BANCA D'ITALIA
EUROSISTEMA

Questioni di Economia e Finanza

(Occasional Papers)

Digitalisation in Italy: evidence from a new regional index

by Andrea Benecchi, Carlo Bottoni, Emanuela Ciapanna, Annalisa Frigo, Aldo Milan and Elisa Scarinzi

December 2021

Number 662

BANCA D'ITALIA
EUROSISTEMA

Questioni di Economia e Finanza

(Occasional Papers)

Digitalisation in Italy: evidence from a new regional index

by Andrea Benecchi, Carlo Bottoni, Emanuela Ciapanna, Annalisa Frigo, Aldo Milan and Elisa Scarinzi

Number 662 – December 2021

The series Occasional Papers *presents studies and documents on issues pertaining to the institutional tasks of the Bank of Italy and the Eurosystem. The* Occasional Papers *appear alongside the* Working Papers *series which are specifically aimed at providing original contributions to economic research.*

The Occasional Papers *include studies conducted within the Bank of Italy, sometimes in cooperation with the Eurosystem or other institutions. The views expressed in the studies are those of the authors and do not involve the responsibility of the institutions to which they belong.*

The series is available online at www.bancaditalia.it .

ISSN 1972-6627 (print)
ISSN 1972-6643 (online)

Printed by the Printing and Publishing Division of the Bank of Italy

DIGITALISATION IN ITALY:
EVIDENCE FROM A NEW REGIONAL INDEX

by Andrea Benecchi†, Carlo Bottoni‡, Emanuela Ciapanna§,
Annalisa Frigo‡, Aldo Milan**, Elisa Scarinzi*

Abstract

The geographic digital divide has a significant, though largely unexplored, dimension within a country. This paper proposes an index of digital development for the Italian NUTS2 regions (rDESI) based on the European Commission's Digital Economy and Society Index (DESI). The rDESI monitors the regional digital divide across five dimensions: (i) the infrastructure and the network usage (connectivity), (ii) the population's digital skills, (iii) the use of internet services by households, (iv) the integration of ICT by firms, and (v) the level of digital services offered by local government. Southern regions tend to lag behind in most of these dimensions, even if infrastructures and the quality of connectivity appears quite homogeneous across the country. In the last part of the paper, we highlight the limitations of the DESI methodology, proposing some improvements.

JEL Classification: C43, C80, L96, R10.
Keywords: digitalisation, connectivity; DESI, regional divergence.
DOI: 10.32057/0.QEF.2021.0662

Contents

1. Introduction 5
2. Building a regional index of digital development 9
   2.1 DESI 1 Connectivity 11
   2.2 DESI 2 Human Capital 12
   2.3 DESI 3 Use of Internet services 13
   2.4 DESI 4 Integration of digital technology 14
   2.5 DESI 5 E-government 15
3. Analysis of the rDESI 18
4. Statistical properties of the rDESI 21
5. Methodological gaps and way forward 25
6. Conclusions 29
References 30
Appendices 32

§ Bank of Italy, Economic Research and International Relations. † Bank of Italy, Trieste branch. ‡ Bank of Italy, Venezia branch. * Bank of Italy, Roma branch. ** Agcom, Autorità per le garanzie nelle comunicazioni.

Introduction[1]

Given the well-established transforming power of digital technologies for advanced economies, addressing deficits in the digital domain is at the heart of many current policy interventions.

At the European level, the Next Generation EU aims at supporting European countries towards the digital transition, among other goals. As part of the next long-term EU budget (i.e. the Multi-annual Financial Framework 2021-2027), the Digital Europe Programme will provide strategic funding to projects in five key capacity areas: super-computing, artificial intelligence, cybersecurity, advanced digital skills, and ensuring a wide use of digital technologies across the economy and society, including through Digital Innovation Hubs. With a planned overall budget of 7.5 billion euros,[2] the programme aims to shape the digital transformation of Europe, benefiting the entire economy and in particular small to medium enterprises (SMEs).

At the national level, the Italian Government has allocated 27% of the Recovery and Resilience Facility budget to the digital transition.[3]

Developing an accurate mapping and monitoring tool for specific regions is beneficial to the design of public policies that promote investment in ultrafast broadband infrastructures, favour the accumulation of digital human capital, and foster ICT adoption by firms.

Since 2015, the European Commission monitors the level of digitalisation of the economies and societies of the Member States by computing the Digital Economy and Society Index (DESI). This composite indicator summarises relevant aspects of Europe's digital performance and tracks the evolution of EU member states across five main dimensions: connectivity, human capital, use of the internet, integration of digital technologies, digital public services.

According to the DESI indicator, in 2020 Italy ranked 25th out of 28 countries, with only Romania, Greece and Bulgaria further behind in the digital development. The poorest performance is recorded in the human capital sub-index, where our country ranks last, followed by internet usage, where it has steadily occupied the 26th position for the last five years. Italy's ranking slightly improves when looking at the other three sub-indices: in particular the figures related to digital public services and connectivity are close to the EU average, having shown significant progress throughout the last years.

[1]We thank for useful comments Federico Cingano, Paolo Chiades, Giuseppe Albanese and participants at the Bank of Italy seminar. The views expressed here are our own and do not necessarily reflect those of the Bank of Italy.

[2]The Digital Europe Programme will complement the funding available through other EU programmes, such as the Horizon Europe programme for research and innovation, the Connecting Europe Facility for digital infrastructure, the Recovery and Resilience Facility, and the Structural funds.

[3]See the detailed allocation in the PNRR document.

Besides the general delay in digitalisation, which involves the whole country, heterogeneity across Italian territories represents a crucial factor, informing the Government's future action within the Italian digital agenda. Along this line, the prompt availability of disaggregated data, providing a mapping of the digital development within regions represents a fundamental tool to establish priorities of intervention and to efficiently guide investment plans. The aim of this paper is to propose a NUTS2-level index of digital development, called *rDESI*, closely following the methodology of the 2020 DESI.

Our synthetic measure is built over a wide range of data and allows to identify the gaps across Italian regions in the same five fields included in the national composite indicator.

The scope of our study is twofold: first, we aim at providing a methodological description of the composite regional DESI by presenting the data sources, the technical characteristics and shedding light on strengths and flaws of our metrics; second, we put our index at work, show our descriptive results and the consequent ranking for the Italian regions.

According to our study, the sub-indices are highly correlated, the strongest relationship being retrieved between the components referred to human capital and to the use of internet services, confirming the low performance on both fronts at the country level. High correlation is also retrieved between internet usage and both integration of ICT by firms, and the level of digital services offered by local public administrations, mainly reflecting the dominance of the demand-side sub-indices within each composite measure.

As of geographical heterogeneity, we find a significant degree of cross-regions variability in all the sub-indices, with Lombardy, Emilia Romagna and Lazio outperforming the rest of the country and Northern and Central-Western regions showing on average better figures than Southern and Central-Eastern ones. Similarly to the 2020 DESI, the data we rely on to construct the regional indicator refer to 2019. Clearly, in 2020 many changes have occurred, in response to the extraordinary demand shock of digital applications that the Covid-19 pandemics and the associated Governments' prevention measures have brought about. These structural changes, together with the acceleration towards a digital transition impressed by the National Recovery and Resilience Plan, are not reflected in the current rDESI yet, but they have such a scale that we expect a strong boost in all its sub-components as well as in the country-level DESI for years to come.

We conclude our analysis by proposing some amendments to the European Commission methodology, in order to better capture key aspects of digitalisation and to address some of the current data limitations. These proposals are in line with the ongoing revision of the general index by a Working Group of the European Commission, aiming at constructing an *enhanced* DESI. This should include further Key Performance Indicators and represent a more advanced and reliable monitoring tool for the Digital Decade.

The topic considered in this article relates to and extends on two strands of the literature. First, it pertains to the field studying the effects of the digital divide. The latter refers to the gap between those able to benefit from digital technologies and those unable, and it is often measured in terms of availability of broadband infrastructure (i.e. spatially) or in terms of digital literacy (e.g. across genders or cohorts).[4] There is growing evidence that the lack of access to the internet can have consistent repercussions in the education attainments, social capital, and economic domain.[5]

In the European context, Szeles (2018) and Szeles and Simionescu (2020) analyse regional- and country level determinants of the regional digital divide in the EU. According to their findings, a mix of regional and national measures (e.g. increasing the tertiary education attainments, boosting R&D expenditure, and discouraging early leaving from education) could successfully reduce the regional digital divide in the EU. García *et al.* (2012) finds that regional governments' policies on broadband expansion have partly bridged the digital divide within the EU; however, regional level interventions are crucial for the improvement of broadband access. With respect to the Italian broadband sector, Nucciarelli *et al.* (2013) examined three regional initiatives and concluded that the major threats to local broadband initiatives may come from the projects' wide geographical extension (leading to a misalignment between public interests and private business opportunities) and from having set up weak incentives to private investments.[6]

Though our paper is close to the aforementioned literature insofar as regional heterogeneities are at stake, our current focus is mostly on the methodology to measure the digital divide, as we first introduce an indicator measuring the degree of digitalisation at the local level, and then employ the latter to provide insights on the Italian geographical distribution of the different components. In this sense, it belongs to the stream of literature attempting to construct digitalisation metrics in the wake of the European Commission's work.

