

UNIVERSITÀ CATTOLICA DEL SACRO CUORE

Campus of Milan

Faculty of Economics

Master program in Economics



Europe's Digitalization and the EDIH initiative: What leads firms to participate?

Supervisor

Professor Maria Luisa Mancusi

Dissertation by

Federico Vicentini

Id number: 5112604

Academic Year 2023/2024

Contents

1	Introduction	3
2	Literature Review	7
2.1	Innovation, its drivers and barriers	8
2.2	Open Innovation Ecosystems	9
3	The EDIH Initiative	13
3.1	Before EDIHs: The DIH Initiative	13
3.2	The birth of the EDIH Network	15
3.3	Objectives and functioning of the EDIHs	16
3.4	Descriptive statistics on the EDIHs	19
4	The DMA Survey	24
4.1	Design and structure of the survey	24
4.2	Analysis of the survey responses	29
5	What are the drivers of participation?	48
5.1	Matching of treated firms with ORBIS	51
5.2	Creating the Control Group	53
5.3	Treated - Control Comparison	54
5.4	Geocoding, distance and density computation	58
5.5	Regression Model and Results	63

6 Discussion and Conclusions	74
-------------------------------------	-----------

A Additional Tables and Figures	78
--	-----------

Chapter 1

Introduction

In modern advanced economies, digitalization has become a fundamental driver of innovation and economic growth. However, Europe is lagging behind its global competitors across every metric regarding digitalization. Neither the public nor the private sector shines in this regard. The newly published Report on the Future of European Competitiveness (also known as the *Draghi report Draghi (2024)*) stresses the need to close the digitalization gap, in order to boost productivity in the EU and reach levels of GDP growth on par with those seen in the US¹.

Long before being stressed as a pivotal challenge by the Draghi report though, the European Commission has tried to tackle the issue of lagging digitalization in the EU through the Digital Europe Programme (DIGITAL), launched in 2021 as part of the Multiannual Financial Framework (2021-2027).

One of the main initiatives launched through DIGITAL was the financing and creation of a network of European Digital Innovation Hubs, defined as one-stop shops supporting companies (specifically SMEs) and public sector organisations (from now on, PSOs) to respond to digital challenges and become more competitive. Since the start of the EDIH initiative in early 2023, any private company or PSO across Europe

1. For a prior thorough examination of the causes of the productivity gap between the US and the EU, also blaming part of the gap on Europe's lagging digitalization, see Van Ark, O'Mahony, and Timmer (2008)

can interact with these hubs to receive support for digitalization initiatives.

Aiming to measure the impact of the EDIH initiative itself, the European Commission's Directorate-General for Communications Networks, Content and Technology (also known as DG CNECT) tasked the Joint Research Centre's Digital Economy unit with the design of a Digital Maturity Assessment Tool (DMA). The DMA has been built to give an indication of the effectiveness of the EDIH's intervention on the firms they help. It also provides the firm with information about their relative level of digital maturity compared to their peers across regional and sectoral dimensions.

Every SME or PSO that is a customer of the EDIH network has to complete the DMA survey prior to the EDIH's intervention. The survey is then repeated 1 year later, and then a final time 2 years after the second submission. Since the EDIH Network has started operating in early 2023, at the moment most of the firms are still in the initial stages of intervention².

Unfortunately, at the current stage, it is impossible to have sufficient data to investigate the impact of the EDIH initiative on the firms that participated. Therefore, I will first present the current state of the firms that entered the program, providing information on their current level of digital maturity across the dimensions measured by the DMA. After that, I will investigate the drivers that led these firms to participate in the program itself.

In order to do this, I am going to match the firms that participated in the program with the ORBIS database, to retrieve firm-level financial information. I will then construct a synthetic control group of untreated firms, taken randomly from the ORBIS database, to compare with the treated sample by running a probit regression on firm-level information.

The research question I will therefore try to address with this thesis is the following:
What are the drivers of firms' participation in the EDIH initiative?

I will investigate this aspect of the EDIH policy by looking at financial and sectoral

2. As a consequence of this, my analysis will be mostly limited to T_0 observations, coming from the very first submission of the surveys, prior to EDIH intervention.

characteristics of the firms taken from the ORBIS database, as well as geospatial information, such as the distance of firms from the nearest EDIH and the "firm density" in the area around the firm's location.

