Evaluation of Digital Development Based on the International Digital Economy and Society Index 2020 Data

Janka Ariella Tarjáni[1] | *Budapest University of Technology and Economics, Budapest, Hungary*
Noémi Kalló[2] | *Budapest University of Technology and Economics, Budapest, Hungary*
Imre Dobos[3] | *Budapest University of Technology and Economics, Budapest, Hungary*

Received 18.5.2023 (revision received 21.7.2023), **Accepted** (reviewed) 26.7.2023, **Published** 15.9.2023

Abstract

Digitalization and technological advances develop together with our society. The European Commission intends to monitor this development using the International Digital and Society Index (I-DESI) and provide an objective comparison of the participating countries. This comparison of the counties can be an important part of the roadmap of digital transformation for companies and other participants in the market. The aim of this study is to further analyze the 2020 data of I-DESI using multivariate statistical methods not included in the official report. Our objective is to show whether there are differences between the EU and non-EU countries (discriminant analysis, variance analysis), whether the dimensions of the I-DESI index are overlapping (correlation analysis, factor analysis), and whether different country groups can be formed (cluster analysis). Answering these questions, we can give a useful tool to companies for a more successful digital transformation.

Keywords	***DOI***	***JEL code***
International Digital Economy and Society Index (I-DESI), digital transformation, Information and Communication Technology (ICT), analytics, Multivariate Statistical Analysis	*https://doi.org/10.54694/stat.2023.21*	*C3, H83, L86, L96*

1 Department of Economics, Faculty of Economic and Social Sciences, Budapest University of Technology and Economics, Műegyetem rkp. 3, 1111 Budapest, Hungary. E-mail: tarjani.janka@gtk.bme.hu. ORCID: <https://orcid.org/0000-0002-3912-4022>.

2 Department of Economics, Faculty of Economic and Social Sciences, Budapest University of Technology and Economics, Műegyetem rkp. 3, 1111 Budapest, Hungary. E-mail: kallo.noemi@gtk.bme.hu. ORCID: <https://orcid.org/0000-0003-3193-081X>.

3 Department of Economics, Faculty of Economic and Social Sciences, Budapest University of Technology and Economics, Műegyetem rkp. 3, 1111 Budapest, Hungary. E-mail: dobos.imre@gtk.bme.hu. ORCID: <https://orcid.org/0000-0001-6248-2920>.

INTRODUCTION

Monitoring the digital and related social changes in different countries can provide steady and controlled development of our society. The European Commission uses a complex indicator, the International Digital Economy and Society Index (I-DESI), to monitor the digital development of EU countries and provide comparisons with the rest of the world (European Commission, 2021a). This monitoring activity of technological progress became a key responsibility of the European Commission, as it gives valuable waypoints for the individual countries regarding their potential improvements, and it ensures the EU's competitiveness against other countries such as the US, Japan and South Korea.

The I-DESI indicator consists of five main dimensions. These are *Connectivity*, which represents the high-speed internet access and the mobile network coverage, *Human Capital*, which represents the ability of the population to consume online content and to take part in online activities, *Use of Internet* and *Integration of Digital Technology*, which illustrates the internet usage of citizens and businesses, and *Public Services*, which shows the demand and supply for online services in the public administration field. The main dimensions are interpreted and measured through 24 subdimensions. The data is therefore a continuous multivariate data set measured on an interval scale, where several samples (countries) on each unit are measured with several variables. I-DESI uses a similar weighting system as used for DESI, another index for only EU Member States. The reasons for the differences are that ultrafast broadband and e-healthcare data are not available from certain non-EU countries, so these sub-indicators have been omitted from the I-DESI indicator. According to the report of the European Commission (2021c), the correlation between the main dimensions and subdimensions of DESI and I-DESI is strong with a value of 0.89 between the scores for 2015–2018 data and the country ranking thus a comparison of the two indicators is reliable.

The European Commission's official report (European Commission, 2021a) for 2020 is based on a trend analysis of data collected since 2015 for the 27 EU Member States and 18 non-EU participants (Australia, Brazil, Canada, Chile, China, Iceland, Israel, Japan, Mexico, New Zealand, Norway, Russia, Serbia, South Korea, Switzerland, Turkey, the United Kingdom and the United States) According to the trend analysis, Finland leads the ranking, and the most advanced non-EU country is the United States.

In this study, we analyze the I-DESI 2020 data set using multivariate statistical methods not included in the official report. The aim is to compare the results of EU Member States with those of non-EU participants. For successful digital transformation, it is important to know for the businesses how well the different regions are performing in digitalization. Furthermore, analytics is one of the most important tools of digitalization, so it seems to be implicit in analyzing all data about digital intensity (as a dimension of digital maturity) of different regions.

Accordingly, we thoroughly analyze I-DESI data, and the structure of our study is as follows. In the literature review, we provide a brief overview of the latest publications on measuring digitization and analyzing international indicators. Next, we summarize our research questions and the statistical methods used. Analyses and results include the following statistical analyses: discriminant analysis, comparison of means, correlation analysis, principal component analysis, partial correlation analysis, and cluster analysis. Finally, we summarize our Conclusion.

1 LITERATURE REVIEW

There are only a few publications in the international literature examining the I-DESI indicator. Sources typically compare the DESI system with other measurement methods and ranking procedures, look for a link between DESI and other indicators, or present results in a national or regional analysis. The I-DESI is rarely in the centre of attention.

A detailed summary was prepared of the measurement methods of the digitalized economy (Kokh and Kokh, 2019). Besides the DESI and I-DESI, the authors included the ICT (Information and Communication

Technology) Development Index, the Huawei Global Network Index, the eGovernment Development Index, the Boston Consulting Group Economic Digitization Index, the Global Digital Competitiveness Index, the Digital Evolution Index, and the Ivanov Digital Index. After analyzing and comparing the calculation methodology of the indices, the authors concluded that all listed indices are global and suitable to characterize the countries in terms of digital development. The authors further explain that there are no indicators that measure the level of digitization of individual industries and services that could be used for sector-specific analysis.

The relationship between the DESI and other indices was also examined. The impact of consumption index growth, purchasing power parity and unemployment on the DESI between 2013 and 2018 were investigated in a study (Stavytskyy et al., 2019). Their results confirm that 98% of DESI values are determined by previous years' data and that a 1% increase in unemployment is a 0.2% decrease in DESI, a 1% increase in the consumption index comes with a 0.2% increase in DESI, so is an increase in the market index is accompanied by an increase in the DESI value. DESI dimensions were also used to assess the impact of financial markets and institutions on digital development (Ha, 2022). Among the DESI dimensions, the role of Human Capital was highlighted as the main factor and digitalization was found to have a significant effect on financialization. The relationship between DESI and GDP was also confirmed (Turuk, 2021) and the current situation of digital enterprises in Central and Eastern Europe using the DESI indicator and the GDP was examined. Using data collected from 2015 to 2019, it showed the relationship between countries' GDP per capita and the DESI. According to his calculations, among the DESI dimensions, the Use of Internet, the Integration of Digital Technologies and Public Services have a positive effect on the development of GDP per capita. The study did not show a significant relationship between the impact of internet access and human capital on GDP per capita.