In fact, the interest raised by the DESI brought about several reports investigating national-level performance more in detail, and commenting specific aspects of its methodology and results (see e.g. Bánhidi *et al.* (2020) for a recent reviewed of this literature).[7]

[4]See e.g. Pereira (2016) for a more formal definition of digital divide.

[5]See, for instance,Guriev *et al.* (forthcoming), Schaub and Morisi (2020), and Campante *et al.* (2018). Similar results are obtained by Mammadli and Klivak (2020).

[6]It is not just about broadband investments. Other works consider firms related policies. For example, Liberati *et al.* (2016) using difference-in-differences estimation argue that Italian science and technology parks have partly improved the economic performance and innovative capacity of firms located around them.

[7]For example, Bak (2020) recently tested the so-called Internet Skill Scale in Hungary to assess the level of digital knowledge. Nagy (2019) have compared the DESI of Hungary and Ukraine and concluded that Hungary is more developed with respect to key dimensions of digitalisation. Moroz (2017) exploits the DESI

Other works have investigated the degree of digitalisation of the economy and society at the local level.

Some scholars and practitioners have replicated the DESI indicator on specific case studies. For instance, RUSSO (2020) provides an application for Abruzzo, while RUIZ-RODRÍGUEZ *et al.* (2018) built the Enterprise Digital Development Index (EDDI) for EU states and Spanish regions.[8] With respect to the digital skills, BAK (2020) has used the Internet Skill Scale (ISS) to analyse the performance of Hungary.[9] Differently from our work, these examples are limited to a single digital dimension and do not offer a comprehensive portrait of the digital development.

Some studies are closer in spirit to the proposed rDESI, as they attempt to measure the digital performance in Italy at the regional level: Politecnico di Milano, Piedmont Region, and a CENSIS-TIM partnership have tried to replicate the DESI, while Unioncamere and Ernst and Young have created their own regional index to measure Italian firms' digital level.

The Politecnico di Milano produced a regional version of the DESI, employing 34 out of 44 DESI variables in the 2019 version and 35 out of 37 DESI variables in the 2020 version. In their latest work, the index measuring firms' integration of digital technologies does not have regional variation as it is based on data at the NUTS1 level. Also, some series used for the connectivity index lack regional detail.

Compared to the Politecnico, we improved the index, by employing regional data for firms' integration of digital technologies (see Section 2 for details), and by resorting to new data sources for connectivity and e-government.[10]

The regional centre for ICT of Piedmont has conducted an analysis in 2019 and in 2020 to describe the performance of Piedmont compared to the other Italian regions. In the 2019 edition, they accounted for all the five dimensions of the DESI and added three e-health variables to the e-government sub-component. However, their results are presented by sub-index, while the overall composite indicator is not available.

In December 2020, Censis (Center for Social Investments Studies) jointly with TIM, issued a report, investigating the pandemic-induced developments in the use of the internet in the Italian society. Quoting the DESI, they computed a province-level composite indicator

and the Networked Readiness Index (NRI), finding that Poland is less digitally developed than its peers.

[8] The key variables are identified with a factor analysis on series retrieved from the "Community survey on ICT usage and e-commerce in enterprises" of Eurostat.

[9] This skills measurement framework has been developed and validated by VAN DEURSEN *et al.* (2016) and is based on 35 questions organised into 5 factors (operational, information navigation, creative, social and mobile).

[10] With respect to measuring the e-government services, it is hard to understand whether our methodology is more accurate compared to the Politecnico's as they do not explain how this is computed. However our sub-index is highly correlated to theirs.

of digitalisation, resulting from the combination of 15 variables. The main flaw of this index, compared to ours, is that it entirely neglects the connectivity and human capital dimensions.

In the same period, Ernst & Young released a report presenting its Digital Infrastructure Index, a composite indicator measuring the degree of digital infrastructure within a territory, distinguishing between the connectivity infrastructure and the diffusion of the IoT.[11] Its main strength lies in providing a broader view of the technological factors enabling development.[12] Differently from other studies, they enriched the analysis providing some statistics at the NUTS3 level. However, due to the close focus on infrastructure, their work neglects other aspects of digitalisation.

Focusing on the digital competence of firms, Unioncamere appraises the businesses' digital maturity though a self-assessment test. The regional average represents the maturity level of the area.[13] This said, no unbiased assessment can be drawn from their data, since the sample consists of respondents who are self-selected by accessing a specific section of the Unioncamere's website.

On the whole, compared to previous attempts, our work provides the closest replica of 2020 DESI methodology at the regional level for Italy. In particular, we have improved the data sources used for the connectivity index and for the index of firms' digitalisation.

The rest of the paper unfolds as follows. Section 2 describes the methodology to construct the rDESI and its components; Section 3 illustrates the Italian regions' digital development according to our indicators; Section 4 investigates the main correlations among these indices and presents some robustness checks with respect to the aggregation weights; Section 5 shows the methodological limitations of the DESI and proposes possible improvements. The final section offers some concluding remarks.

2 Building a regional index of digital development

Closely following the methodology of the index elaborated by the European Commission, the regional DESI (rDESI) is structured along five dimensions: connectivity, human capital, use of internet, integration of digital technology, and digital public services.[14] Each dimension is measured by a composite index, which summarises several indicators, as reported in Table

[11] The connectivity infrastructure mainly refers to investments by TLC operators. The diffusion of Internet of Things (IoT) mainly depends on the degree of digitisation of the other types of infrastructures present in the territory: transport, energy, and environmental networks.

[12] More in detail, they considered fixed connectivity (from ADSL to FTTH), wi-fi and mobile connectivity (from LTE to 5G), IoT technologies (networks and sensors), for a total of 30 combined indicators.

[13] In their last report, Trentino Alto Adige ranked first for the digitalisation of small to medium enterprises (SMEs), while Molise ranked last.

[14] See the 2020 methodological note.

1.

All the indicators are normalised between 0 and 1 with the minimum-maximum method, where the lowest and highest value are taken from the Italian regional values. These indicators are in turn an aggregation of individual data series (appropriately normalised) mostly referred to 2019.[15] Therefore, the indicator provides a snapshot of the pre-pandemic situation.

The rDESI index is computed as follows:

$$\begin{aligned} rDESI = \; & 25 \text{ Connectivity } + 25 \text{ Human Capital} + \\ & 15 \text{ Use of Internet services } + 20 \text{ Int. of digital technology} \\ & + 15 \text{ E-government.} \end{aligned} \tag{1}$$

Table 1: Components of the connectivity index

DESI dimension	**main sub-component**
1 Connectivity	1a Fixed broadband take-up 1b Fixed broadband coverage 1c Mobile broadband 1d Broadband price index
2 Human capital	2a Internet User Skills 2b Advanced Skills and Development
3 Use of internet services	3a Internet use 3b Activities online 3c Transactions
4 integration of digital technology	4a Business digitisation 4b e-Commerce
5 E-government	5a E-government users 5b Pre-filled forms 5c Online service completion 5d Digital public services for businesses 5e Open data

In what follows, we present and describe the data series employed to construct each dimension of our metrics. Data for all 21 Italian NUTS2 regions were collected;[16] whenever the information for 2019 was not available, we relied on earlier data, as indicated below. No other imputation techniques were adopted. We also underline all instances in which the NUTS2 breakdown of the original DESI variable was not available and had to be replaced

[15] Earlier years are used if 2019 figures are not available.

[16] In many instances, we attributed to the provinces of Trento and Bolzano the regional value of Trentino Alto Adige because the provincial breakdown was unavailable.

with a close proxy, leading to minor discrepancies between the DESI and the rDESI.

2.1 DESI 1 - Connectivity

The connectivity dimension represents the essential infrastructure of a digital economy. In fact, it is well-known that in those rural areas characterised by low population density and high deployment costs, private investments are discouraged, creating a vicious circle of limited capacity, high prices, and low service demand (so called white clusters or market failure areas). Table 2 lists the variables that compose this indicator.

