	6	19	5	10
	Category D: online services	90	2	21	32	35
	Category E: internet related ICT	353	2	107	85	159
2017	Total[1)]	7 774	268	2 057	1 362	3 730
	Internet category					
	Category A: no website	1 037	182	598	177	81
	Category B1: passive online presence	2 601	52	819	558	1 172
	Category B2: active online presence	3 278	27	494	502	2 256
	Category C: online stores	49	6	20	8	17
	Category D: online services	94	1	21	29	43
	Category E: internet related ICT	358	2	106	88	163
2018	Total[1)]	7 989	265	2 084	1 391	3 897
	Internet category					
	Category A: no website	1 010	184	590	173	62
	Category B1: passive online presence	2 663	54	829	560	1 219
	Category B2: active online presence	3 436	18	516	531	2 371
	Category C: online stores	50	5	19	8	18
	Category D: online services	106	1	19	30	56
	Category E: internet related ICT	374	1	111	90	171

Source: CBS
[1)] Categories do not add up to the grand total because not every employee could be matched to a company.

6. Jobs in the internet economy broken down by size, 2016-2018

		Size				
		Total[1)]	1 employed person	2-49 employed persons	50-249 employed persons	250 or more employed persons
		x 1 000				
2016	Total[1)]	8 023	278	2 119	1 372	3 835
	Internet category					
	Category A: no website	1 066	191	631	178	66
	Category B1: passive online presence	2 814	57	835	562	1 360
	Category B2: active online presence	3 223	20	500	505	2 199
	Category C: online stores	42	6	20	6	10
	Category D: online services	96	2	23	35	37
	Category E: internet related ICT	363	2	110	87	163
2017	Total[1)]	8 255	294	2 182	1 422	3 959
	Internet category					
	Category A: no website	1 108	197	642	186	83
	Category B1: passive online presence	2 731	57	863	580	1 231
	Category B2: active online presence	3 497	29	526	527	2 415
	Category C: online stores	52	6	21	8	17
	Category D: online services	100	2	22	32	44
	Category E: internet related ICT	367	2	108	89	168
2018	Total[1)]	8 476	290	2 208	1 449	4 134
	Internet category					
	Category A: no website	1 077	200	631	180	65
	Category B1: passive online presence	2 799	61	874	581	1 284
	Category B2: active online presence	3 659	21	549	557	2 532
	Category C: online stores	52	6	20	8	18
	Category D: online services	111	1	20	32	58
	Category E: internet related ICT	383	2	114	91	176

Source: CBS
[1)] Categories do not add up to the grand total because not every job could be matched to a company.

7. Production value, value added and employment in the internet economy, 2016-2018

		Production value (basic prices)	Value added (basic prices)	Employed persons	Employed persons (FTE)
		Billion euro		x 1 000	
2016	Total[1)]	974	409	5 770	4 593
	Internet category				
	Category A: no website	198	79	1 245	966
	Category B1: passive online presence	454	168	2 249	1 852
	Category B2: active online presence	227	115	1 750	1 336
	Category C: online stores	7	2	58	41
	Category D: online services	16	8	82	65
	Category E: internet related ICT	72	37	386	332
2017	Total[1)]	1 037	429	5 932	4 732
	Internet category				
	Category A: no website	211	85	1 266	988
	Category B1: passive online presence	479	172	2 227	1 846
	Category B2: active online presence	242	122	1 893	1 443
	Category C: online stores	7	3	68	48
	Category D: online services	18	8	91	72
	Category E: internet related ICT	80	38	388	335
2018	Total[1)]	1 100	453	6 108	4 883
	Internet category				
	Category A: no website	227	97	1 231	943
	Category B1: passive online presence	507	180	2 331	1 812
	Category B2: active online presence	260	127	1 975	1 390
	Category C: online stores	8	3	69	60
	Category D: online services	18	8	90	46
	Category E: internet related ICT	81	38	411	297

Source: CBS

[1)] Total of the business economy. This does not include NACE categories A, K, O-U.