The relationship between the dimensions of DESI and labour market indicators were also investigated (Başol and Yalçin, 2020). In this, the authors compared 2018 DESI data dimensions with positive employment indicators (personal earnings, employment rate) and negative employment indicators (labour market insecurity, long-term unemployment rate) with correlation analysis and regression calculation. They concluded that with the increase in DESI, both the employment rate and personal earnings would increase, as well as the long-term unemployment rate and labour market insecurity, so digital development will improve positive employment indicators. Others studied the relationship between digitalization and labour productivity and the global competitiveness index using cluster analysis (Polozova et al., 2021). The analysis identified four clusters, leaders, prospective countries, followers, and transition countries. The Nordic countries are also at the forefront of this analysis. They managed to prove the relationship between DESI and labour productivity, while the relationship between DESI and the competitiveness index was not clear. The relationship between unemployment and digital development was also proved (Mirzaei and Soleimani, 2021). These indicators have a saddle-shaped connection. With digital development unemployment increased in this study to a certain maximum, but this effect is said to be possible to prevent with cautious digital expansion. The effect of digitalization on public health was also studied using DESI dimensions and the Eurobarometer survey (Moreno-Llamas et al., 2020). Sedentary behaviour, such as the number of hours a person sits, was found to be in a positive linear relationship with the indices of digital development and e-device ownership at the country level.

The relationship between sustainability and digitization in the data of the Visegrád countries was analyzed in another regional study (Esses et al., 2021). In their study, the authors proved correlations between the dimensions of DESI, countries' GDP, Human Development Index (HDI) and Social Progress Index (SPI). Their results cover the effects of Covid-19, which led to a leap forward in the digitization of countries in 2020 compared to the previous years. Another example of a regional analysis proposed to rebuild the technological development monitoring framework of the region in Abruzzo, Italy, following the European guidelines for the DESI index (Russo, 2020). In his study, he maps the digital economy of Abruzzo, describes the course of digitization over time and creates a unique data collection system to track the progress. An analysis of the state of digitization was also prepared in Romania (Gurău, 2021).

Earlier the 2019 DESI results were studied with multivariate statistical methods (Bánhidi et al., 2020; Bánhidi et al., 2021). Similar to the present study, the authors examined linear relationships between dimensions using correlation analysis as well as principal component analysis, and a causal chain is also established by partial correlation analysis. This study only included the data of EU Member States. These countries were grouped by cluster analysis. In our study, we supplement this series of studies by discriminating the data of the EU member states and non-EU countries and comparing their results. An alternative ranking method was presented based on the statistical properties of the DESI composite index data sets (Bánhidi et al., 2021). They used Data Envelopment Analysis (DEA) and composite indices (CI), and the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS). Based on the results, the Nordic countries (Finland, Sweden, Denmark) lead the ranking just as the original DESI ranking.

From different data and from the I-DESI index, it is clear that most countries are ready for digital transformation. There are differences but digital transformation is well-researched in different countries (e.g., Russia (Nissen et al., 2018), Greece (Bousdekis and Kardaras, 2020), Slovenia and Hungary (Erjavec et al., 2018), or Denmark (Scupola, 2018)) and different fields (e.g., education (Teixeira et al., 2021), health care (Burton-Johnres et al., 2020), management consultancy (Tarjáni et al., 2021) or banking (Cuesta et al., 2015)). This really short overview of studies shows that there are big differences in the depth and advances of digital transformation in the different regions of the world. Making it necessary to analyze the data on digital characteristics as thoroughly as possible.

2 STATISTICAL METHODS

In this paper, we analyzed the data of the International Digital Economy and Society Index published in December 2020 (European Commission, 2021c) using multivariate statistical methods. The data set is shown in the Appendix (Table A1). We selected the methods based on an analysis of the previous year (Bánhidi et al., 2020), in order to compare the results from 2019 and 2020. Our research questions and selected methods are:

Q1. Is it possible to separate the dataset into groups of EU and non-EU countries?
(discriminant analysis)

Q2. Is there a difference between EU and non-EU averages?
(analysis of variance, ANOVA)

Q3. What linear relationships can be detected between the dimensions of the indicator system?
(correlation analysis)

Q4. How can we reduce the number of model variables?
(principal component analysis)

Q5. What causal relationships can be considered between the model variables?
(partial correlation analysis)

Q6. How can we group the studied countries?
(cluster analysis)

We began the data analysis with graphical studies. Then we used Wilks's λ and canonical correlation in the discriminant analysis. To compare the means, the conditions of the analysis of variance were checked with the QQ plot, Shaphiro-Wilk, Kolmogorov-Smirnov, and Box test. Pearson correlation coefficients were used for the correlation test. Outliers were evaluated based on Mahalanobis distances. Due to the strong correlation revealed between the dimensions (0.35–0.82), the number of variables was reduced by principal component analysis without rotation and with Varimax rotation. The suitability of the sample was checked with a Kaiser-Meyer-Olkin measure (0.804), and the model was evaluated with Bartlett's test.

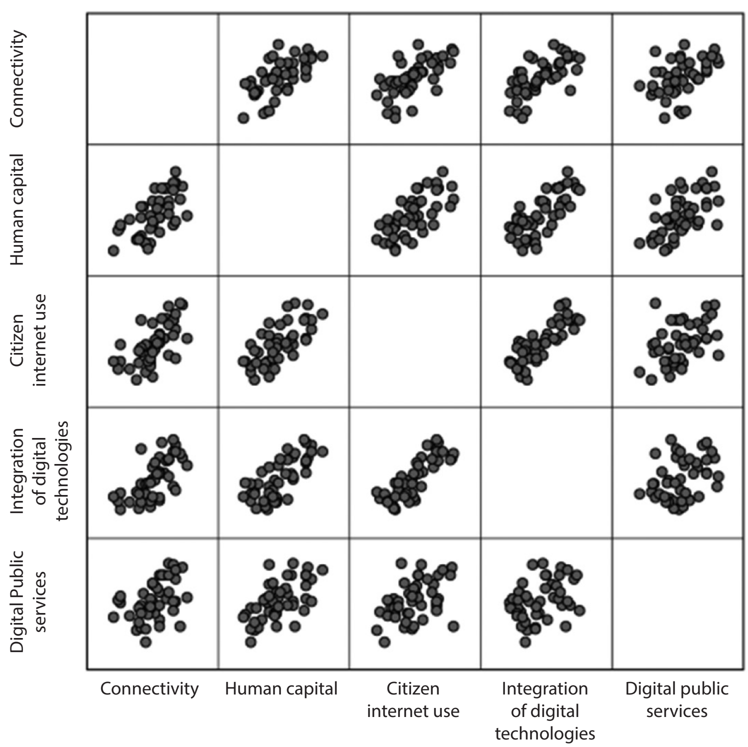
The presumed causal chain between the variables was established by a partial correlation test performed at a 15 percent significance level. Hierarchical cluster analysis was used to group the countries. The basis of group formation was the relationship within the group. The squared Euclidean distance was used to generate the distances. The result was plotted on a dendrogram. For the calculations, we used the IBM Statistics Package for Social Sciences (SPSS) statistical software.