Table 2: Components of the connectivity index

id	basic sub-component	weight
1a1	Overall fixed broadband take-up	0.125
1a2	At least 100 Mbps fixed broadband take-up	0.125
1b1	Fast broadband (NGA) coverage	0.125
1b2	Fixed Very High Capacity Network (VHCN) coverage	0.125
1c1	4G coverage	0.117
1c2	Mobile broadband take-up	0.117
1c3	5G readiness	0.117
1d1	Broadband price index	0.150

The first two sets of sub-components in the connectivity index relate to fixed broadband take-up and coverage. They are measured as of 2019 and made available by the Authority for Communications Guarantees, henceforth AGCOM. With respect to the take-up, we employ the percentage of households subscribing to any fixed broadband (1a1) and fixed broadband of at least 100 Mbps (1a2). As of the fixed broadband coverage, we consider the percentage of households covered by fixed broadband of at least 30 Mbps in download (1b1) and the percentage of households covered by any fixed very high capacity network (VHCN, at a minimum speed of 1 Gbps) (1b2). The next three sub-components pertain to mobile connections: the percentage of populated areas with 4G coverage - measured as the average coverage of telecommunication operators in a given area (1c1), the number of mobile data subscriptions per 100 people (1c2), and the amount of spectrum assigned and ready for 5G use by the end of 2020 within the so-called 5G pioneer bands (1c3) - this indicator is available only nation-wide.

The last sub-component is the broadband price index, which measures the prices of representative baskets of fixed, mobile and converged broadband offers (1d1). We use the national indicator as we are unaware of any regional difference in broadband prices, nor local price discrimination policies.

2.2 DESI 2 - Human capital

The endowment of human capital is key to foster digitalisation, as it sustain both the demand and the supply of ICT tools. Interestingly, in the labour market, despite an increasing demand for ICT related skills and the soaring unemployment, firms suffer from a shortage of digital skills (Mahida and Ramadas, 2013; Jackman *et al.*, 2021).

Fully adhering to the DESI methodology, our regional composite indicator is based on the rich set of variables listed in Table 3.

Table 3: Components of the human capital index

id	basic sub-component	weight
2a1	At least basic digital skills	0.166
2a2	above basic digital skills	0.166
2a3	At least basic software skills	0.166
2b1	ICT specialists	0.166
2b2	female ICT specialists	0.166
2b3	ICT graduates	0.166

The first sub-set of the human capital index consists of indicators of individual digital competences and skills. The component is calculated as the weighted average (with equal weights) of the following three normalised indicators: (i) at least basic digital skills, (ii) above basic digital skills, and (iii) at least basic software skills.

According to the Eurostat definition, the measurement of skills is based on selected activities related to internet or software use, performed by individuals aged 16-74 in four specific areas (information, communication, problem solving, and software skills).

It is assumed that individuals having performed certain activities have the corresponding skills. According to the variety or complexity of activities performed, two levels of skills (basic and above basic) are constructed for each of the four dimensions. The source used by the European Commission for the international comparison is the *Community survey on ICT usage in Households and by Individuals* made available by Eurostat. The respective Italian survey is called *Aspetti della vita quotidiana*.[17] We rely on the NUTS2-level data, provided by Istat by aggregating micro-data.

The second component of human capital refers to advanced skills and development. Also in this case, three variables are combined averaging over equal weights. The reference statistics are: (i) share of ICT specialists, (ii) share of female ICT specialists, and (iii) share of ICT graduates.

[17] All activities listed in the methodological manual of Eurostat are present in the corresponding Istat survey starting from 2019.

The source of information for these variables is the 2019 ISTAT Labour Force Survey (*Indagine sulle forze lavoro*). ICT specialists are identified through the categories of the International Standard Classification of Occupations (ISCO-08) by the International Labour Organisation (ILO), which groups jobs according to the tasks and duties undertaken. The categories of workers considered as ICT specialists are the following: ICT Service managers, Software and multimedia developers and analysts, Database specialists and systems administrators, ICT operations and user support technicians, Communications technicians, and Electronics, Telecommunications Installers and Repairers.[18]

The variable on graduates in ICT degrees is computed from the data made available each year by the Ministry of Education, University and Research (MIUR). The indicator is the share of ICT graduates over the total number of graduates, based on the region of students' residence.

2.3 DESI 3 - Use of Internet services

The index component capturing the use of internet services is closely related to the digital skills' one (see Subsection 4: an individual with low digital skills is less likely to use extensively internet services. Due to this redundancy, the European Commission has established to drop the indicator by the next DESI release, in 2021 (see Section 5). As of the present edition, the index focuses on three different dimensions: (a) general Internet usage; (b) basic online activities; (c) online transactions (banking, buying and selling goods or services). Consistently with the EU-DESI, all data refer to 2019.

[18] It should be noted that the European Commission's DESI is based on 4-digit ISCO codes, and therefore it includes also Electronic engineers (2152), Telecommunication engineers (2153), Graphic and multimedia designers (2166), Information technology trainers (2356), ICT sales professionals (2434), Electronics engineering technicians (3114).

Table 4: Components of the use of Internet services index

id	basic sub-component	weight
3a1	Individuals who never used the internet	0.125
3a2	Internet users (at least once a week)	0.125
3b1	Reading online news sites, newspapers or news magazines	0.083
3b2	Playing online or download games, images, films or music	0.083
3b3	Using online video on demand services	0.083
3b4	Making telephone or video calls (e.g. Skype)	0.083
3b5	Participating in social networks	0.083
3b6	Doing an online course (on any subject)	0.083
3c1	Using online banking	0.083
3c2	Ordering goods or services online	0.083
3c3	Selling goods or services online	0.083

Table 4 shows the list of variables and their weights. Several data sources are employed: for 3a1, 3a2, 3b5, 3c1, 3c2 and 3c3 we retrieved regional data from Eurostat, the same source employed by the European Commission. For all the other series, we resort to Istat; in particular, for 3b1, 3b2, 3b3 and 3b4, data are sourced from the survey *Cittadini e ICT*, while for 3b6 from *Multipurpose survey on households: aspects of daily life - general part*. The main difference is represented by the age group each statistic refers to: Eurostat data refer to 16-74 years old people, whereas *Cittadini e ICT* considers the 14-74 years old range and the *Multipurpose survey of households* includes individuals aged 6 and over. All indices are referred to the share of individuals who used Internet in the last 3 months, excluding 3c2, which relates to individuals who used Internet in the last year, and 3a1 and 3a2, which consider all individuals.

2.4 DESI 4 - Integration of digital technology

The digital transformation is also a fundamental pillar for industrial growth, as the economy is increasingly digitalised. This subsection presents the source of data used to monitor the adoption of digital technologies by Italian firms. In particular, the indicator looks at the use of big data, cloud services, and e-commerce, while overlooking other technologies (see Section 5 for a discussion of the potential limitations of this approach).

In order to construct our indicator, we rely on data collected by Istat in the context on the *Survey on the Use of ICT by Businesses*, for the time range 2018-2020.[19] Firms are located according to their legal headquarters.

[19] The statistics have been published in the *Rapporto sul territorio 2020. Ambiente, economia e società.*

Table 5: Components of the integration of digital technology index

id	basic sub-component	weight
4a1	Electronic information sharing	0.10
4a2	Social media	0.10
4a3	Big data	0.20
4a4	Cloud	0.20
4b1	SMEs selling online	0.13
4b2	e-Commerce turnover	0.13
4b3	Selling online cross-border	0.13

Table 5 shows the list of variables and their weights. More in detail, the Digital Integration component is the average of the percentage of firms with Enterprise resource planning and electronic information sharing in 2019 (4a1); the percentage of firms present on at least 2 social media in 2019 (4a2); the percentage of firms analysing big data in 2018 (4a3); the share of businesses purchasing cloud computing services in 2020 (e.g.: hosting of the enterprise's database, accounting software applications, CRM software, computing power) (4a4); the share of SMEs selling online at least 1% of their turnover (4b1); the ratio of e-commerce turnover over total turnover in 2020 (4b2); and the percentage of firms selling online cross-border in 2019 (4b3). Differently from Eurostat, which refers to companies with 10 to 249 employees, our data also include larger firms (i.e. with 10 or more employees).

2.5 DESI 5 - E-government

E-government refers to the digital provision of public and administrative services, which requires a joined-up technology to run both front-end and back-end operations in a cohesive fashion. For example, the virtual ID, in Italy called SPID, is intended to make easier to fulfil basic administrative tasks online, such as registering a new residency, compiling income statements, registering a car, or applying for public child-care. To fulfil these requests, official databases need to be integrated and accessible to the national or local authority the citizen is interacting with. Similar infrastructures are required to develop e-health services based on electronic patient databases, such as the European Union's "digital green certificate", whose functioning depends on Covid-19 vaccine data being digitised and collated.