8a. Local branches in the internet economy broken down by province, 2016

		Internet category					
	Total	Category A: no website	Category B1: passive online presence	Category B2: active online presence	Category C: online stores	Category D: online services	Category E: internet related ICT
Total	1 726 880	1 067 430	429 520	172 560	31 540	7 500	18 340
Province							
Groningen	50 530	30 040	13 210	5 490	1 010	230	540
Friesland	62 940	38 500	16 080	6 440	1 240	230	460
Drenthe	44 340	26 270	11 750	4 870	910	190	360
Overijssel	104 230	60 990	27 360	12 040	2 170	470	1 200
Flevoland	39 250	23 850	10 090	3 850	860	180	430
Gelderland	203 440	121 130	53 860	21 760	3 790	810	2 090
Utrecht	144 180	87 230	36 840	14 990	2 350	680	2 080
Noord-Holland	342 250	217 480	81 020	32 480	5 840	1 670	3 770
Zuid-Holland	347 830	221 490	81 700	33 350	6 130	1 530	3 640
Zeeland	36 360	23 300	8 190	3 960	610	120	190
Noord-Brabant	256 570	158 940	65 000	23 870	4 950	1 040	2 770
Limburg	94 970	58 210	24 430	9 470	1 690	360	820

Source: CBS

8b. Local branches in the internet economy broken down by province, 2017

	Internet category					
	Total	Category A: no website	Category B1: passive online presence	Category B2: active online presence	Category C: online stores	Category D: online services	Category E: internet related ICT
Total	1 799 040	1 099 240	456 160	185 180	33 660	6 900	17 910
Province							
Groningen	52 080	30 520	13 970	5 860	1 060	180	500
Friesland	64 860	38 710	17 140	7 020	1 340	210	430
Drenthe	45 830	26 580	12 540	5 220	980	160	360
Overijssel	107 710	61 910	29 220	12 780	2 290	390	1 120
Flevoland	40 950	24 630	10 540	4 280	930	160	420
Gelderland	210 910	123 570	57 430	23 190	3 950	770	2 010
Utrecht	150 370	90 180	38 990	15 960	2 560	660	2 020
Noord-Holland	360 060	228 210	85 630	34 610	6 260	1 540	3 820
Zuid-Holland	365 010	230 760	86 910	35 760	6 540	1 460	3 580
Zeeland	37 420	23 300	8 880	4 280	660	100	200
Noord-Brabant	265 760	162 050	68 890	25 850	5 300	950	2 720
Limburg	98 080	58 820	26 040	10 360	1 810	320	740

Source: CBS

8c. Local branches in the internet economy broken down by province, 2018

	Internet category					
	Total	Category A: no website	Category B1: passive online presence	Category B2: active online presence	Category C: online stores	Category D: online services	Category E: internet related ICT
Total	1 882 350	1 129 070	493 150	200 190	33 670	7 380	18 890
Province							
Groningen	53 820	30 740	15 010	6 200	1 070	230	580
Friesland	66 890	38 980	18 450	7 450	1 320	220	480
Drenthe	47 230	26 620	13 440	5 590	980	160	440
Overijssel	111 530	62 180	31 540	13 920	2 290	440	1 160
Flevoland	43 020	25 320	11 500	4 610	970	190	430
Gelderland	219 890	125 610	62 260	24 970	4 040	900	2 110
Utrecht	158 020	92 630	42 720	17 320	2 520	690	2 150
Noord-Holland	379 460	236 850	92 900	37 770	6 310	1 680	3 940
Zuid-Holland	386 710	242 470	93 820	38 780	6 510	1 360	3 770
Zeeland	38 850	23 760	9 470	4 620	650	120	220
Noord-Brabant	275 860	164 550	74 240	27 930	5 260	1 060	2 810
Limburg	101 070	59 370	27 790	11 020	1 760	340	790