Doing this will be helpful to assess the effectiveness of the EDIH initiative in multiple ways. For instance, the analysis will help us understand whether the program is effectively achieving its stated objective of targeting SMEs that would not have digitalized by themselves. For instance, if we find that firms that are more indebted are more likely to participate in the program, this could be an indication of efficacy in targeting financially constrained firms, which are less likely to have the resources to invest heavily in digitalization. Moreover, investigating the drivers of participation will be instrumental in understanding how to eventually increase the participation in the program itself, if future analysis will indeed show that the program is in fact effective in helping firms' digitalization. For example, if we find that firms that are closer to an EDIH are more likely to participate, this could indicate that an increase in the number of EDIHs or a better geographical distribution of them, can lead to a higher number of firms participating in the program.

Moving forward, this thesis is structured in 4 parts. In the first part, I will present a brief review of the literature on digitalization. The review will mainly focus on studies covering digitalization initiatives by governments, the relevance of firms' characteristics on their level of innovation (for example the role of financial constraints), and the role of Open Innovation ecosystems, such as the EDIH Network.

In the second part, I will take an in-depth look at the EDIH initiative, by covering the objective of the program, its birth and development, and its current state 1 year after its official launch.

With the third part, I will explore the Digital Maturity Assessment Tool, presenting the design and structure of the survey. Here I will also present a breakdown of the survey responses, which will give a snapshot of the level of digital maturity and the adoption of specific technologies in the sample of treated firms.

Throughout the fourth part, I will present the core of the analysis. Here the main

research question, regarding the drivers of participation in the EDIH initiative by firms in the EU, will be explored. I will describe how treated firms have been matched with ORBIS firm-level data, then how the sample of control firms is constructed. After that, a comparison of the two sample is presented, followed by the explanation of the regression model utilized for the analysis and the results from the probit regression.

The fifth and last part will deal with the discussion of results and conclusions stemming from them, as well as possible avenues for further research.

Chapter 2

Literature Review

Presenting in this chapter a literature review on the exact subject of this thesis would have been impossible, since the EDIH initiative is still young. Thus, no academic papers cover this specific digitalization initiative by the European Commission.¹

Moreover, few studies deal with the research question the thesis poses, which is concerned with the drivers of firms' participation in public initiatives designed to increase digitalization in the private sector.

Thus, in this chapter I will first present a review of the literature on the topic of innovation more broadly, presenting the main drivers of innovation in firms and also some of the barriers to the production and adoption of innovation. After that, I will move on to present the concept of Open Innovation Ecosystem, such as the EDIH Initiative, which we will cover more in-depth in Chapter 3.

1. The few studies that do exist on the matter are publications from the European Commission's Joint Research Centre, or DG CNECT itself. I am not going to present them here, as they will be treated in the chapter describing the EDIH initiative

2.1 Innovation, its drivers and barriers

The study presented in this thesis will be focused mainly on the process of digitalization. However, we can think of digitalization first and foremost as a particular type of innovation, and this is why I will start by focusing the attention on the process of innovation more broadly.²

The economic literature on innovation goes back to Schumpeter's seminal work, *The Theory of Economic Development (1934)*. In the book, innovation is defined as the introduction of a new product or production process (or a significant improvement of an existing one). Innovation is conceived as the driver of economic development, through the so-called process of "creative destruction".

In the 90 years that have passed since the publication of Schumpeter's book, the innovation landscape has changed. As a result of scientific progress, and the process of innovation itself, it has become more costly and more difficult to achieve breakthroughs. Bloom et al. (2020) argue that this is true in a vast number of fields of knowledge, noting that in recent years, increasing resources devoted to R&D did not translate to proportionally higher innovation output.

In this landscape, SMEs face much bigger challenges compared to larger firms. Since R&D is more costly and less productive than it used to, small firms are incentivized to not invest (or to invest less) on R&D. In fact, Crépon, Duguet, and Mairesse (1998) showed that the probability of engaging in research for a firm increases with size, market share and diversification for example. Thus, SMEs find themselves at a disadvantageous position: moreover, there is evidence that financial constraints play a role in determining firms' R&D decisions, as shown in Mancusi and Vezzulli (2014). In this study, the authors utilize both survey and accounting data from AIDA on firms in Italy. They find that need for external finance has a significant negative effect on the probability

2. Furthermore, let's not forget that the EDIHs are designed to support firms in Digitalization and Innovation initiatives. While the focus of this thesis is on Digital Maturity and Digitalization, some of the services provided by EDIHs, such as experimentation of new products or technologies, would fall under the broader category of innovation initiatives.

of investing in R&D, pointing to the role of credit constraints in limiting innovation in SMEs. One policy implication of these findings is that easing credit constraints to SMEs is of course, crucial to boosting their R&D investment.