Table 6: E-government index components

id	main sub-component	weight
5a	E-government users	0.20
5b	Pre-filled forms	0.20
5c	Online service completion	0.20
5d	Digital public services for businesses	0.20
5e	Open data	0.20

As mentioned earlier, our methodology for the e-government index diverges substantially from the European Commission's due to a lack of comparable data at the regional level. At this level it is necessary to consider the quality of digital services offered by local authorities such as municipalities (the Commission uses a sample of few large cities).[20] For this reason we propose several proxies to replicate the dimensions that compose the original index.

Table 6 shows the five main components of the e-government index. For all indices but the first we employed several proxies to avoid relying on a single source. In this context, data are collected only once in a while, so we combine data for 2018 and 2020.

The first item (e-government users, 5a) is defined as the share of individuals aged 14 and above who have submitted completed forms to any public authority in the last 12 months. Every year, Istat reports this figure at the regional level.

The second indicator of e-government (5b) is meant to measure the amount of data which is pre-filled in public service online forms. An essential step to this purpose is the creation of authentic sources, i.e. base registries used by governments to automatically validate or fetch data on citizens or businesses. Since we do not know the exact amount of data actually pre-filled for e-government users at the regional level, we employ three proxies, measuring the development and usage of authentic sources at the local authority level. In particular, we compute a simple average of the following five statistics: (i) the share of population living in municipalities which have joined the national registry (Anagrafe nazionale);[21] (ii) the average number of management systems that are integrated within a local authority (2020 survey managed by the Bank of Italy);[22] (iii) the average number of data-sets used by local

[20]The cities sampled by the European Commission are: Bari, Bologna, Brescia, Catania, Firenze, Genova, Messina, Milano, Modena, Napoli, Padova, Palermo, Prato, Reggio-Calabria, Roma, Taranto, Torino, Trieste, Venezia, and Verona.

[21]A main purpose of the Anagrafe nazionale is to allow e-government users to be more easily identified without asking them for official certificates, e.g. residence certificate.

[22]The management systems considered are: accounting; register of suppliers; HR management; procurement; secretariat, document protocol; treasury (with SIOPE+); registry office (for municipalities only); local taxes record.

authorities (2019 survey commissioned by the "Corte dei Conti");[23] (iv) the share of local authorities which allow authentication to online services via SPID (2019 survey commissioned by the "Corte dei Conti"); (v) the share of municipalities that have received payments through pagoPA in 2020.[24]

As a third element, the European Commission produced the online service completion indicator (5c). This measures the share of administrative steps that can be carried out online in several situations, such as changing job or moving to a new city. We substitute this indicator with the simple mean of three indices: (i) index of online service availability for services offered by municipalities in 2018 (computed by Istat); (ii) completeness of online services by local authorities (2020 survey run by the Bank of Italy); (iii) share of local authorities offering at least one online service (2019 survey commissioned by the "Corte dei Conti").

The fourth index (5d) broadly reflects the share of public services needed to open a business and conduct regular business operations that can be performed online by domestic as well as foreign users. This dimension is hard to capture at the regional level; in order to do so, we resort to two data-sets: (i) the share of municipalities offering online services of the "Sportello Unico Attività Produttive" (SUAP) and of the "Sportello Unico Edilizia" (SUE; 2020 survey run by the Bank of Italy);[25] (ii) the share of municipalities offering online services for SUAP and SCIA-VIA (2019 survey commissioned by the "Corte dei Conti").[26] In both cases, we assign one point if the municipality offers both services online, half point if it offers only one service and zero if neither are offered.

The European Commission's open data index (5e) measures to what extent countries have an open data policy in place (including the transposition of the revised PSI Directive), the estimated political, social and economic impact of open data and the characteristics (functionalities, data availability and usage) of the national data portal. Our proposed regional version of this metrics relies on data regarding specific results achieved by local

[23]The dataset included in the list are: (1) national registry of territorial data; (2) registry office (ANPR); (3) criminal record; (4) business record; (5) immigration and asylum seekers record; (6) agricultural business record; (7) motor vehicles registry (PRA); (8) federated national information system of infrastructures (SINFI); (9) national archive of house numbers of urban streets (ANNCSU); (10) land registry.

[24]pagoPa is a new payment method, introduced by the Italian Code for Digital Administration and by the Decree-law 179/2012 with the aim of guaranteeing safe and reliable electronic payments to the public administration. It is managed by PagoPA S.p.A., a public company owned by the Italian Ministry of Economy and Finance.

[25]The SUAP is an office that every Italian municipality must run since all companies and entrepreneurs need to interact with their local SUAP to open, transfer, change or close a business. The SUE is another office that can be found in every Italian municipality and it is in charge of monitoring all requests of construction works.

[26]The SCIA ("Segnalazione certificata di inizio attività") and the VIA ("Valutazione impatto ambientale") are two formal protocols that are required to start construction works.

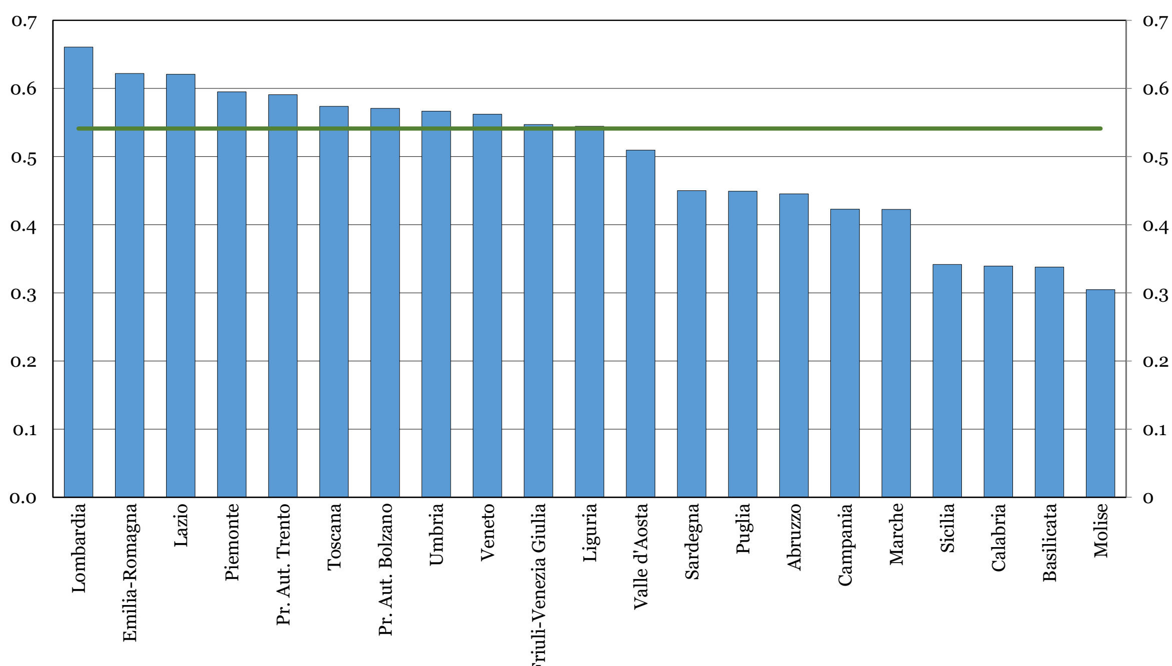
public authorities. In particular, we consider three indices: (i) the regional open data index released by the government and based on a basket of datasets selected for their relevance; (ii) the share of municipalities that provided open data online in 2018 (survey run by Istat); (iii) the share of local authorities providing online free-access open data in 2020 (survey run by the Bank of Italy).

While representing a valid alternative to the European Commission index, our measure of regional e-government is not exempt from flaws. The main concern is the uncertainty in future data collection, which might prevent updating and monitoring changes in regional e-government services and in their quality (see Section 5).

3 Analysis of the rDESI

The regional DESI is plotted in Figure 1, where regions are sorted from the highest to the lowest score, and the green line represents the national average.