3 ANALYSES AND RESULTS

Graphical examinations help to form an overall impression of the data set, whether there are outliers, differences between groups, or other anomalies. To illustrate multivariate data sets, Rencher (2002) suggests profiles, cobweb diagrams, characteristic signs, and boxes. The statistical software package only supports profiles (bar charts) among these. Figure A1 shows the I-DESI data of the studied countries.

Visual inspection of the data reveals that there are countries with low I-DESI values among both the EU Member States (such as Greece and Poland) and non-EU countries (such as Brazil and Mexico). High values are also present in EU countries (such as Finland and the Netherlands) and non-EU countries (such as Korea and the United States). Some countries have a balanced profile with all I-DESI components close to each other. Examples of such balanced EU countries are Austria, Ireland, Luxembourg and Sweden. Similar examples of the non-EU group are the United Kingdom and the United States. In contrast, there are countries where the value of one component is twice as high as that of the other. In Greece, the value

Figure 1 Bivariate scatter plots



Source: Calculation on the European Commission data (EC, 2021b)

of Connectivity and Public Services are three times the value of Human Capital and twice the value of the Use of Internet and the Integration of Digital Technologies. A similar phenomenon can be observed in Brazil and China. The value of Connectivity is typically high in EU countries, but in some cases, the value of Public Services or the Integration of Digital Technologies is the highest among the dimensions (e.g. Finland, Germany, the Netherlands and Sweden). According to Figure A1 the values of Public Services are generally high in non-EU countries, in four cases Connectivity has the highest value. The Integration of Digital Technologies ranks first in Israel and Switzerland, and the Use of Internet in Iceland is the highest of the five indicators.

Possible outliers are Denmark and the USA based on graphical analysis, as they are outside of the +1.5 interquartile range. If these two data are proved to be outliers, it is better to exclude them from further analyses. To prove the case of outliers, graphical and numerical tests are used. The graphical test is based on the bivariate scatter plots in Figure 1. The plots show possible correlations of the variables, probably because they are all related to digitalization. These correlations are not investigated further in this paper. The most scattered relationship is between citizen internet use and public services and between business technology and public services. Possible outliers are only present in the plot between citizen internet use and public services, but these data points do not belong to Denmark or the USA. Based on the bivariate scatter plots, the cases of outliers for Denmark and the USA are not proved.

Numerical analysis of the outliers is based on the Mahalanobis distances. Mahalanobis distances are calculated for all data points, and the highest values are present in Table 1. These distances are compared with a $\chi 2$ distribution of the same degrees of freedom to provide a p-value in the table. Even the highest values of the Mahalanobis distances are not significant, which confirms the result of the graphical analysis with no outliers presumed.

Table 1 Mahalanobis distances

	Mahalanobis distance	Sig.
Iceland	12.17295	0.0325
Korea	10.24161	0.0687
France	8.20925	0.1451
Lithuania	8.18205	0.1465
Israel	7.92057	0.1607

Source: Calculation on the European Commission data (EC, 2021b)

3.1 Separation of the dataset (discriminant analysis)

After the graphical evaluation, we examined whether the data of the EU member states differ from the results of the non-EU participants by discriminant analysis. Since the two samples did not have the same number of items, we corrected the data for the size of the groups. The eigenvalue was 0.208, and the canonical correlation was 0.415. For two groups, the canonical correlation is the most useful measure in the table and coincides with the Pearson correlation coefficient between scores and groups. The value of the canonical correlation is weak, below 0.5, thus, it is not possible to separate the two data sets based on the data. Wilks's λ value was 0.828 with a significance of 0.177. The high value of Wilks's λ and the low significance score shows that there is no difference between the averages of EU and non-EU countries according to the classification with the grouping variable, and the groups cannot be separated based on the data set alone. The related χ^2 test examines the hypothesis that data from EU and non-EU countries come from the same population. As the result is not significant, this grouping of the EU and non-EU countries could even be random.

3.2 Comparison of country groups (analysis of variance)

Analysis of variance (ANOVA) examines the difference between means. The assumptions required for the reliability of the analysis are the independence of the samples, the multivariate normality, and the equality of the covariance matrices (homogeneity of variance). The independence of the samples can be assumed in our case because the different samples come from different countries. Multivariate normality and equality of variance were checked from graphical and numerical studies.

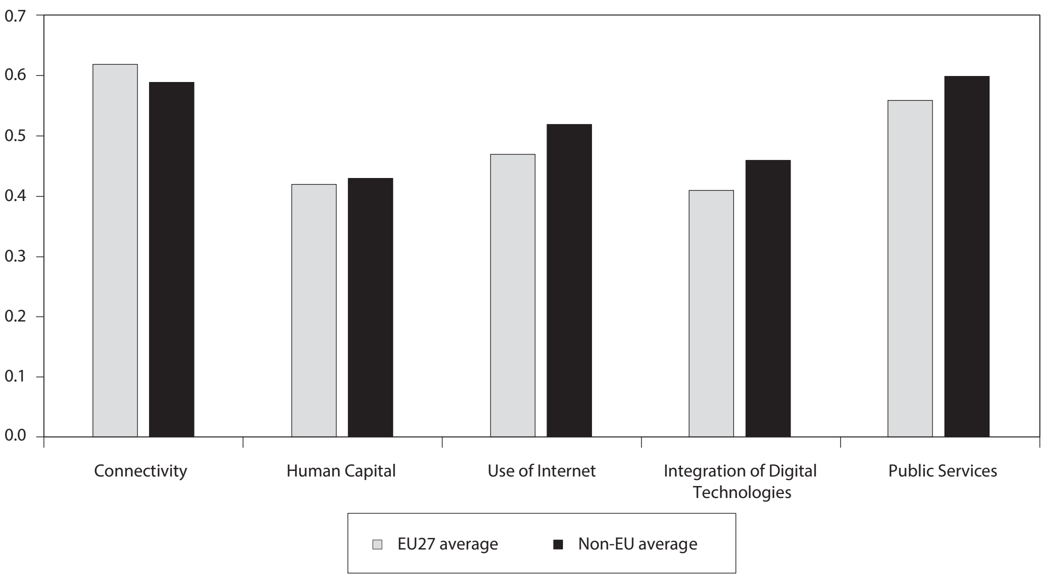
To assess multivariate normality, Rencher (2002) suggests checking each variable separately. If normality can be assumed for all variables, then the multivariate distribution is also considered normal. Univariate normality testing is usually done by graphical analysis as it is a fast and reliable method to diagnose abnormalities. Based on QQ plots, we have no reason to assume a deviation from normality. The Kolmogorov-Smirnov test was not significant for any of the variables, and the Shapiro-Wilk test showed a weakly significant result for the Integration of Digital Technologies at 0.05. This test is for small

Table 2 Tests for normality

	Kolmogorov-Smirnova			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Connectivity	0.077	45	0.200*	0.973	45	0.366
Human capital	0.093	45	0.200*	0.970	45	0.291
Citizen internet use	0.078	45	0.200*	0.975	45	0.436
Integration of digital technologies	0.120	45	0.105	0.945	45	0.034
Public services	0.080	45	0.200*	0.979	45	0.584

Notes: * this is a lower bound of true significance. a Lilliefors significance correction.
Source: Calculation on the European Commission data (EC, 2021b)

Figure 2 Differences between the means



Source: Own design based on the data of the National Bank of Hungary sample sizes and has higher statistical power than the Kolmogorov-Smirnov test. For this reason, normality is assumed for the univariate and multivariate cases. Test results are shown in Table 2.