Source: CBS

9a. Local branches in the internet economy broken down by COROP-area, 2016

	Internet category						
	Total	Category A: no website	Category B1: passive online presence	Category B2: active online presence	Category C: online stores	Category D: online services	Category E: internet related ICT
Total	1 726 880	1 067 430	429 520	172 560	31 540	7 500	18 340
COROP-area							
Oost-Groningen	11 440	6 960	2 790	1 330	260	40	70
Delfzijl en omgeving	3 480	2 210	820	370	50	10	20
Overig Groningen	35 610	20 880	9 610	3 790	700	180	450
Noord-Friesland	29 770	18 300	7 570	2 980	600	110	210
Zuidwest-Friesland	15 100	9 360	3 880	1 440	270	50	100
Zuidoost-Friesland	18 070	10 830	4 630	2 020	380	70	150
Noord-Drenthe	17 220	9 930	4 780	1 900	370	80	160
Zuidoost-Drenthe	14 430	8 930	3 580	1 490	260	50	110
Zuidwest-Drenthe	12 700	7 410	3 390	1 480	270	60	90
Noord-Overijssel	34 810	20 580	9 050	4 000	660	160	370
Zuidwest-Overijssel	13 740	7 810	3 740	1 690	270	70	160
Twente	55 680	32 600	14 570	6 360	1 250	240	670
Veluwe	67 260	40 090	17 380	7 510	1 250	300	740
Achterhoek	39 560	23 390	10 720	4 230	780	130	310
Arnhem/Nijmegen	69 020	40 080	19 070	7 510	1 290	310	760
Zuidwest-Gelderland	27 590	17 580	6 700	2 510	470	70	270
Utrecht	144 180	87 230	36 840	14 990	2 350	680	2 080
Kop van Noord-Holland	36 860	23 410	8 750	3 650	690	110	240
Alkmaar en omgeving	24 470	14 550	6 260	2 760	520	110	270
IJmond	17 690	10 920	4 390	1 800	360	80	140
Agglomeratie Haarlem	28 320	17 460	7 000	2 880	510	150	330
Zaanstreek	15 230	9 600	3 670	1 470	320	40	130
Groot-Amsterdam	185 410	120 300	42 210	16 740	2 850	1 000	2 310
Het Gooi en Vechtstreek	34 270	21 240	8 750	3 170	580	170	360
Agglomeratie Leiden en Bollenstreek	39 990	24 700	9 780	4 290	660	130	430
Agglomeratie 's-Gravenhage	84 810	56 250	18 320	7 770	1 280	350	840
Delft en Westland	23 490	14 550	5 870	2 280	420	100	280
Oost-Zuid-Holland	32 590	20 190	8 050	3 270	600	110	360
Groot-Rijnmond	130 870	83 170	31 000	12 210	2 470	710	1 320
Zuidoost-Zuid-Holland	36 090	22 650	8 670	3 540	700	130	400
Zeeuwsch-Vlaanderen	9 480	6 250	2 030	960	160	30	50
Overig Zeeland	26 880	17 050	6 160	3 000	450	90	140
West-Noord-Brabant	62 400	38 970	15 520	5 810	1 190	280	640
Midden-Noord-Brabant	46 300	28 700	11 810	4 270	920	180	410
Noordoost-Noord-Brabant	70 400	43 880	17 690	6 470	1 370	260	730
Zuidoost-Noord-Brabant	77 470	47 390	19 980	7 320	1 480	320	990
Noord-Limburg	24 370	14 830	6 330	2 500	470	60	180
Midden-Limburg	22 200	13 700	5 680	2 130	420	90	190
Zuid-Limburg	48 400	29 680	12 410	4 850	800	210	450
Flevoland	39 250	23 850	10 090	3 850	860	180	430