Figure 1: Regional DESI



The leading regions in the rDESI index are Lombardy, Lazio, and Emilia Romagna, while Calabria, Basilicata, and Molise are the least digitalised areas. More generally, the index confirms the gap among Northern and Southern Italian regions that we observe for a large range of statistics (e.g. GDP per capita, activity rate, etc.). With respect to the rDESI, almost all Northern regions score above the Italian average,[27] while the Southern ones are all below the national mean. Figure 2 presents a set of regional maps of Italy where each indicator is displayed through a colour map of the regional values.[28]

The connectivity index rDESI 1 (Figure 2b) has the lowest variance among all five sub-indicators and does not show the North-South gap described before, displaying a more homogeneous distribution across all the Italian regions. This result can be at least in part ascribed to the natioal Government's policies[29] that have financed the construction of a fixed broadband infrastructure mostly in those areas where undertaking the investment was regarded as unprofitable (so called white areas or market failu zones). Moreover, mobile 4G connection is nowadays highly widespread over the whole nation. Within this index, Lazio, Liguria ad Emilia Romagna are the top three regions, while Valle d'Aosta, Basilicata and Molise lie at the bottom of the distribution.

The human capital index rDESI 2 (Figure 2c) displays the highest variance among all five sub-indicators, showing a large gap between Northern and Southern regions, both in the basic individual digital competences and with respect to ICT specialists and graduates. Recall that in the nationwide metrics, Italy ranks last among the European countries, pointing out critical issues in the education domain. In the regional ranking, Lazio, Lombardy and Provincia Autonoma di Trento are at the top, while Basilicata, Calabria and Sicily show the lowest scores.

The use of internet index rDESI 3 (Figure 2d) follows a similar pattern as the previous indicator (see Section 4). The top/last three regions ranking is almost confirmed: Lombardy, Emilia Romagna and Tuscany are the best three regions, while Basilicata, Calabria and Sicily are the last three.

The integration of digital technology by firms, indicator rDESI 4, (Figure 2e) shows the North-South gap for the first component (digital integration), which disappears for the e-commerce leg. As it will be discussed in Section 5, this evidence is in line with the broadly accepted stylised fact that the *e-commerce* component tends to relatively favour economies with higher weight of the service sector. In particular, both *e-commerce* sub-components (number of SMEs that have sold their products online and share of turnover from online sales) are more than double in private services than in industry.[30] The top three regions in

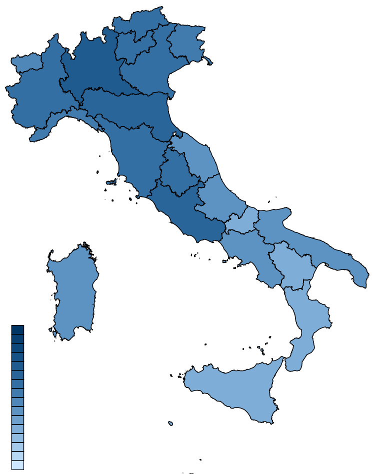
[27] Valle d'Aosta is the only exception, which can be partly explained by the poor results in the connectivity index, since it is a small mountain region.

[28] Darker shades of blue refer to higher values of the indicator. In the appendix, Figure A.1 reports for each region the values of the rDESI and of its main sub-components for a more detailed comparison.

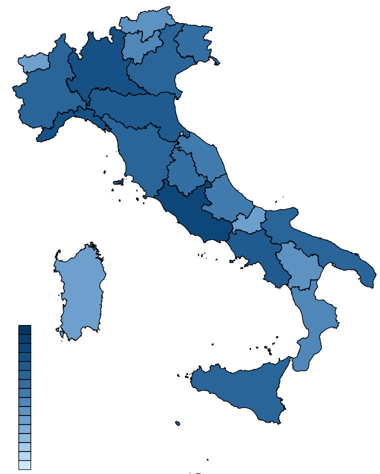
[29] Since 2009, the government launched the Piano Nazionale Banda Larga and Progetto Strategico Banda Ultralarga.

[30] See for instance the data from Istat (2020) Survey on the Use of ICT by Businesses.

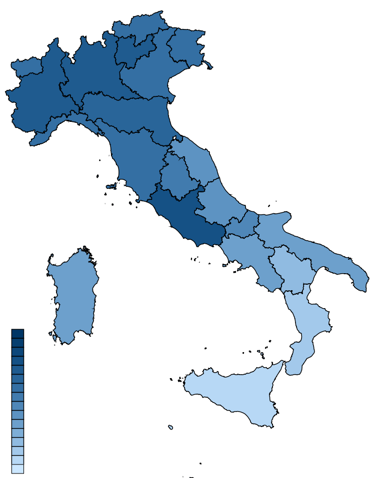
Figure 2: Regional digitalisation levels



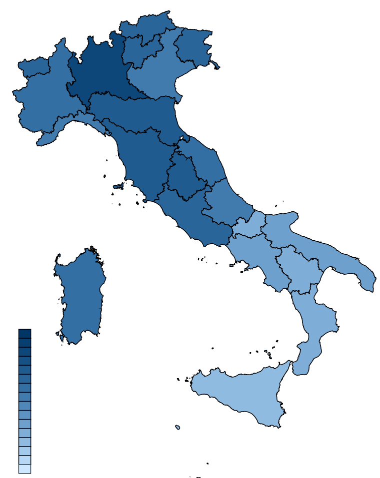
(a) rDESI



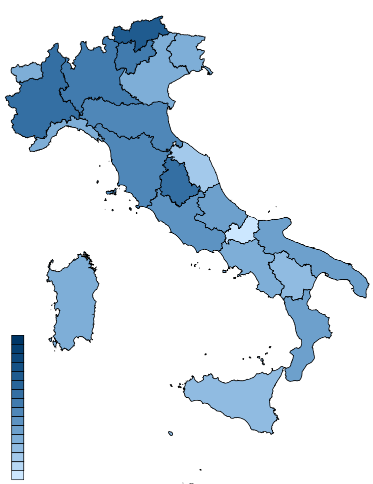
(b) Connectivity



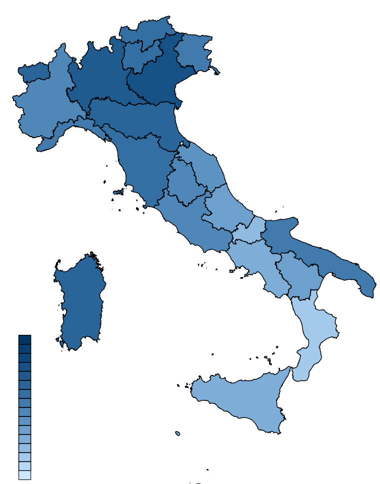
(c) Human capital



(d) Use of internet



(e) ICT integration



(f) e-government

this case are Provincia Autonoma di Bolzano, Umbria and Piedmont, while Sicily, Marche and Molise occupy the bottom of the distribution.

Lastly, the e-government index rDESI 5 (Figure 2f) reveals a better performance for Northern regions compared to the South-Center area, mostly driven by the supply of online services provided by local authorities. The top three regions in the ranking are Veneto, Lombardy and Emilia Romagna, while Campania, Molise and Calabria score in the last three.

4 Statistical properties of the rDESI

The different dimensions of a digital economy and society are interconnected and, as such, only concerted improvements in all of them would have a significant impact on people's life. In this section, we analyse the correlations among the rDESI components to highlight possible interplay and enabling roles.[31]

Table 7: Correlation matrix of DESI components

	(1)	(2)	(3)	(4)	(5)
(1) Connectivity	1.00	0.25	0.22	0.31	0.22
(2) Human Capital	0.25	1.00	0.82	0.52	0.64
(3) Use of Int. services	0.22	0.82	1.00	0.62	0.77
(4) Integration of dig. tech.	0.31	0.52	0.62	1.00	0.49
(5) E-government	0.22	0.64	0.77	0.49	1.00

Table 7 shows the correlation matrix of the five aggregate indicators. Correlations are strong and larger than 0.5, except for the connectivity index, due to its supply components (as shown below). The strongest relationship is observed between the human capital index and the use of internet services and might suggest that skills are essential to the widespread use of digital services.[32] Interestingly though, it is the use of internet that more strongly correlates with the other two indices (integration of digital technology and e-government), pointing to the importance of the demand side in the development of a digital society. At the same time, this evidence shows that the third indicator could be redundant – in fact, the European Commission intends to exclude it from the future version of the DESI.

[31]In the literature, Bánhidi *et al.* (2020) have proposed a similar multivariate statistical analysis of the original DESI across European countries.

[32]In this regard, it is worth remembering that Italy's human capital index is far behind the EU average.

Table 8: Correlation of connectivity sub-components

	(1a) Fixed broadband take-up	(1b) Fixed broadband coverage
(2) human capital	0.64	0.04
(3) use of internet services	0.68	-0.02
(4) integration of digital tech.	0,52	0.10
(5) e-government	0.47	-0.03

Table 8 suggests that the low correlation of connectivity with the rest of the rDESI components is mainly driven by the low correlation of supply of fixed broadband (1b) despite it has a 0.5 correlation with broadband take-up.[33] Mobile connectivity is not considered here because some components are available only at the national level and because the 4G indicators display very little variation (likely produced by sheer geographical factors – see Subsection 2.2).