Homogeneity of variance was examined by comparing the standard deviations of the variables and using the Box test. The standard deviation of the five variables is similar in the two groups, suggesting homogeneity of variance. The most significant difference between the groups was in Internet access, where the variance of the EU group is half (0.05101) of the variance of the non-EU group (0.10427). The results of the Box test to verify multivariate variance homogeneity were not significant, so we assumed equality of covariance matrices. With this, we checked the conditions of the multivariate analysis of variance, the results can be considered authentic. The averages are illustrated in Figure 2.

The value of Connectivity is generally higher in the EU Member States, while the Use of Internet services, the Integration of Digital Technologies and Public Services are higher in non-EU countries. There is no significant difference in Human Capital between the two groups. In MANOVA, the null hypothesis is tested that the values of the five dimensions, measured in the EU and in non-EU countries, can be derived from the same population using F tests. None of the five dimensions were significant at the 0.05 level, thus we have no reason to assume that there is a significant difference between the two groups. The maximum value of $\eta 2$ for internet services was 0.031, so this dimension alone only explains 3.1% of the difference between the EU and non-EU countries. Therefore, the data set can be examined together.

3.3 Relation of the index's dimensions (correlation test)

After proving no difference between EU and non-EU data, the correlation analysis was performed on the entire data set. The result is shown in Table 3. The correlation between the variables is positive in all cases, so they move together. This is also apparent from the context of the data, as all five dimensions are meant to represent an aspect of digital development (European Commission, 2021c). In nine of the ten cases examined, the correlation between the Integration of Digital Technologies and Public Services is significant at only a 0.05 level. In the case of the 2019 data set (Bánhidi et al., 2020), the correlation was significant at a 0.01 level in all cases. Examining the data of the EU member states only, the p-value is 0.467, which is also not significant at the level of 0.01. Looking only at data from non-EU countries, Public Services do not show a significant correlation with any of the other DESI dimensions, suggesting that the Public Services is less related to the other four indicators.

The multicollinearity between the variables were also examined. The variance inflation factors (VIF) of all dimensions have a value of less than 5, the variables are not multicollinear, and the linear relationships

Table 3 Matrix of the Pearson correlation coefficients

Dimensions	Human capital	Use of internet	Integration of digital technologies	Public services
Connectivity	0.626** 0.000	0.641** 0.000	0.656** 0.000	0.433** 0.003
Human capital		0.705** 0.000	0.730** 0.000	0.602** 0.000
Use of internet			0.823** 0.000	0.444** 0.002
Integration of digital technologies				0.355* 0.017

Notes: ** correlation is significant at the 0.01 level (two-way). * correlation is significant at the 0.05 level (two-way).
Source: Based on calculation of Tarjáni et al. (2022)

between them are not significant. For the sake of brevity, we will not go into details. The application of the VIF method is described in Vörösmarty and Dobos (2020).

3.4 Reduction of dimensions as model variables (principal component analysis)

Due to strong correlations between the I-DESI dimensions, we performed principal component analysis to simplify the model and explore latent variables. The Kaiser-Meyer-Olkin (KMO) score was 0.804, indicating the adequacy of the sampling and suggesting the existence of latent variables. The result of the Bartlett test was significant, which proves the relationship between the variables, which makes the data set suitable for principal component analysis. The communalities ranged from 0.6 to 0.9, thus, the principal components explain most of the variance. According to our calculations without rotation, the first factor explains 68.8 percent of the variance. Factor weights were above 0.810 with four variables indicating strong correlation, but with Public Services, the value of the factor weight was only 0.653. Two principal components explained 83.4 percent of the variance. With the second factor, Public Services showed the highest correlation with a value of 0.737.

To confirm the results, we used the Varimax rotation method, which explained 83.4 percent of the variance in the same way for the two main components but resulted in different factor weights. The factor weights obtained after rotation are summarized in Table 4. The rotation converged after three iterations.

Table 4 Factor weights after rotation

Dimensions	Factors	
	Digital capability	Digital applications
Integration of digital technologies	0.932	0.136
Use of internet	0.881	0.240
Connectivity	0.767	0.295
Human capital	0.718	0.539
Public services	0.210	0.962

Source: Based on calculation of Tarjáni et al. (2022)

The first main component is strongly correlated with the Integration of Digital Technologies, the Use of Internet, and Connectivity, the impact of Human Capital is moderate, and Public Services have a weak correlation with this component. In the analysis of the previous year's data (Bánhidi et al., 2020), the Integration of Digital Technologies was more correlated with the second main component. In our opinion, the change shows the impact of the Covid-19 pandemic on digitalization, as the digital applications of businesses have begun to mix with the characteristics of residential use. Assuming the transient nature of the effect, the previous names of the components were retained.

3.5 Relationships between dimensions as model variables (partial correlation analysis)

We examined the causal relationships between I-DESI dimensions by partial correlation analysis. Of the ten coefficients, six were found to be significant at the less strict 0.15 significance level, which is listed in Table 5. The values of the significant coefficients ranged from 0.2 to 0.6, thus, the system is characterized by weak and moderate causal relationships.

Based on the partial correlation coefficients, we set up the causal chain between the variables shown in Figure 3.