This can be also one reason to argue for public subsidies to R&D in firms, especially in countries where Venture Capital is not particularly developed (e.g. European countries). Czarnitzki and Hussinger (2004) look at R&D subsidies in Germany and find that they significantly influence R&D expenditure (and also encourage patenting by firms). The study also rejects the possibility of crowding-out effects, which have been instead suggested by other works in the literature.³

Mina et al. (2021) analyzes the SME Instrument, a public initiative by the European Commission designed to help innovative SMEs. They find that although firms that are successful in obtaining the funds are growing faster than the average SME, they nonetheless present lower profit levels, indicating that they were in fact in need of external funding.

The picture painted by these studies on R&D investment and innovation is clear: firms that are financially constrained are less likely to invest in R&D, and public subsidies can help alleviate the problem, usually with limited crowding-out effects. We can extrapolate these findings to the digitalization process itself, since as we said, digitalization is a particular type of innovation. Thus, firm-level characteristics such as debt levels, profitability and size will be included in the main analysis of the thesis described in Chapter 5, in order to see if they drive firms' participation in the EDIH initiative.

2.2 Open Innovation Ecosystems

Another barrier to the process of digitalization is the problem of innovation adoption. This is particularly crucial for those technologies which take advantage of positive net-

3. A complementary relationship between privately and publicly financed R&D is indicated by most of the preceding literature; see for reference David, Hall, and Toole (2000). Successive literature has shown that crowding-out effects can occur in specific contexts, depending on firm characteristics as well as the economic environment

work effects (and thus necessitate of widespread adoption to achieve their full innovative potential), but is in general true for any innovative technology. In fact, on a macro level, the productivity-enhancing effect of new production technologies can be sizable only if those new methods become widespread (as discussed in Juhász, Squicciarini, and Voigtländer (2020)).

One recent model of the production of innovation that takes into account the importance of network effects while also addressing new challenges posed by the new innovation landscape is the Open Innovation model.

The seminal work on the Open Innovation concept comes from a book written by Henry William Chesbrough in 2003 (Chesbrough (2003)). In this book, Open Innovation is defined in contrast to the existing paradigm of Closed Innovation.

Chesbrough describes Closed Innovation as a process where:

1. Companies finance their own private R&D labs
2. These labs produce research leading to new intellectual property
3. The intellectual property is developed and capitalized by the company
4. Profits from the intellectual property are used to finance further research

It is clear how this paradigm represents the "classic" way innovation is funded and achieved in a capitalist economic system, as described also in the studies covered in the previous sections. The book however, argues that there are two major factors challenging this paradigm: the growing mobility of skilled professionals, and the growing presence of Venture Capital. These two factors contribute to the escape of intellectual property from the company's control, making the model of Closed Innovation less viable.

As a new paradigm designed to profit from these new characteristics of the innovation environment, the author proposes to adopt the model of Open Innovation, described as follows:

1. Firms should utilize both internal and external knowledge in their innovation processes

2. To achieve this, a more collaborative knowledge ecosystem is needed: firms and research institutions need to collaborate
3. Internal knowledge should be shared in the network, in order to provide for a faster rate of innovation that will in turn benefit also the firm itself

After publication of the book, subsequent research expanded open innovation, initially focused on large technology-intensive firms, to include the services sector, SMEs and Public Sector Organizations (PSOs). Open Innovation has gradually become central to many regional, national and supranational innovation strategies.

One paper that looks at the effectiveness of the Open Innovation model is Carmona-Lavado Antonio et al. (2021); the study employs fuzzy-set Qualitative Comparative Analysis (fsQCA) to investigate which configurations lead to high performance in innovation among biotech firms in Spain. The authors analyzed data collected through structured questionnaires filled by key figures in the companies, such as CEOs or R&D managers. The analysis finds that high-performing configurations are characterized by a high number of alliances, diversity of partners, investment in external R&D, coordination and learning capabilities and complementary organizational assets (e.g. human capital).

These experimental findings show one of the reasons why public intervention has been thought of as beneficial to Open Innovation: if you need to reach a high number of partners and achieve diversity in those partners, having public-private partnerships is critical.