Table 9: Correlation of use of internet services

	(3a) Internet use	(3b) Activities online	(3c) Transactions
(2a) Internet user skills	0.72	-0.05	0.94
(2b) Advanced skills	0.52	0.11	0.33
(4a) Business digitisation	0.49	0.41	0.31
(4b) E-commerce	0.49	-0.16	0.45
(5a) E-government users	0.65	-0.02	0.80
(5b-d) E-government offer	0.53	0.02	0.63

Table 9 shows the correlation of the components of the use of internet services with components of the other indices. In particular, it can be noticed that internet use (3a) and transactions (3c) strongly correlate with other indicators, while online activities (3b) does not. Furthermore, basic skills of internet users tend to be much more correlated both with use of internet and transactions, compared to the endowment of advanced skills.

[33] As we discuss later on, there is an issue of consistency between these two indicators due to the use of different data sources (AGCOM and Istat).

Table 10: Correlation of integration of digital technology

	(4a) Business digitisation		(4b) E-commerce	
	correlation	partial correlation	correlation	partial correlation
(2a) Basic digital skills	0.28	0.13	0.44	0.42
(2b) Advanced skills	0.42	0.33	0.27	0.19
Share of industry	0.33	0.23	-0.05	-0.22

The partial correlation between y and x1 is an attempt to estimate the correlation that would be observed between y and x1 if the other x's did not vary. This squared correlation can also be interpreted as the decrease in the model's R2 value that results from removing x1 from the full model.
Industry includes : (B) mining, (C) manufacture, (D) utilities and (E) waste collection and water supply.
The share of industry source is Istat, ASIA ("Registro statistico delle imprese attive") for 2018.

Table 10 shows correlation and partial correlation coefficients for the two components of the integration of digital technology index (4a and 4b) with respect to skills and to the share of employees in the Industry sector at the regional level.[34] We observe a negative correlation between e-commerce and the Industry share, supporting the idea that this sub-index favours service-based economies. On the other hand, the share of manufacturing is positively associated to the business digitalisation index, confirming the idea that digital innovations, such as cloud or big data are more likely to be adopted by manufacturing firms. In either case, we do not observe a strong relationship. Moreover, advanced skills seem to be more correlated with business digitalisation than with e-commerce, whereas the opposite holds true for basic digital skills, suggesting that different kinds of skills can trigger different digital transitions.

Table 11: Correlation of integration of digital technology

	(5a) e-gov. users		(5b-5e) e-gov. services	
	correlation	partial correlation	correlation	partial correlation
(2a) Basic digital skills	0.82	0.83	0.50	0.45
(2b) Advanced skills	0.28	0.12	0.12	0.01
Share of old age residents	0.20	-0.40	0.31	0.09

The partial correlation between y and x1 is an attempt to estimate the correlation that would be observed between y and x1 if the other x's did not vary. This squared correlation can also be interpreted as the decrease in the model's R2 value that results from removing x1 from the full model.
The share of old age residents is defined as the ratio between residents at January,1st 2020 older than 64 and those older than 14; the source is Demoistat.it.

Finally, table 11 reports the correlation and partial correlation coefficients for the demand and for the supply indicators of e-government services with respect to skills and share of

[34] We use only partial correlation as an exploration of the data because the limited number of observations prevent robust estimations of e.g. OLS coefficients.

senior residents (defined as share of 65 and older over the population aged 15 or more).[35] The share of e-government users is highly correlated with the presence of basic digital skills, while the correlation with the endowment of advanced skills is very low. This suggests that e-government instruments can be successful only if they can rely on a wide mass of citizens with basic digital skills. Moreover, if skills were constant across regions, the share of old age residents would unsurprisingly have a negative correlation with the e-government users index. The offer of online public services is less correlated with basic digital skills and the correlation with the advanced ones is close to zero.[36] These findings suggest that supply might be often demand-driven, inducing a vicious circle that could lead to a *digital poverty trap*: local authorities pursue digital innovation where there is more demand for e-government services, while do not seem able to foster new demand.

Table 12: Correlation between e-government sub-indices

	(5a)	(5b)	(5c)	(5d)	(5e)
correlation with (5a) e-gov. users	-	0.11	0.57	0.26	0.43
coefficient of variation	0.50	0.26	0.40	0.43	0.33

The coefficient of variation is the ratio of the standard deviation to the mean.

In order to shed further light on this intuition, we consider the sub-components of e-government, distinguishing between supply and demand-side ones and analyse both their correlations and their degree of variability. Table 12 shows that the share of e-government users is more correlated with online service completion and less with pre-filled forms and open data sub-indices. Moreover, the latter are more homogeneous among regions, showing a relatively lower coefficient of variation. These two facts taken together suggest that e-government services, presenting a higher degree of regional heterogeneity, might endogenously reflect differences in demand, whereas advancements in pre-filled forms and open data, whose development follows policies set by the national Government, are quite exogenous with respect to demand characteristics.

Given the strong correlations among DESI components, in Appendix B we consider the application of the principal component analysis to these sub-indices. The digitalisation metrics resulting from the PCA has a 97 percent correlation with the regional DESI. This result logically descends from the strong correlations among almost all the series; as such, the weighting procedure has only a minor impact on the ranking, since a poor (or good) performance in one indicator is likely to be replicated in the other components. However,

[35] Note that we calculate "e-government supply" as the simple mean of indicators 5b, 5c, 5d, and 5e, since they all represent different aspects of local government services provision.

[36] Unfortunately we are not able to control for education or skills of local civil servants.

this also suggests that the aspects of demand are over-represented in the DESI framework and that greater weighting could have been assigned to the supply indicators in order to improve the effectiveness of the indicator in capturing both sides of the market.

5 Methodological gaps and way forward

The Commission's Communication "2030 Digital Compass: the European way for the Digital Decade" of 9 March 2021 laid out the vision for a successful digital transformation of the European Union by 2030. The EU's goal is to be digitally sovereign in an open and interconnected world, and to pursue digital policies that empower people and businesses to seize a human centred, inclusive, sustainable and more prosperous digital future. This includes addressing vulnerabilities and dependencies as well as accelerating investment. In the light of these ambitions and challenges, the Commission proposes the Policy Programme "Path to the Digital Decade". In terms of targets and indicators, this relies on an *enhanced* version of the DESI, serving as the analytical basis for the Report on the state of the Digital Decade (RSDD). In this regard, Member States expressed strong support for enhancing DESI as a monitoring tool for the Digital Decade, with consistent but adaptable indicators to facilitate comparisons across time and space. When selecting targets, the Commission looked at existing Key Performance Indicators (KPIs), having in mind that targets need to be measurable in order to be monitored. The choice of lead KPIs was guided by existing indicators, currently in the DESI, appropriately modified and improved upon, which will be enriched with new potential KPIs, should studies (ongoing or future) or other sources be needed, methodologies still be developed or relevant data still be acquired to measure progress towards the 2030 targets.

In this spirit, the present section contributes to the debate on what might be revised of the current DESI index. We make some suggestions for a better representation of national and regional digital levels, addressing issues regarding both the methodology and the data sourcing.

The DESI is thought as an instrument to conduct policy evaluation in a standardised way across European countries and regions. To this end though, it is important that data are provided at yearly frequency and that no imputation is required to fill gaps. This would enable more accurate analyses of policy impacts. Moreover, there is room for improvements in terms of data harmonisation, as some countries still lag behind in collecting data, affecting the overall index quality.

Digital competences and technical solutions are rapidly evolving as new technologies get spread, contributing to the obsolescence of some indicators. In this sense, we face a tradeoff between the relevance of the variables and the continuation of time series, particularly relevant in monitoring digital progress. Furthermore, requiring the national and European statistical offices to add new survey questions might result into an excessive burden to respondents.

In general, greater attention should be placed to avoid conceptual ambiguity. At the moment, there might be a measurement issue stemming from the lack of clarity in some survey questions. For example, concerning ICT adoption by firms, cloud services have evolved over time into a variety of sophisticated applications, to the point that the most basic online storage systems might be omitted by the respondents when listing the adopted cloud services. A more stringent definition should be provided when asking about the adoption of particular technologies, in order to avoid different interpretations across countries or across sectors.