Source: CBS

9b. Local branches in the internet economy broken down by COROP-area, 2017

	Internet category						
	Total	Category A: no website	Category B1: passive online presence	Category B2: active online presence	Category C: online stores	Category D: online services	Category E: internet related ICT
Total	1 799 040	1 099 240	456 160	185 180	33 660	6 900	17 910
COROP-area							
Oost-Groningen	11 730	6 990	2 960	1 440	260	30	60
Delfzijl en omgeving	3 470	2 150	860	370	60	0	20
Overig Groningen	36 890	21 380	10 140	4 050	740	150	430
Noord-Friesland	30 710	18 320	8 150	3 290	650	100	200
Zuidwest-Friesland	15 570	9 450	4 080	1 600	280	60	100
Zuidoost-Friesland	18 580	10 950	4 910	2 140	410	60	130
Noord-Drenthe	17 870	10 120	5 080	2 070	380	70	150
Zuidoost-Drenthe	14 800	8 980	3 800	1 580	290	40	120
Zuidwest-Drenthe	13 160	7 490	3 660	1 570	300	50	90
Noord-Overijssel	36 140	20 930	9 750	4 270	710	130	350
Zuidwest-Overijssel	14 280	8 040	3 960	1 790	290	50	160
Twente	57 290	32 940	15 510	6 720	1 300	210	620
Veluwe	69 780	40 930	18 470	8 060	1 350	270	700
Achterhoek	40 530	23 410	11 400	4 510	760	130	320
Arnhem/Nijmegen	72 230	41 480	20 410	7 980	1 350	310	710
Zuidwest-Gelderland	28 370	17 760	7 140	2 640	490	60	280
Utrecht	150 370	90 180	38 990	15 960	2 560	660	2 020
Kop van Noord-Holland	38 220	23 750	9 160	4 220	760	120	220
Alkmaar en omgeving	25 570	15 130	6 610	2 940	560	90	240
IJmond	18 500	11 380	4 610	1 940	390	60	130
Agglomeratie Haarlem	29 610	18 190	7 370	3 060	550	130	300
Zaanstreek	16 060	10 070	3 900	1 570	360	50	120
Groot-Amsterdam	196 500	127 820	44 760	17 520	3 010	910	2 480
Het Gooi en Vechtstreek	35 610	21 870	9 210	3 390	630	180	340
Agglomeratie Leiden en Bollenstreek	41 760	25 520	10 440	4 560	710	130	400
Agglomeratie 's-Gravenhage	90 030	59 710	19 340	8 370	1 440	320	850
Delft en Westland	24 410	14 840	6 380	2 360	410	100	310
Oost-Zuid-Holland	34 000	20 870	8 500	3 500	670	120	340
Groot-Rijnmond	137 440	86 660	33 130	13 100	2 590	680	1 290
Zuidoost-Zuid-Holland	37 380	23 160	9 120	3 870	730	110	400
Zeeuwsch-Vlaanderen	9 670	6 220	2 190	1 020	170	20	50
Overig Zeeland	27 750	17 070	6 690	3 260	490	80	150
West-Noord-Brabant	64 590	39 700	16 300	6 420	1 300	230	640
Midden-Noord-Brabant	48 010	29 400	12 480	4 590	980	160	400
Noordoost-Noord-Brabant	72 830	44 600	18 750	7 040	1 470	260	720
Zuidoost-Noord-Brabant	80 320	48 350	21 360	7 800	1 550	300	970
Noord-Limburg	25 190	15 010	6 760	2 730	470	70	160
Midden-Limburg	23 000	13 900	6 070	2 320	460	80	170
Zuid-Limburg	49 900	29 910	13 210	5 320	880	170	400
Flevoland	40 950	24 630	10 540	4 280	930	160	420