In fact, public policy has aimed to encourage public-private partnerships, university-industry collaborations and the creation of innovation hubs. The EDIH network is one of these initiatives, that we can define as Open Innovation Ecosystems.⁴

Some other notable examples of Open Innovation ecosystems are:

- Manufacturing USA Institutes and the National Science Foundation's Innovation

4. The specific characteristics of EDIHs as Open Innovation Ecosystem will be covered more in-depth in the chapter on the EDIH Initiative

Corps in the United States of America

- Creative Economy Innovation Centers (CEICs) in South Korea
- Open Innovation Platform in Singapore
- Innovation Superclusters in Canada
- Digital Catapult in the United Kingdom
- Fraunhofer Institutes in Germany

The literature on Open Innovation Ecosystem initiatives is still young, since these are recent programmes, mostly launched in the 2010s and 2020s. However, we can already find a thorough review of the literature on Open Innovation initiatives in Idrissi Fakhreddine and Castonguay (2024). They analyze 99 articles about Open Innovation in SMEs. They find a significant gap in the literature, specifically regarding public policy initiatives to foster OI in SMEs, which is addressed only by 3% of studies in their sample.

This thesis will hopefully fill part of that gap, by providing insights into the effectiveness of one specific example of Open Innovation Ecosystem: the EDIH initiative, which I am now going to describe in the next chapter.

Chapter 3

The EDIH Initiative

The European Digital Innovation Hubs (EDIH) initiative, launched by the European Commission in 2021, represents a pivotal step towards enhancing digital transformation across Europe, particularly for small and medium-sized enterprises (SMEs). Established in response to the pressing need for increased digitalization within the EU, where only 20% of SMEs are considered highly digitized, EDIHs serve as one-stop shops designed to support businesses in the adoption of advanced digital technologies. This chapter will explore the origins of the EDIH network, delineate its main objectives and present descriptive statistics that illustrate the current landscape of EDIHs across Europe.

3.1 Before EDIHs: The DIH Initiative

The EDIH initiative emerged from a broader effort by the European Commission to accelerate digital transformation across the European Union, especially helping SMEs, which have been falling behind on digitalization metrics.

The roots of the program can be found in the *Digitising European Industry* document (European Commission and DG CNECT (2016)), starting implementation in April 2016, which for the first time created Digital Innovation Hubs (DIHs) across Europe.

DIHs were designed as regional cooperations with multiple participating entities, such as research and technology centres, universities, industry associations, chambers



Figure 3.1: Geographical Distribution by country of Digital Innovation Hubs in Europe as of January 2018, 2 years after the launch of the initiative. Source: *Digitising European Industry - Progress so far, 2 years after the launch* (European Commission and DG CNECT (2018))

of commerce, incubators/accelerators, regional development agencies and vocational training institutes.

These DIHs were precursors to the EDIHs later instituted. These hubs were designed as regional support centers assisting SMEs in digital transformation initiatives by providing access to expertise, test-before-invest opportunities, training, and networking. Their focus was local; as a consequence of this, they were typically financed through national or regional public funds. This of course led to having different levels of support from governments to DIHs across Europe.

As can be seen from the map, most of the DIHs were concentrated in central and western Europe. In fact, the country with the most DIHs is Germany, with 28 hubs, closely followed by the Netherlands with 25, France and Spain, both with 22 DIHs, and

then Belgium with 17. However, if we account for the difference in population sizes, we can see how one region (apart from the Benelux block) is over-represented in terms of DIHs: the Baltics. In fact, Estonia has 5 hubs, Latvia has 2 and Lithuania has 3. Considering they have a combined population of around 7.5 million people, their 10 hubs represented a much higher concentration of hubs than other countries in Europe. For example, Italy has 8 times the population of the Baltic states, while having only 1 additional DIH¹.

It's clear from the brief description provided before that DIHs were not sufficient to face the challenges of digitalization in the European private sector. Although effective at providing regional support, DIHs struggled to extend their services across regional boundaries, let alone the confines of individual member states. These shortcomings were the main reasons leading to the birth of the EDIH network, to better and improve the effectiveness of Europe's digitalization programme.

3.2 The birth of the EDIH Network

As part of the European Commission's "DIGITAL Europe Programme" (European Commission (2021)), it was decided to create a network of European Digital Innovation Hubs (EDIHs, for short). The main aim of the network was to overcome the shortcomings of the DIH Initiative. In fact, the main differences from DIHs to EDIHs are:

- DIHs' funding is at the regional or national level. In fact, funding for DIHs was idiosyncratic to the specific regional or national authority to which it was dependent on. On the other hand, the EDIHs' financing comes directly from the EU budget for digitalization, and constitutes 50% of the overall funding of each hub, with the remaining 50% coming from regional or national funds².

1. We will see in the next chapters how the high level of penetration of the DIH initiative in the baltic region will have a positive effect on the relative number of EDIHs coming from the region, as well as their productivity in terms of services provided to SMEs so far.