Moving to the DESI 1 indicators measuring connectivity, they do not properly take into account the actual speed, nor the quality of connections. While we do have information on the maximum navigation speed ensured by a given technology, it would be more appropriate to base the indicator on estimates of the actual performance. These data are available for Italy and could be sourced from AGCOM, but many countries do not have equivalent disaggregated high quality information, making the transition towards more accurate indicators quite problematic.[37]

Sources between take-up (Eurostat) and coverage (i.e. supply, sourced by AGCOM) should be coherent, using either a single source or at least the same definitions of technology or connection speed. This would ensure a more consistent comparison between demand and supply, within and across countries. Finally, the broadband (2 Mbps) take-up and the 4G coverage are now clearly outdated. In DESI 2021 European Commission will introduce the 5G coverage and the at least 1 Gbps take-up, without removing the outdated ones.

With respect to the human capital dimension, the DESI methodology annex could better clarify the exact source of information needed to measure the share of ICT graduates in the population. While the ISCED level qualifying a graduate person is specified, the list of ICT degrees is not available.[38] With respect to the proxy based on the share of ICT graduates, two alternative definitions could be used. The index could be based on the share of ICT graduates over the total tertiary educated population (i.e. the stock of graduates). This

[37]The theoretical speed approach penalises countries that adopt alternative technologies to address local challenges (e.g. geographical ones) with cost-effective solutions, such as internet dongles, fixed wireless access, or consumer grade satellite Internet service.

[38]The degrees we identified as relevant for ICT are informatics, information, IT and telecommunication engineering. The degrees in Italian are: Informatica, Ingegneria informatica, Ingegneria dell'informazione, Ingegneria delle telecomunicazioni, Sicurezza informatica, Scienze e tecnologie informatiche, Metodologie informatiche per le discipline umanistiche, Specialistiche in ingegneria informatica, Specialistiche in ingegneria delle telecomunicazioni, Specialistiche in informatica.

interpretation of the definition implies a very persistent indicator, with only marginal annual updates. As an alternative, the index could employ the share of the population graduating in ICT subjects over the total number of students graduating in a given solar year (i.e. the annual flow of graduates). In the latter case, the relevant region could be the region of origin of the student, or the region where the university granting the degree is located. We point out that which of the two definitions is chosen is particularly important in the computation of the rDESI for Italy, but also in the methodology of the European Commission.

Apart from the issue of whether is best to measure the stock or the flow of graduates, the current methodology also misses human capital relocation, i.e. the very fact that an increasing share of graduates moves to a different country to work.[39] As a starting point, the methodology should explicitly handle incoming foreign students and/or people moving abroad to study.[40]

When accounting for the digital skills, one may question whether the DESI is effectively capturing the relevant "soft skills".[41] As a matter of fact, it is not clear whether nowadays browsing on the internet for various purposes corresponds to an actual communication and/or information skill. In parallel, mastery of digital skills should include the ability to verify the truthfulness of web content. Next, we point out that deficiencies in software skills (2a3)[42] accrue three times in the overall formula; indeed, according to the definitions, having at least basic digital skills (2a1) amounts to scoring "above basic" in all 4 domains, including software skills. Similarly, having above basic digital skills (2a2) implies having at least one "basic" but no "no skills" in all 4 domains.

Regarding the use of internet, we have observed a strong correlation of this index with the human capital one. Although it is not surprising that skills and internet usage are strongly associated, we argue that methodologically this is also due to an overlap of some sub-indicators. As a matter of fact, within the DESI 3, we measure the proportion of citizens who have recently performed a given set of online activities. Some of these activities coincide with those used to measure the digital skills in one indicator of the DESI 2, e.g. the percentage

[39] This is the so-called *brain drain* or *brain gain* phenomenon that has grown drastically as the tertiary education markets have become global (see, for example, the work of Giousmpasoglou and Koniordos, 2017).

[40] Consider that the European Commission in 2021 has almost doubled the budget of the European student mobility programme (Erasmus).

[41] Before the introduction of DESI, these important concepts of digital competence and literacy were defined by Ferrari (2012).

[42] The activities to be performed to be considered skilled are the following: using word processing software; using spreadsheet software; used software to edit photos, video or audio files; creating presentation or document integrating text, pictures, tables or charts; using advanced functions of spreadsheet to organise and analyse data (sorting, filtering, using formulas, creating charts); have written a code in a programming language.

of individuals who used the internet to make telephone or video calls (3b4). This can be considered as a double counting and should be addressed. Yet, it should be noticed that the forthcoming methodology (2021) will entirely drop the third index from the enhanced DESI.

The range of applications of ICT technology is very broad and multifaceted. In some business contexts, the corporate use of cloud infrastructure or administrative software is extremely common. Nevertheless, it does not necessarily correspond to innovative technologies. Additionally, the DESI 4 methodology is likely to penalise economies with low services weight: the e-commerce and the social network indicators are lower in manufacturing (as suggested by the negative correlation between the share of workforce in the industry sector and the e-commerce performance of the region presented in Table 10).

In the forthcoming DESI 2021 methodology, this concern will be attenuated since the adoption of other digital technologies will be taken into account (i.e. artificial intelligence, e-invoices). Finally, DESI 4 indicators measure the integration of digital technologies, but there is no evaluation of the ability to create new digital technologies. Some possible indicators could be the number of ICT patents or the share of innovative SMEs.[43]

The e-government index also presents some worrying issues. First, the supply-side indicators are fed from a small-size survey; second, though the digital level of a country depends on both national- and regional-level digital services, this second dimension is highly overlooked. A more effective methodology would distinguish between the set of online services provided by the local governments and those provided by the national Government (possibly through its agencies) all over the country. Currently, the DESI is unable to capture the heterogeneity in the quality of digital services for local administrations, as the component related to local administrations is based on a non-representative sample data (few large cities), and likely underestimates territorial disparities.

This appears even more problematic regarding for those services supplied by regional public administrations, such as the management of the healthcare system in Italy, which represents the major activity for regional governments. To provide a more complete picture, it would be desirable to enrich the index with variables describing the level of e-health.[44]

[43] Paragraph 25 of the Italian law 221/2012 contains the Italian definition of innovative SMEs. The Italian Ministry of Economic Development collects quarterly the number of innovative firms by region.

[44] Even in this case, it would be best to distinguish national e-health services from regional ones, to achieve analyses at both levels.

6 Conclusions

In this document, we have presented the regional DESI for Italy, explaining differences and similarities with other works in the digitalisation literature. We have reviewed the methodology to build the regional indicators by adapting the European Commission's approach to the data available for the Italian regions.

Furthermore, we have shown some statistical properties of the rDesi index and its components; in particular, we have observed that weighting has little impact on the overall ranking of regions since the different dimensions are highly correlated. Based on this and other evidence, we proposed some amendments to the original methodology, to better capture some key aspects of digitalisation and to address some of the current data limitations.

The Covid-19 pandemic has greatly affected consumption habits and the organisation of work. While we do not have the data to consider the impact on the rDESI dimensions yet, we expect that the positive shock to the demand of digital services will lead to a general increase of the sub-components in the next edition. In this context, the PNRR can potentially leverage private investments in ICT and digital technologies. In July 2021, the Italian Government has committed almost 34 billion euros to the innovation and digital transition of firms and public administration; for firms resources will be allocated through the Transition 4.0 Plan. The rDESI's results suggest that the large gap in human capital compared to the other European countries may represent the greatest challenge for the investment plan of the Italian Government in years to come.

In this work, we have also pointed out the wide regional disparities in the digitalisation levels across Italy. We believe that more efforts should be made at every administrative level to properly measure the advancements in the relevant digital dimensions and, for this reason, we have proposed some ways forward. Along the same line, with a view to monitoring the digital developments of European regions and more broadly in the context of cohesion policy assessment, the European Commission might consider to transition towards a regional version of the DESI.

References

Bak G. (2020) Measurement of digital intelligence (dq), *Management, Enterprise and Benchmarking in the 21st Century* pp. 1–8.

Bánhidi Z., Dobos I. and Nemeslaki A. (2020) What the overall digital economy and society index reveals: A statistical analysis of the desi eu28 dimensions, *Regional Statistics* **10**(2), 42–62.

Campante F., Durante R. and Sobbrio F. (2018) Politics 2.0: The multifaceted effect of broadband internet on political participation, *Journal of the European Economic Association* **16**(4), 1094–1136.

Ferrari A. (2012) Digital competence in practice: An analysis of frameworks, *Sevilla: JRC IPTS.(DOI: 10.2791/82116)* .

García P. P., Thapa B. and Niehaves B. (2012) Bridging the digital divide at the regional level? the effect of regional and national policies on broadband access in europe's regions, Tech. rep., 13th International Conference on Electronic Government (EGOV), Dublin, Ireland.

Giousmpasoglou C. and Koniordos S. (2017) Brain drain in higher education in europe: Current trends and future perspectives .