Source: CBS

9c. Local branches in the internet economy broken down by COROP-area, 2018

	Internet category						
	Total	Category A: no website	Category B1: passive online presence	Category B2: active online presence	Category C: online stores	Category D: online services	Category E: internet related ICT
Total	1 882 350	1 129 070	493 150	200 190	33 670	7 380	18 890
COROP-area							
Oost-Groningen	11 150	6 450	2 930	1 410	250	30	70
Delfzijl en omgeving	3 530	2 140	930	380	50	10	20
Overig Groningen	39 150	22 140	11 160	4 410	770	190	490
Noord-Friesland	31 090	18 090	8 600	3 430	640	110	220
Zuidwest-Friesland	16 750	9 990	4 600	1 710	290	50	120
Zuidoost-Friesland	19 050	10 900	5 250	2 310	390	50	150
Noord-Drenthe	18 450	10 100	5 460	2 290	380	50	170
Zuidoost-Drenthe	15 160	8 970	4 030	1 660	300	60	140
Zuidwest-Drenthe	13 630	7 560	3 950	1 640	310	50	130
Noord-Overijssel	37 660	21 280	10 520	4 640	730	140	350
Zuidwest-Overijssel	14 900	8 160	4 310	1 900	290	70	160
Twente	58 980	32 740	16 710	7 380	1 270	230	650
Veluwe	72 900	41 580	20 130	8 700	1 380	340	760
Achterhoek	41 760	23 350	12 320	4 880	760	140	320
Arnhem/Nijmegen	75 790	42 570	22 130	8 600	1 390	340	760
Zuidwest-Gelderland	29 450	18 110	7 670	2 790	510	80	290
Utrecht	158 020	92 630	42 720	17 320	2 520	690	2 150
Kop van Noord-Holland	39 570	24 010	9 910	4 490	740	140	270
Alkmaar en omgeving	27 100	15 680	7 270	3 230	550	110	270
IJmond	19 510	11 780	5 060	2 040	410	70	150
Agglomeratie Haarlem	31 390	18 780	8 150	3 380	580	150	340
Zaanstreek	17 050	10 590	4 260	1 710	320	50	130
Groot-Amsterdam	207 790	133 810	48 310	19 230	3 050	970	2 440
Het Gooi en Vechtstreek	37 050	22 200	9 950	3 700	660	190	350
Agglomeratie Leiden en Bollenstreek	43 720	26 170	11 340	4 950	720	140	400
Agglomeratie 's-Gravenhage	96 930	64 220	21 060	9 000	1 430	380	840
Delft en Westland	25 690	15 480	6 790	2 630	400	80	310
Oost-Zuid-Holland	35 610	21 490	9 290	3 710	670	120	340
Groot-Rijnmond	145 660	91 190	35 580	14 310	2 590	510	1 490
Zuidoost-Zuid-Holland	39 100	23 920	9 770	4 180	700	130	390
Zeeuwsch-Vlaanderen	9 900	6 220	2 340	1 080	180	30	60
Overig Zeeland	28 950	17 540	7 140	3 540	470	90	170
West-Noord-Brabant	67 010	40 360	17 490	6 930	1 290	270	660
Midden-Noord-Brabant	50 120	30 060	13 500	4 980	970	190	420
Noordoost-Noord-Brabant	75 240	44 910	20 210	7 680	1 460	250	740
Zuidoost-Noord-Brabant	83 490	49 220	23 040	8 340	1 540	350	1 000
Noord-Limburg	26 060	15 330	7 160	2 880	450	60	170
Midden-Limburg	23 710	13 980	6 480	2 530	460	80	190
Zuid-Limburg	51 290	30 060	14 150	5 610	850	200	420
Flevoland	43 020	25 320	11 500	4 610	970	190	430

Source: CBS

10. Turnover developments (growing and declining) in the internet economy, 4th quarter 2016-2018

		Businesses with a growing turnover	Businesses with a stable turnover	Businesses with a declining turnover
		%		
2016 4th quarter	**Total**[1)]	41	22	37
	Internet category			
	Category A: no website	39	24	37
	Category B1: passive online presence	45	18	37
	Category B2: active online presence	47	18	34
	Category C: online stores	50	15	36
	Category D: online services	47	16	37
	Category E: internet related ICT	47	16	37
2017 4th quarter	**Total**[1)]	40	22	37
	Internet category			
	Category A: no website	38	25	37
	Category B1: passive online presence	44	19	38
	Category B2: active online presence	46	19	36
	Category C: online stores	46	14	40
	Category D: online services	48	15	37
	Category E: internet related ICT	48	15	37
2018 4th quarter	**Total**[1)]	42	23	35
	Internet category			
	Category A: no website	40	25	35
	Category B1: passive online presence	44	20	36
	Category B2: active online presence	47	19	34
	Category C: online stores	47	15	38
	Category D: online services	47	17	36
	Category E: internet related ICT	49	16	35

Source: CBS

[1)] Total of the business economy. This does not include NACE categories A, K, O-U.

11. International trade of goods in the internet economy; import and export values, 2016-2018

		Import value	Export value
		million euro	
2016	**Total**[1)]	368 864	423 236
	Internet category		
	Category A: no website	44 540	54 069
	Category B1: passive online presence	135 308	160 445
	Category B2: active online presence	67 141	58 949
	Category C: online stores	2 689	1 195
	Category D: online services	709	425
	Category E: internet related ICT	24 717	20 786
2017	**Total**[1)]	408 871	467 435
	Internet category		
	Category A: no website	50 313	62 859
	Category B1: passive online presence	151 529	174 828
	Category B2: active online presence	69 253	59 768
	Category C: online stores	3 518	1 662
	Category D: online services	563	642
	Category E: internet related ICT	28 429	22 571
2018	**Total**[1)]	441 278	497 903
	Internet category		
	Category A: no website	55 703	67 435
	Category B1: passive online presence	150 135	172 909
	Category B2: active online presence	83 267	74 459
	Category C: online stores	3 377	1 819
	Category D: online services	435	392
	Category E: internet related ICT	31 048	24 555

Source: CBS

[1)] Categories do not add up to the grand total because the total also includes the exports and imports of foreign companies without a Dutch establishment, but of which the trade does go via the Netherlands.