2. This could lead the reader to believe that the idiosyncratic element is still present even in the

- While DIHs were for the most part operating exclusively for one specific region in which they were headquartered, the EDIHs are designed to work as a network. This means that, in theory, one firm requesting support in Portugal could be directed by the Portuguese EDIH to a different hub in Spain for example, if the Portuguese EDIH does not have the expertise needed for that specific sector.
- Furthermore, single EDIHs can benefit from the network by sharing best practices, expertise and know-how in order to level regional disparities across Europe. One example can be that an EDIH located in Estonia that helps firms to digitalize could share best practices with hubs in Bulgaria for example, a country that has much less experience with digitalization of the private and public sector compared to Estonia.

Each EDIH would be structured as a consortium of different entities, such as private companies, Higher Education Institutions, Research Centres and Public Organizations³. It is possible to have EDIHs with member entities belonging to different Member States of the EU. Thus, the EDIH network is not only a network of hubs, but also a network of networks, with each hub acting as a node in the network.

3.3 Objectives and functioning of the EDIHs

As partially covered previously, the EDIH initiative aim to fulfil the following objectives⁴:

1. **Support the Twin Transition:** by providing services to support the digital and environmental transition of SMEs and PSOs.

EDIH network. However, the difference with the EDIH funding is that, even if coming from regional or national budgets, the rules and regulation in place are still set by the EU regulation on the EDIHs. Thus, the degree of standardization across EU Member States is much more advanced in the EDIH than it was for the former DIH initiative

3. Composition of the average EDIH will be covered in the Descriptive Statistics Section of this Chapter

4. As stated by the European Commission in the Digital Europe Programme (European Commission (2021) and European Commission, DG CNECT (2021))

2. **Act as Access Points to the whole network:** single EDIHs should direct firms to the most suitable hub for their needs, even if that hub is located in a different Member State. This objective aims at addressing the limitations of the DIH initiative.
3. **Networking and Knowledge Sharing:** EDIHs should share best practices, expertise and know-how amongst themselves, in order to level regional disparities across Europe.
4. **Brokering Role:** EDIHs may also act as brokers between public administrations and companies providing e-government technologies.
5. **Experimentation and Testing:** EDIHs should provide firms with the opportunity to test new technologies before investing in them, by making available testing facilities and expertise.
6. **Digital Capacities Multiplier:** EDIHs will act as a multiplier and widely diffuse the use of all the digital capacities built up under the different specific objectives of the Digital Europe Programme on High Performance Computing, Artificial Intelligence, Cybersecurity, Advanced Digital Skills and Accelerating the best use of technologies. To achieve this, EDIHs will participate in a "Train the Trainer programme", where specialists in these digital capacities will train the EDIHs on how to use them.
7. **Interface for European Commission Policies:** EDIHs will be the interface between the European Commission and the firms, providing feedback on the effectiveness of the policies and the needs of the firms in the specific sectors they operate in.

The allocation of the over 320 million euros budgeted for the EDIH initiative is done through a competitive process, where each EDIH (after having obtained the endorsement of one Member State) has to submit a proposal to the European Commission. The

proposals are then evaluated by a panel of experts, and the best ones are selected to receive funding, which is then disbursed in the form of grants, covering the functioning of the hub. More specifically, all the proposals that pass the evaluation are considered eligible for funding, and are then ranked and given funding according to availability. If a proposed EDIH is deemed eligible but no funding is available, it will receive the Seal of Excellence certification: it will then be a full member of the EDIH network, but the funding for its operation will have to come entirely from regional or national funds.

To evaluate the performance of individual EDIHs, the Commission will rely on a set of results indicators, such as:

- Number of SMEs and PSOs that accessed the EDIH's services
- Amount of additional investments successfully triggered by the EDIH's intervention
- Number of collaboration with other EDIHs and stakeholders outside the region
(this indicator addresses the network dimension of the EDIH initiative)

Furthermore, some additional impact indicators will be collected and analysed by the Digital Transformation Accelerator and the Joint Research Centre of the European Commission. These will be:

- The market maturity and market creation potential of innovations, as measured by the JRC's Innovation Radar methodology.
- The increase in the level of Digital Maturity of organization supported by the EDIHs, as measured by the Digital Maturity Assessment Tool (DMA) developed by the JRC⁵.

5. The design, structure and analysis of survey results from the DMA will be covered in the next chapter

3.4 Descriptive statistics on the EDIHs

As of today, the number of EDIHs operating across the EU's 27 member states is 151; however, when accounting also for Seal of Excellence Hubs, the number climbs up to 227 individual members of the network.

The map reported in Figure 3.2 shows the geographical distribution of EDIHs and Seal of Excellence Hubs across Europe. Addresses of the hubs were retrieved from the EDIH Catalogue, while the geocoding process was performed following the procedure that will be described in Chapter 5.