Guriev S., Melnikov N. and Zhuravskaya E. (forthcoming) 3g internet and confidence in government, *Quarterly Journal of Economics* .

Jackman J. A., Gentile D. A., Cho N.-J. and Park Y. (2021) Addressing the digital skills gap for future education, *Nature Human Behaviour* pp. 542—-545.

Liberati D., Marinucci M. and Tanzi G. M. (2016) Science and technology parks in italy: main features and analysis of their effects on the firms hosted, *The Journal of Technology Transfer* **41**(4), 694–729.

Mahida D. and Ramadas S. (2013) The digital talent gap: developing skills for today's digital organizations, Tech. rep., Capgemini.

Mahida D. and Ramadas S. (2017) *Principal Component Analysis (PCA) based Indexing*, pp. 54–56.

Mammadli E. and Klivak V. (2020) Measuring the effect of the digitalization, *Available at SSRN 3524823* .

Mooi E., Sarstedt M. and Mooi-Reci I. (2018) *Principal Component and Factor Analysis*, pp. 265–311, Springer Singapore, Singapore.

Moroz M. (2017) The level of development of the digital economy in poland and selected european countries: a comparative analysis, *Foundations of management* **9**(1), 175–190.

Nagy S. (2019) Digital economy and society. a cross country comparison of hungary and ukraine, *arXiv preprint arXiv:1901.00283* .

Nucciarelli A., Castaldo A., Conte E. and Sadowski B. (2013) Unlocking the potential of italian broadband: Case studies and policy lessons, *Telecommunications Policy* **37**(10), 955–969.

Osborne J. W. and Costello A. B. (2004) Sample size and subject to item ratio in principal components analysis, *Practical Assessment, Research, and Evaluation* **9**(11).

Pereira J. P. R. (2016) Broadband access and digital divide, in *New advances in information systems and technologies*, pp. 363–368, Springer.

Ruiz-Rodríguez F., Lucendo-Monedero A. L. and González-Relaño R. (2018) Measurement and characterisation of the digital divide of spanish regions at enterprise level. a comparative analysis with the european context, *Telecommunications Policy* **42**(3), 187–211.

Russo V. (2020) Digital economy and society index (desi). european guidelines and empirical applications on the territory, in *Qualitative and Quantitative Models in Socio-Economic Systems and Social Work*, pp. 427–442, Springer.

Schaub M. and Morisi D. (2020) Voter mobilisation in the echo chamber: Broadband internet and the rise of populism in europe, *European Journal of Political Research* **59**(4), 752–773.

Szeles M. R. (2018) New insights from a multilevel approach to the regional digital divide in the european union, *Telecommunications Policy* **42**(6), 452–463.

Szeles M. R. and Simionescu M. (2020) Regional patterns and drivers of the eu digital economy, *Social Indicators Research* pp. 1–25.

Van Deursen A. J., Helsper E. J. and Eynon R. (2016) Development and validation of the internet skills scale (iss), *Information, Communication & Society* **19**(6), 804–823.

Appendices

A rDesi detailed ranking

In Figure A.1 we report the value of the rDESI and its components for each Italian region.

	rDESI	connectivity	human capital	internet users	integration of digital technologies	e-government
Lombardia	0.66	0.66	0.70	0.76	0.53	0.67
Emilia-Romagna	0.62	0.68	0.63	0.70	0.47	0.64
Lazio	0.62	0.75	0.72	0.63	0.44	0.47
Piemonte	0.60	0.61	0.68	0.58	0.57	0.48
Provincia Aut. Trento	0.59	0.54	0.69	0.64	0.51	0.57
Toscana	0.57	0.61	0.55	0.69	0.48	0.57
Provincia Aut. Bolzano	0.57	0.45	0.57	0.61	0.67	0.60
Umbria	0.57	0.59	0.50	0.67	0.59	0.49
Veneto	0.56	0.66	0.57	0.55	0.33	0.71
Friuli-Venezia Giulia	0.55	0.61	0.59	0.65	0.35	0.53
Liguria	0.54	0.73	0.57	0.53	0.31	0.50
Valle d'Aosta	0.51	0.45	0.56	0.63	0.33	0.64
Sardegna	0.45	0.46	0.37	0.56	0.32	0.63
Puglia	0.45	0.62	0.36	0.37	0.35	0.51
Abruzzo	0.45	0.54	0.40	0.52	0.37	0.39
Campania	0.42	0.67	0.35	0.35	0.34	0.32
Marche	0.42	0.51	0.40	0.59	0.23	0.41
Sicilia	0.34	0.68	0.11	0.27	0.26	0.33
Calabria	0.34	0.51	0.23	0.30	0.36	0.24
Basilicata	0.34	0.45	0.27	0.34	0.27	0.36
Molise	0.30	0.32	0.49	0.34	0.05	0.27

Figure A.1: Detailed rankings of the regional DESI

B Robustness checks: PCA weights

In principle, the purpose of principal component analysis (PCA) is to reduce a large number of series into a set of few variables containing most of the information of the original data.[45] In our setting, this methodology might shed some light on the limitations of relying on pre-set arbitrary weights to aggregate the main components. It should be noted that PCA has two main limitations: firstly, since each component is a linear combination of all underlying series, it is generally harder to interpret; secondly, since PCA relies on estimated correlations,

[45]See Mooi *et al.* (2018) for an introduction to PCA with STATA applications.

we do not have enough regions to use this methodology on a large set of variables.[46]

For these reasons we attempt PCA on all aggregated sub-indices with regional variation, except for (1c) and (3b), which perform very poorly on the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy.[47] The eigenvalues we obtain are reported in Table B.1.

Table B.1: Eigenvalues of the principal components

	eigenvalue	proportion of extracted variance	cumulative
Component 1	4.90	0.49	0.49
Component 2	2.15	0.21	0.70
Component 3	0.97	0.10	0.80
Component 4	0.74	0.07	0.88
Component 5	0.51	0.05	0.93
Component 6	0.26	0.03	0.95
Component 7	0.24	0.02	0.98
Component 8	0.13	0.01	0.99
Component 9	0.07	0.01	1.00
Component 10	0.04	0.00	1.00

We select the first four principal components, whose eigenvalues are largest, thus keeping most (88%) of the original variability in the data.[48] Table B.2 shows the corresponding eigenvectors (only the most significant, i.e. that are larger than 0.3 in absolute value). The first components seem to represent mostly demand factors (broadband take-up, internet use, e-government users) and basic skills. The second one correlates most with both connectivity indices and the advanced skills index. The indicators of business digitalisation, e-commerce and the group of e-government services, are more closely related to components three and four which barely contribute to the extracted variance (0.17 versus 0.70 of the first two components).

[46] In the literature, a commonly used rule of thumb is to have at least five observations for each input series – see e.g. Osborne and Costello (2004).

[47] The KMO estimates how much of the variables' variance could be due to a common underlying factor. While values close to 1 are ideal (i.e. validating the use of PCA), the minimum that should be achieved from this test is a value of 0.5. Overall, our KMO score is barely sufficient (0.7), consistently with the low number of observations.

[48] In the literature it is common to keep only the components whose eigenvalues are larger than one, however we have relaxed this rule of thumb to make sure all original components contribute to the composite index.

Table B.2: Eigenvectors of the principal components

	comp.1	comp.2	comp.3	comp.4	unexplained
(1a) Fixed broadband take-up	0.31	0.41			0.13
(1b) Fixed broadband coverage		0.62			0.13
(2a) Basic skills	0.38				0.10
(2b) Advanced skills		0.43		-0.33	0.14
(3a) Internet use	0.41				0.18
(3c) Transactions	0.40				0.05
(4a) Business digitisation				0.76	0.11
(4b) E-commerce			-0.59	0.40	0.15
(5a) E-gov. users	0.37				0.15
(5b-5e) E-gov. Services			0.71		0.11

This methodology can be also used to produce an aggregate index based on correlation of its components (Mahida and Ramadas, 2017). In particular, once PCA is performed, the weight of each indicator i can be calculated according to the formula:

$$W_i = \sum_{j=1}^{4} |L_{i,j}| E_j,$$

where E_j is the eigenvalue of factor j, and $L_{i,j}$ is the loading value of the i-th unit of grouping on the factor j.

The digitalisation index calculated from the PCA has a 97 percent correlation with the regional DESI. This result could be explained by the strong correlations among almost all the series; as such, weighting has only a minor impact on ranking since a poor (or good) performance in one indicator is likely to be replicated in the other indicators.[49]

[49] PCA-based weights are evenly distributed: the methodology assigns 18 and 22 percent weight to connectivity and skills, respectively, and around 20 percent weight to each of the other three dimensions. Within connectivity, take-up has a slightly higher weight than coverage.