Appendix D - Indicators

Indicator	Definition
(Number of) businesses	The actual transactor in the production process characterised by independence in decisions about the process and by providing products to others.
Age	The number of years between December 1 of the reporting year and the founding date of the business. This date is corrected for mergers, take-overs, spin-offs etc.
Size	The size of a business in terms of number of employed persons. The number of "employed persons" consists of: - employees and (managing) directors on the payroll; - employees on the payroll of another enterprise or organisation, but working for the company itself and therefore in fact employees (hired staff/ personnel); - employed owners, members of the firm, members of a limited liability partnership, partners and employed family members; - temps. In this study we group BUs into the following 4 size classes: 1 employed person, 2-49 employed persons, 50-249 employed persons and 250 or more employed persons.
Sector	“Sector of industry” or “branch” is the common term used to define businesses sharing the same main activity. Statistics Netherlands uses the classification by main activity, the so-called Standard Industrial Classification (SIC). Businesses in a sector of industry or branch may also be engaged in other activities (subsidiary activities).
Employees	A person who has a contract with an economic unit to carry out work in return for financial remuneration.
Jobs of employees	A job is an explicit or implicit employment contract between a person and an economic unit which specifies that labour will be carried out in return for (financial) remuneration.
Province region	Administrative part of the Dutch territory. The division of the Netherlands into 12 Province areas is the equivalent of the European NUTS 2 level.
COROP region	One or more contiguous municipalities in a province. The division of the Netherlands into 40 COROP areas is the equivalent of the European NUTS 3 level.
Turnover	Value of sales of goods and services to third parties, VAT excluded. Turnover includes main activity as well

	as secondary activities. Third parties are consumers and enterprises not belonging to (the Dutch part of) the own group of companies.
Value added	The value of all goods and services produced (production value or output), minus those that have been intermediately used upon production (intermediate consumption). Value added is rated at basic prices: purchaser's prices minus trade and transport margins and taxes on products paid and plus subsidies on products received. Intermediate consumption is rated at purchaser's prices minus non-deductible VAT. Included is the output by all kind-of-activity units residing in the Netherlands, also those that are held by foreign owners.
Production value	The value of the goods and services produced, valued at basic prices.
Export value	The value of all goods, including costs insurance freight (CIF) up to the Dutch border. The goods are domestically produced or imported.
Import value	The value of all goods brought into the Dutch economy for consumption. This is the case when import duties and national taxes have been paid. For imports from EU countries this is the value of the goods including freight and insurance up to the Dutch border (CIF). For imports from non-EU countries this is the value including freight and insurance up to the border of the European Union.
Employed persons	Someone working for a company located in the Netherlands, or an institution or private household in the Netherlands.
Full-time equivalents	A measure of labour volume, calculated by converting all full-time and part-time jobs to full-time jobs

Explanation of symbols

Empty cell	Figure not applicable
.	Figure is unknown, insufficiently reliable or confidential
*	Provisional figure
**	Revised provisional figure
2017–2018	2017 to 2018 inclusive
2017/2018	Average for 2017 to 2017 inclusive
2017/'18	Crop year, financial year, school year, etc., beginning in 2017 and ending in 2018
2013/'14–2017/'18	Crop year, financial year, etc., 2013/'14 to 2017/'18 inclusive

Due to rounding, some totals may not correspond to the sum of the separate figures.

Colophon

Publisher
Statistics Netherlands
Henri Faasdreef 312, 2492 JP Den Haag
www.cbs.nl

Prepress
CCN Creatie, The Hague

Design
Edenspiekermann

Information
Telephone: +31 88 570 7070
Via contact form: www.cbs.nl/information