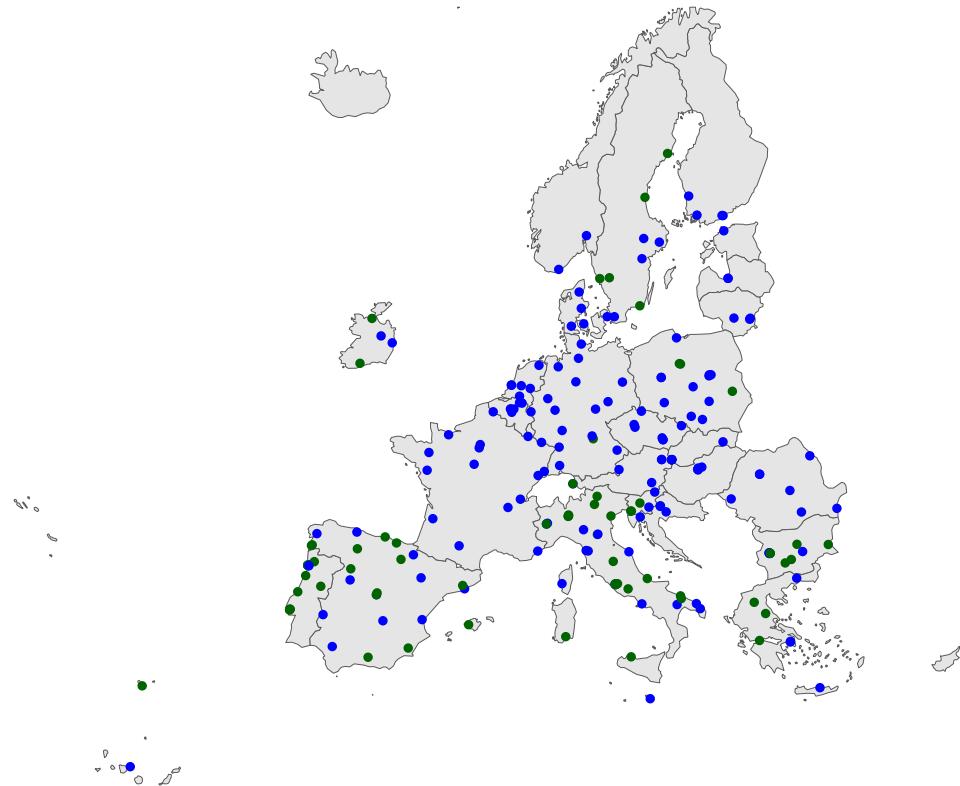


Figure 3.2: Geographical Distribution of European Digital Innovation Hubs (blue dots) and Seal of Excellence Hubs (green dots) in Europe as of March 2024. Source: Author's elaboration based on the geocoding of hubs in the EDIH Catalogue

The map shows a clear concentration of hubs in central and western Europe, with

a lower number of hubs in the eastern and southern regions. This is in line with the distribution of DIHs, as shown in Figure 3.1. The Baltic region, however, is still overrepresented, with 4 EDIHs and no Seal of Excellence Hubs. We will see in the next chapters how this will impact the number of firms in the sample that have interacted with the Baltic EDIHs.

It is easy to note that most of the Seal of Excellence hubs are located in the Southern and Eastern regions of the EU. This means that, while a lot of proposed hubs in these regions were deemed substantially worthy to be considered eligible, their proposals were considered less effective than those of the hubs that received funding. This could potentially have an impact on the effectiveness of the initiative in these particular regions, if the hubs do not manage to secure the necessary funding from national or regional budgets.

The JRC's report (De Nigris Sarah, Kalpaka Annita, and Nepelski Daniel (2023)) also provides some data on the current state of the EDIH network. On the regional distribution, it stresses the success of the reported regional coverage of EDIHs, stating that EDIH services are available in 223 regions, roughly 90% of the NUTS2 regions making up the EU. Furthermore, 213 regions host at least one organization that participates in the EDIH network.

Looking at the composition of the EDIHs from the table in Figure 3.3, we see that the average EDIH is composed of 11 participants, with 3 of them being Private Companies, 2 being Higher Education Institutions, 2 being Research Centres and another 2 being Public Organizations. An average of 3 participants per EDIH are labeled as "Other". This composition underlines a somewhat strong presence of the private sector in the EDIH network, which is in line with the objective of the initiative to foster public-private partnerships. The total number of participant entities in the EDIH network is 2355 unique organizations as of the writing of the JRC Report on the EDIH network.