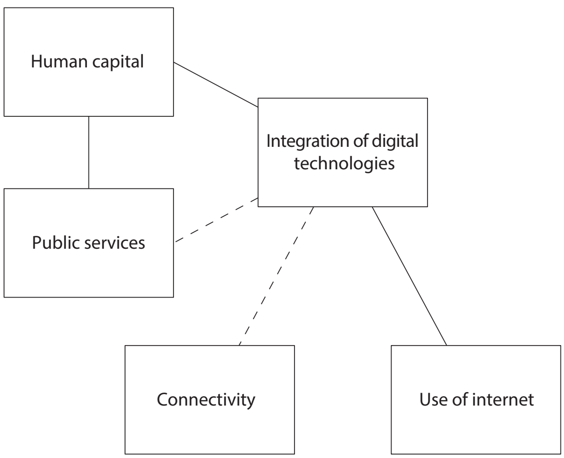
Compared to the figure based on the previous year's data (Bánhidi et al., 2020), it is apparent that Connectivity can still be considered an independent variable but is no longer related to the Use of Internet,

Table 5 Partial correlation coefficients matrix

Dimensions	Human capital	Use of internet	Integration of digital technologies	Public services
Connectivity	0.143 0.365	0.147 0.353	0.228 0.147*	0.136 0.390
Human capital		0.119 0.425	0.380 0.013*	0.469 0.002**
Use of internet			0.519 0.000**	0.150 0.342
Integration of digital technologies				– 0.252 0.107*

Notes: ** correlation is significant at the 0.01 level (two-way). * correlation is significant at the 0.15 level (two-way).
Source: Based on calculation of Tarjáni et al. (2022)

Figure 3 Causal relations between the variables



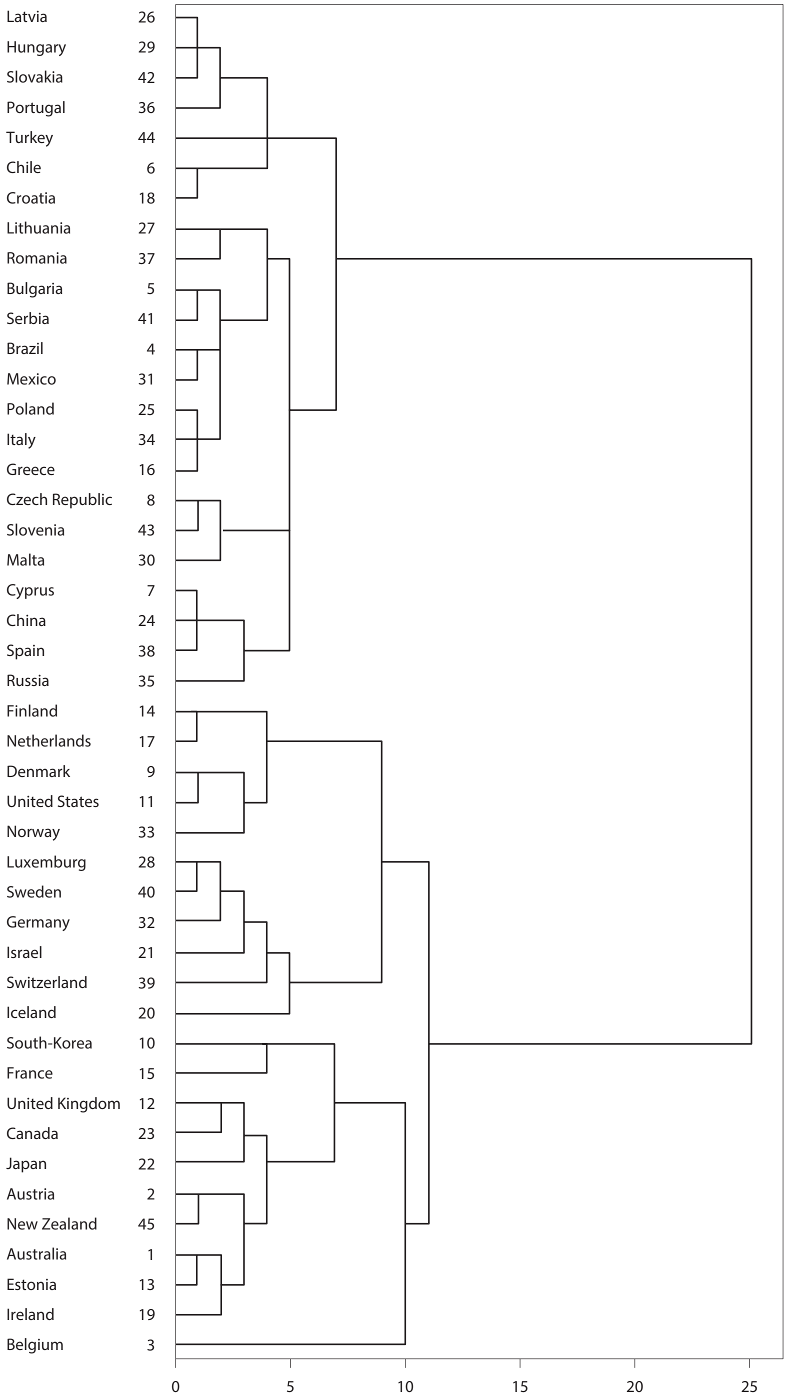
Source: Based on calculation of Tarjáni et al. (2022)

but it is directly related to the Integration of Digital Technologies. In the case of Human Capital, it is no longer a clearly independent variable and is no longer linked to the Use of Internet. However, the Use of Internet is also linked to the Integration of Digital Technologies based on the 2020 data set, as are the other three variables. Data from 2020 show that the Integration of Digital Technologies has shifted closer to the private sector than in the previous year, presumably as a result of the COVID-19 pandemic. The Public Services dimension is no longer strongly linked neither to the Integration of Digital Technologies nor to Human Capital.

3.6 Categorization of the data set (cluster analysis)

For grouping the surveyed countries, we used hierarchical cluster analysis with a square Euclidean distance measure and between-groups linakge, where the group formation was based on the relationship between the groups. The results are on a dendrogram in Figure 4. Based on the results, the countries participating in the survey can be divided into two major groups, the 22 countries that perform better in digitalization

Figure 4 Dendrogram of the cluster analysis



Source: Calculation on the European Commission data (EC, 2021b)

(EU member states: Finland, the Netherlands, Denmark, Luxembourg, Sweden, Germany, France, Austria, Estonia, Ireland, Belgium), and the 23 less digitalized countries (EU Member States: Latvia, Hungary, Slovakia, Portugal, Croatia, Lithuania, Romania, Bulgaria, Poland, Italy, Greece, the Czech Republic, Slovenia, Malta, Cyprus, Spain). For three clusters, the 11 best-performing countries stand out from the more advanced block, including the northern EU member states (Finland, the Netherlands, Denmark, Luxembourg, Sweden, and Germany) and the United States, Norway, Israel, Switzerland and Iceland.

Cluster analysis was also performed with fixed cluster numbers, and the results of this study are shown in Table 6. Dividing the countries into three clusters, the 11 best-performing countries are separated from the other 11 countries in the middle field and the 23 least less digitalized countries. By defining four clusters, only Belgium's ranking changes, sticking out of the midfield. For five clusters, the 11 best-performing countries split into the top five (Denmark, the United States, Finland, the Netherlands, Norway) and the six better-performing countries (Iceland, Israel, Luxembourg, Germany, Switzerland, Sweden).