The report also shows how the participating entities in the EDIHs primarily come from the NACE sector related to Professional, Scientific and Technical Activities, as well as Education and Information and Communication. This is consistent with the

Table 2.1. Number of participants by organisation type per EDIHs

	Minimum	Median	Average	Maximum
Number of participants	1	11	11	46
Number of Private Companies	0	3	4	24
Number of Higher Education Institutions	0	2	2	14
Number of Research Centres	0	2	2	16
Number of Public Organisations	0	2	2	10
Number of Other	0	3	3	16

Figure 3.3: Data on composition of the EDIHs. Source: De Nigris Sarah, Kalpana Annita, and Nepelski Daniel (2023)

focus of the program on upskilling and digitalization of firms.

As far as technological targeting goes, the most targeted technologies are AI (91% of EDIHs), Cybersecurity (70%) and Internet of Things (64%). This reflects well on the adherence of the EDIH network to the stated objectives of the Digital Europe Programme, since these are some of the technologies specifically cited as a focus of the programme itself.

The most targeted sectors by the EDIHs are instead Manufacturing (64%)⁶, Health Care (47%) and Public Sector (44%). Out of the services the EDIHs can provide, the most widely offered are SME support (81%), Ecosystem building (72%) and Technological Innovation (71%).

Note however that all these data are based on self-reporting by individual EDIHs at the start of their operation, so they should be taken with a certain amount of skepticism. Fortunately, by now we have reporting data on actual services that were performed by the EDIHs, so it is possible to directly check the veridicity of the self-reported data. The analysis on the sectors in which supported firms operate, as well as the technologies covered, will be partially covered in the next chapters. Here I will present only a quick rundown on the nature of services provided by the EDIHs so far, as well as a snapshot

6. This will be reflected in the DMA survey responses, as well as in the overall analysis on the drivers of participation in the EDIH initiative

of the sectors supported by EDIHs⁷.



Figure 3.4: Data on sectoral composition of the EDIHs customers (for the sake of readability, only sectors with more than 125 entries are reported).

Looking first at the sectoral snapshot of supported firms and PSOs, we can see that by far the most represented sector is Manufacturing and Processing, which constitutes more than 600 entries in the reporting dataset. Just after that, Personal Services comes as a striking second place, considering that it was never explicitly targeted by EDIHs. It is then followed closely by General Public Services (reflecting the presence of PSOs in the dataset), and then by Telecommunication, Construction and Assembly, Health Care, Biotech and Education. We can see that there is a partial overlap between the sectors most targeted by EDIHs and those where their customers operate, but the overlap is not perfect. For example, Biotech and Personal Services were not explicitly targeted by a large number of EDIHs (only 29% and 11% of EDIHs respectively), yet

7. The attentive reader will note that the number of reported services does not match the number of respondents to the DMA survey that will be reported in the next chapter. This is due to a variety of reasons, but the main one is related to the timing mismatch between the reporting of the services delivered by EDIHs and the submission of the survey to the DTA. It is often the case that the service gets delivered and then some weeks or even a month pass before the survey is loaded on the platform.

they are among the most represented sectors in the dataset.

In the next bar chart we can see some data on the type of services provided so far by EDIHs. The most common service reported is "Networking and access to innovation ecosystem"; this is the category that encompasses all services that help EDIH customers in connecting with potential partners, becoming part of innovation clusters, or organizing networking events. Seeing this service category as the most common can be a good sign, since this is the category that most benefits from the network dimension of the EDIH initiative. Next is the "Test before invest" category, which includes all types of prototyping or technology testing services that the EDIHs offer. "Training and skills development" is a close third, signaling the importance of upskilling in the digitalization process. Lastly, "Support to find investment" is the least popular service category, which could be a sign that the EDIHs are not yet fully exploiting their potential as brokers between firms and investors.