Table 6 Result of the cluster analysis for three, four and five clusters

Countries	3 clusters	4 clusters	5 clusters
Australia	1	1	1
Austria	1	1	1
South Korea	1	1	1
United Kingdom	1	1	1
Estonia	1	1	1
France	1	1	1
Ireland	1	1	1
Japan	1	1	1
Canada	1	1	1
New Zealand	1	1	1
Belgium	1	2	2
Brazil	2	3	3
Bulgaria	2	3	3
Chile	2	3	3
Cyprus	2	3	3
Czech Republic	2	3	3
Greece	2	3	3
Croatia	2	3	3
China	2	3	3
Poland	2	3	3
Latvia	2	3	3
Lithuania	2	3	3
Hungary	2	3	3
Malta	2	3	3
Mexico	2	3	3
Italy	2	3	3

Table 6 (continuation)

Countries	3 clusters	4 clusters	5 clusters
Russia	2	3	3
Portugal	2	3	3
Romania	2	3	3
Spain	2	3	3
Serbia	2	3	3
Slovakia	2	3	3
Slovenia	2	3	3
Turkey	2	3	3
Denmark	3	4	4
United States	3	4	4
Finland	3	4	4
Netherlands	3	4	4
Norway	3	4	4
Iceland	3	4	5
Israel	3	4	5
Luxembourg	3	4	5
Germany	3	4	5
Switzerland	3	4	5
Sweden	3	4	5

Source: Calculation on the European Commission data (EC, 2021b)

Some discrepancies were noted in Figure 4. For example, China ended up in the same group as Spain, which we found interesting. To ensure the results further, we performed a k-means cluster analysis using the five clusters result as an initial. The number of replications was 10, and the Minkowski distance parameter was 2. The results are shown in Table 7.

Table 7 Result of the k-means cluster analysis for five clusters using the five clusters' initial

Countries	5 clusters initial	k-means clusters
Australia	1	1
Austria	1	1
Canada	1	1
Estonia	1	1
France	1	1
Ireland	1	1
Japan	1	1
New Zealand	1	1
South Korea	1	1
Belgium	2	2

Table 7 (continuation)

Countries	5 clusters initial	k-means clusters
Chile	3	2
Croatia	3	2
Czech Republic	3	2
Hungary	3	2
Latvia	3	2
Lithuania	3	2
Portugal	3	2
Slovakia	3	2
Slovenia	3	2
Brazil	3	3
Bulgaria	3	3
China	3	3
Cyprus	3	3
Greece	3	3
Italy	3	3
Malta	3	3
Mexico	3	3
Poland	3	3
Romania	3	3
Russia	3	3
Serbia	3	3
Spain	3	3
Turkey	3	3
Denmark	4	4
Finland	4	4
Netherlands	4	4
Norway	4	4
United States	4	4
Germany	5	5
Iceland	5	5
Israel	5	5
Luxembourg	5	5
Sweden	5	5
Switzerland	5	5
United Kingdom	1	5

Source: Calculation on the European Commission data (EC, 2021b)

China maintained its place with Spain in this analysis, and mainly the second cluster expanded with countries from the third, not leaving Belgium alone anymore. The United Kingdom also changed its place and joined the fifth cluster the other members of the Commonwealth countries back in cluster one.

CONCLUSION

In our study, we examined the five dimensions of the I-DESI index using multivariate statistical methods. We found answers to our research questions:

Q1. *We found that it is possible to separate the dataset into groups of EU and non-EU countries.* Based on the graphical analysis of the data, we found that Connectivity is typically high in the EU member states, while the values of Public Services are generally higher in the non-EU countries.

Q2. *We found that there is no difference between EU and non-EU averages.* Based on the discriminant analysis, we found that the data series for EU member states and non-EU countries did not differ from each other, and when comparing the averages, there was no significant difference between the two groups, thus, we analyzed the two data sets together.

Q3. *We found some linear relationships between the dimensions of the indicator system.* In the correlation analysis, we found that there are strong correlations between the variables, the strongest is between the Use of Internet and the Integration of Digital Technologies, which makes the data suitable for principal component analysis.

Q4. *We found that the number of model variables cannot be reduced.* In the principal component analysis, two principal components were separated. The Integration of Digital Technologies, the Use of Internet, and Connectivity are strongly correlated with the first principal component, and the impact of Public Services is of marginal importance in the case of the second factor. Human Capital is almost equally involved in both factors.

Q5. *We found some relationships between the model variables.* Based on the partial correlation analysis, we set up the causal chain between the variables and found that the Integration of Digital Technologies has shifted closer to the factors of the private sector compared to the previous year's data, presumably due to the COVID-19 pandemic.

Q6. *We could categorize the analyzed countries.* During the cluster analysis, we formed groups from the participating countries. In the case of three clusters, the 11 best-performing countries are separated from 11 of the mid-range countries and the 23 least digitalized countries. For four clusters, Belgium stood out in the middle, and for five clusters, the best-performing countries split from the leading and high-performing countries. The k-means cluster analysis reinforced these results, expanding the fourth cluster and specifying the place of the United Kingdom between clusters one and five.

These results form an important basis for further research related to digitalization. Not only the regional differences and the changes over time are visible, but we can see that EU states are great areas for digitized operations and related research too. To examine the aftermath of the COVID-19 epidemic and to assess the temporary approach of the private sector to the digitization of businesses, it is worth carrying out similar analyses in the upcoming years. In this way, long-term conclusions can be drawn about the relationship between pandemics and digitalization.

References

BÁNHIDI, Z., DOBOS, I., NEMESLAKI, A. (2020). What the overall Digital Economy and Society Index reveals: a statistical analysis of the DESI EU28 dimensions [online]. *Regional Statistics*, 10(2): 42–62. <https://doi.org/10.15196/RS100209>.

BÁNHIDI, Z., DOBOS, I., TOKMERGENOVA, M. (2021). Russia's Place Vis-à-Vis the EU28 Countries in Digital Development: a Ranking Using DEA-Type Composite Indicators and the TOPSIS Method [online]. In: HERBERGER, T. A., DÖTSCH, J. J. (eds.) *Digit. Digit. Transform. Sustain. Glob. Econ.*, Springer, 135–146. <https://doi.org/10.1007/978-3-030-77340-3_11>.

BAŞOL, O., YALÇIN, E. C. (2021). How Does the Digital Economy and Society Index (DESI) Affect Labor Market Indicators in EU Countries? [online]. *Human Systems Management*, 40(4): 503–512. <https://doi.org/10.3233/HSM-200904>.

BOUSDEKIS, A., KARDARAS, D. (2020) Digital Transformation of Local Government: a Case Study from Greece [online]. In: *2020 IEEE 22nd Conf. Bus. Informatics*, 131–140. <https://doi.org/10.1109/CBI49978.2020.10070>.

BURTON-JONES, A., AKHLAGHPOUR, S., AYRE, S., BARDE, P., STAIB, A., SULLIVAN, C. (2020). Changing the conversation on evaluating digital transformation in healthcare: Insights from an institutional analysis [online]. *Inf. Organ.* 30: 100255. <https://doi.org/10.1016/j.infoandorg.2019.100255>.

CUESTA, C., RUESTA, M., TUESTA, D., URBIOLA, P. (2015). The digital transformation of the banking industry [online]. *BBVA Res. Digit. Watch.* <https://www.researchgate.net/profile/David-Tuesta/publication/291357544_The_digital_transformation_of_the_banking_industry/links/56a2cc6f08ae1b65112cbdb9/The-digital-transformation-of-the-banking-industry.pdf>.

ERJAVEC, J., MANFREDA, A., JAKLIČ, J., STEMBERGER, M., FEHÉR, P., SZABÓ, Z., KŐ, A. (2018). Case Studies of Successful Digital Transformation in Slovenia and Hungary [online]. *5th International Conference on Management and Organization, Brdo pri Kranju, Szlovénia: Slovenian Academy of Management,* 219–233. <https://www.researchgate.net/profile/Peter-Feher-2/publication/326010451_Case_Studies_of_Successful_Digital_Transformation_in_Slovenia_And_Hungary/links/5b335ccba6fdcc8506d17551/Case-Studies-of-Successful-Digital-Transformation-in-Slovenia-And-Hungary.pdf>.

ESSES, D., CSETE, M. S., NÉMETH, B. (2021). Sustainability and Digital Transformation in the Visegrad Group of Central European Countries [online]. *Sustainability*, 13(11): 5833. <https://doi.org/10.3390/su13115833>.

EUROPEAN COMMISSION. (2021a). I-DESI 2020: *How digital is Europe compared to other major world economies?* [online]. <https://digital-strategy.ec.europa.eu/en/library/i-desi-2020-how-digital-europe-compared-other-major-world-economies>.

EUROPEAN COMMISSION. (2021b). *Raw data: International Digital Economy and Society Index 2020* [online]. <https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=72353>.

EUROPEAN COMMISSION. (2021c). *Study: International Digital Economy and Society Index 2020* [online]. <https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=72352>.

GURĂU, M. I. (2021). The Impact of Social Media in the Digitization Process [online]. *Global Economic Observer,* 9(1): 139–146. <http://www.globeco.ro/wp-content/uploads/vol/split/vol_9_no_1/geo_2021_vol9_no1_art_016.pdf>.

HA, L. T. (2022). Effects of digitalization on financialization: Empirical evidence from European countries [online]. *Technology in Society*, 68: 101851. <https://doi.org/101851. 10.1016/j.techsoc.2021.101851>.

KOKH, L. V., KOKH, Y. V. (2019). Analysis of existing approaches to measurement of digital economy [online]. *St. Petersburg State Polytechnical University Journal. Economics,* 12(4): 78–89. <https://doi.org/10.18721/JE.12407>.

MIRZAEI ABBASABADI, H., SOLEIMANI, M. (2021) Examining the effects of digital technology expansion on Unemployment: a cross-sectional investigation [online]. *Technol. Soc.*, 64: 101495. <https://doi.org/10.1016/j.techsoc.2020.101495>.

MORENO-LLAMAS, A., GARCÍA-MAYOR, J., DE LA CRUZ-SÁNCHEZ, E. (2020). The impact of digital technology development on sitting time across Europe [online]. *Technol. Soc.*, 63: 101406. <https://doi.org/10.1016/j.techsoc.2020.101406>.

NISSEN, V., LEZINA, T., SALTAN, A. (2018). The Role of IT-Management in the Digital Transformation of Russian Companies [online]. *Foresight STI Gov.*, 12(3): 53–61. <https://doi.org/10.17323/2500-2597.2018.3.53.61>.

POLOZOVA, T., KOLUPAIEVA, I., SHEIKO, I. (2021). Digital Gap in EU Countries and its Impact on Labour Productivity and Global Competitiveness [online]. *Proceedings of the international scientific conference Hradec Economic Days 2021*, University of Hradec Králové. <https://doi.org/10.36689/uhk/hed/2021-01-065>.

RENCHER, A. C. (2002). *Methods of Multivariate Analysis.* Danvers, MA: Wiley-Interscience.

RUSSO, V. (2020). Digital Economy and Society Index (DESI). European Guidelines and Empirical Applications on the Territory [online]. In: SARASOLA SÁNCHEZ-SERRANO, J., MATURO, F., HOŠKOVÁ-MAYEROVÁ, Š. (eds.) Qualitative and *Quantitative Models in Socio-Economic Systems and Social Work. Studies in Systems, Decision and Control*, Springer, 427–442. <https://doi.org/10.1007/978-3-030-18593-0_31>.

SCUPOLA, A. (2018). Digital transformation of public administration services in Denmark: a process tracing case study [online]. *Nord. Balt. J. Inf. Commun. Technol.*, 1: 261–284. <https://doi.org/10.13052/nbjict1902-097X.2018.014>.

STAVYTSKYY, A., KHARLAMOVA, G., STOICA, E. A. (2019). The Analyses of the Digital Economy and Society Index in the EU [online]. *Baltic Journal of European Studies Tallinn University of Technology*, 9(3): 245–261. <https://doi.org/10.1515/bjes-2019-0032>.

TARJÁNI, A. J., KALLÓ, N., DOBOS, I. (2022). Statistical analysis of 2020 data for the international digital economy and society index (in Hungarian) [online]. *Statisztikai Szemle*, 100(3): 266–284. <https://doi.org/10.20311/stat2022.3.hu0266>.

TARJÁNI, A. J., KALLÓ, N., PATAKI, B. (2021). Introducing a Task Management Tool into the Operation of a Management Consulting Firm [online]. *Int. J. Eng. Manag. Sci.*, 6(2): 122–135. <https://doi.org/10.21791/IJEMS.2021.2.11>.

TEIXEIRA, A. F., GONÇALVES, M. J., TAYLOR, M. D. (2021). How Higher Education Institutions Are Driving to Digital Transformation: a Case Study [online]. *Educ. Sci.*, 11(10): 636. <https://doi.org/10.3390/educsci11100636>.
VÖRÖSMARTY, G., DOBOS, I. (2020). Green purchasing frameworks considering firm size: a multicollinearity analysis using variance inflation factor [online]. *Supply Chain Forum: an International Journal*, 21(4): 290–301. <https://doi.org /10.1080/16258312.2020.1776090.>

APPENDIX

Table A1 Basic data of our study

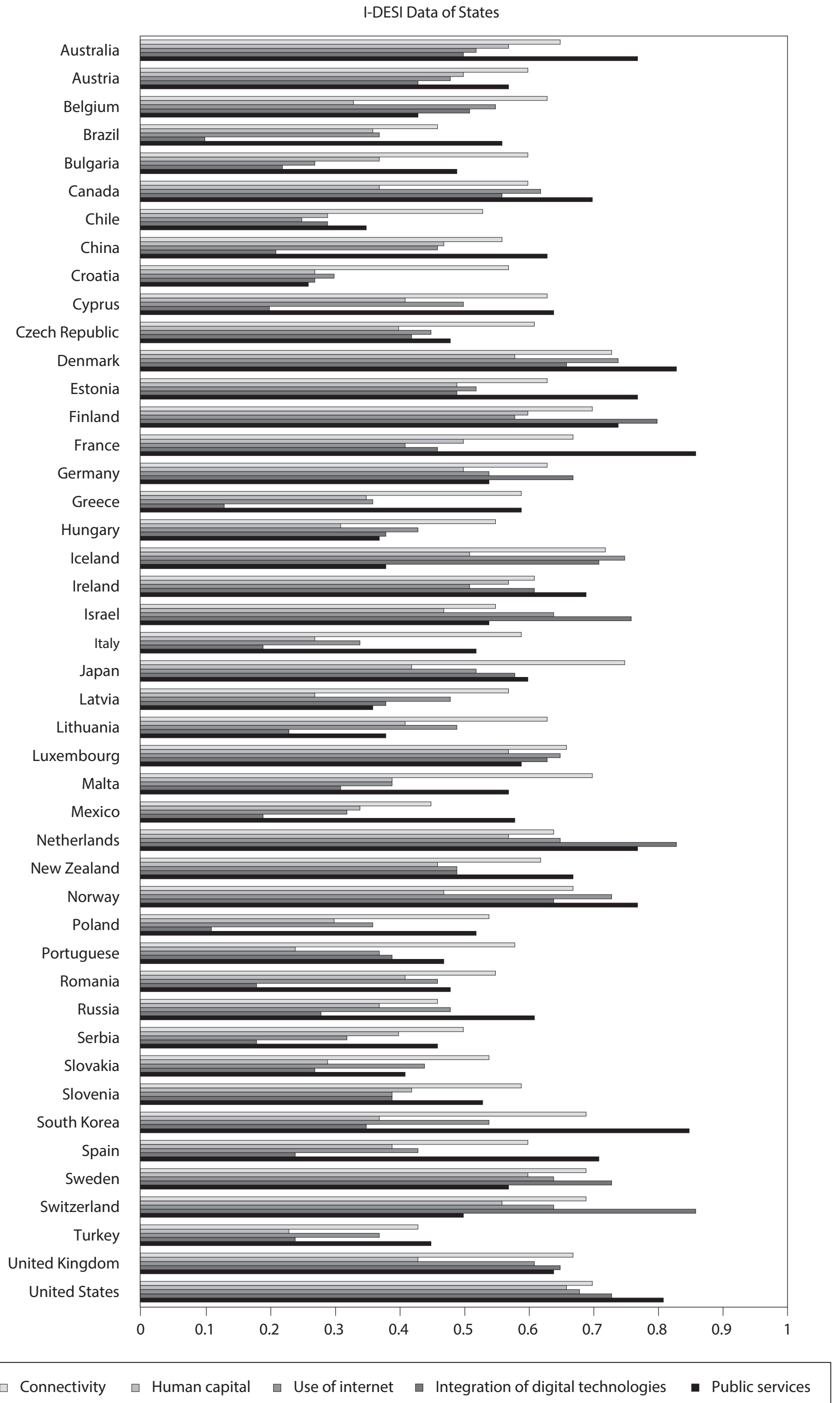
Countries	Dimensions					
	Connectivity	Human capital	Use of internet	Integration of digital technologies	Public services	I-DESI
Weights	0.25	0.25	0.15	0.20	0.15	
EU27 average	0.62	0.42	0.47	0.41	0.56	0.50
Austria	0.60	0.5	0.48	0.43	0.57	0.52
Belgium	0.63	0.33	0.55	0.51	0.43	0.49
Bulgaria	0.60	0.37	0.27	0.22	0.49	0.40
Croatia	0.57	0.27	0.30	0.27	0.26	0.35
Cyprus	0.63	0.41	0.5	0.2	0.64	0.47
Czech Republic	0.61	0.40	0.45	0.42	0.48	0.47
Denmark	0.73	0.58	0.74	0.66	0.83	0.7
Estonia	0.63	0.49	0.52	0.49	0.77	0.57
Finland	0.70	0.60	0.58	0.80	0.74	0.68
France	0.67	0.50	0.41	0.46	0.86	0.57
Germany	0.63	0.50	0.54	0.67	0.54	0.58
Greece	0.59	0.35	0.36	0.13	0.59	0.40
Hungary	0.55	0.31	0.43	0.38	0.37	0.41
Ireland	0.61	0.57	0.51	0.61	0.69	0.60
Italy	0.59	0.27	0.34	0.19	0.52	0.38
Latvia	0.57	0.27	0.48	0.38	0.36	0.41
Lithuania	0.63	0.41	0.49	0.23	0.38	0.44
Luxembourg	0.66	0.57	0.65	0.63	0.59	0.62
Malta	0.7	0.39	0.39	0.31	0.57	0.48
Netherlands	0.64	0.57	0.65	0.83	0.77	0.68
Poland	0.54	0.30	0.36	0.11	0.52	0.36
Portuguese	0.58	0.24	0.37	0.39	0.47	0.41
Romania	0.55	0.41	0.46	0.18	0.48	0.42
Slovakia	0.54	0.29	0.44	0.27	0.41	0.39
Slovenia	0.59	0.42	0.39	0.39	0.53	0.47
Spain	0.60	0.39	0.43	0.24	0.71	0.47

Table A1 (continuation)

Countries	Dimensions					
	Connectivity	Human capital	Use of internet	Integration of digital technologies	Public services	I-DESI
Sweden	0.69	0.60	0.64	0.73	0.57	0.65
Non-EU average	0.59	0.43	0.52	0.46	0.60	0.52
Australia	0.65	0.57	0.52	0.5	0.77	0.60
Brazil	0.46	0.36	0.37	0.1	0.56	0.37
Canada	0.60	0.37	0.62	0.56	0.70	0.55
Chile	0.53	0.29	0.25	0.29	0.35	0.35
China	0.56	0.47	0.46	0.21	0.63	0.46
Iceland	0.72	0.51	0.75	0.71	0.38	0.62
Israel	0.55	0.47	0.64	0.76	0.54	0.58
Japan	0.75	0.42	0.52	0.58	0.60	0.57
Mexico	0.45	0.34	0.32	0.19	0.58	0.37
New Zealand	0.62	0.46	0.49	0.49	0.67	0.54
Norway	0.67	0.47	0.73	0.64	0.77	0.64
Russia	0.46	0.37	0.48	0.28	0.61	0.43
Serbia	0.50	0.40	0.32	0.18	0.46	0.38
South Korea	0.69	0.37	0.54	0.35	0.85	0.54
Switzerland	0.69	0.56	0.64	0.86	0.50	0.66
Turkey	0.43	0.23	0.37	0.24	0.45	0.34
United Kingdom	0.67	0.43	0.61	0.65	0.64	0.59
United States	0.70	0.66	0.68	0.73	0.81	0.71

Source: Calculation on the European Commission data (EC, 2021b)

Figure A1 I-DESI data of the examined states



Source: Calculation on the European Commission data (EC, 2021b)