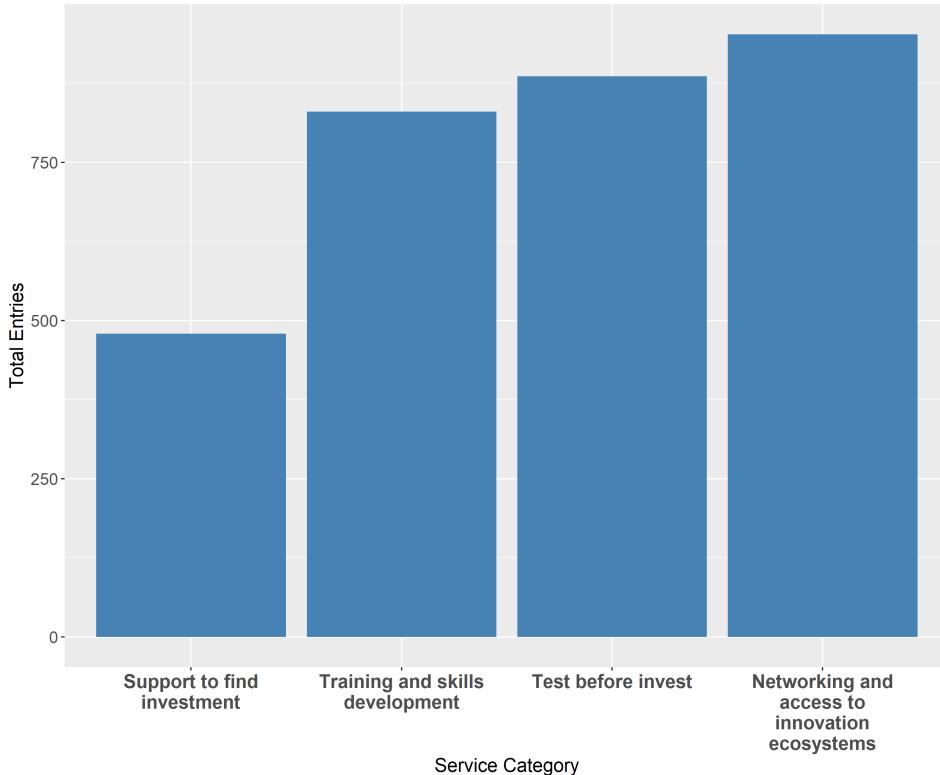


Figure 3.5: Data on the distribution of services provided by the EDIHs.

Chapter 4

The DMA Survey

While the EDIH network was being formed, the European Commission's DG CNECT decided it needed result and impact indicators to evaluate the performance of individual EDIHs, as well as the overall initiative.

Thus, in addition to result indicators to be reported by the individual EDIHs, DG CNECT delegated the JRC to design and validate an assessment questionnaire for measuring the digital maturity of EDIH customers (SMEs).

The validation was performed by testing the survey on a sample of SMEs around Europe and with the already active DIH community, precursor of the EDIH network.

In a second phase, an assessment survey was designed also for Public Sector Organization (PSOs); however, due to our focus on the part of the EDIH initiative that deals with private sector firm, I am not going to cover that version of the survey here.

I will now go over the design process and the overall structure of the survey, and then proceed to present some analysis of the survey responses obtained so far, from the sample of firms that I will then use for the main analysis.

4.1 Design and structure of the survey

The rationale behind the creation of a new DMA framework to be used by EDIHs is that measuring the level of digital maturity of a firm and its evolution over time will

provide a useful indication of the effectiveness of the intervention, and at an aggregated level, it will be possible to assess the effectiveness of the EDIH policy initiative.

Thus, the Digital Maturity Assessment Survey (DMA) was designed to investigate the base digital maturity level of the firm receiving support from the EDIH, before the intervention starts. The survey was then to be submitted two more times: the first one set for 1 year after the initial DMA submission, and the second one 2 years after the first one¹.

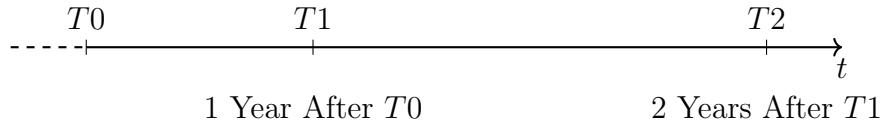


Figure 4.1: Timeline of DMA Submissions

The survey itself was organized in two main modules:

1. Module 1: Customer Data
2. Module 2: Digital Maturity for SMEs

The first module contained basic data to identify the firm, as well as some statistical filters, such as:

- Enterprise identification
- Contact information (Address, Email, etc.)²
- Staff size
- Foundation year

1. Specifically, according to De Nigris Sarah et al. (2023), we have the following definitions:

- T_0 = before EDIH intervention (a point in time no longer than 6 months prior to EDIH support)
 - T_1 = 1 year after T_0 (if EDIH intervention still running, wait until it is concluded. It should be applied not later than 3 months after EDIH support has been delivered.)
 - T_2 = 2 years after T_1 (a further point in time no earlier than 18 months and not longer than 24 months after EDIH support has been delivered)
2. Address information to geocode treated firms was taken from here

- Special SME categories
- Country/Region info (NUTS2, ZIP code)
- Sector of activity (NACE compatible)³

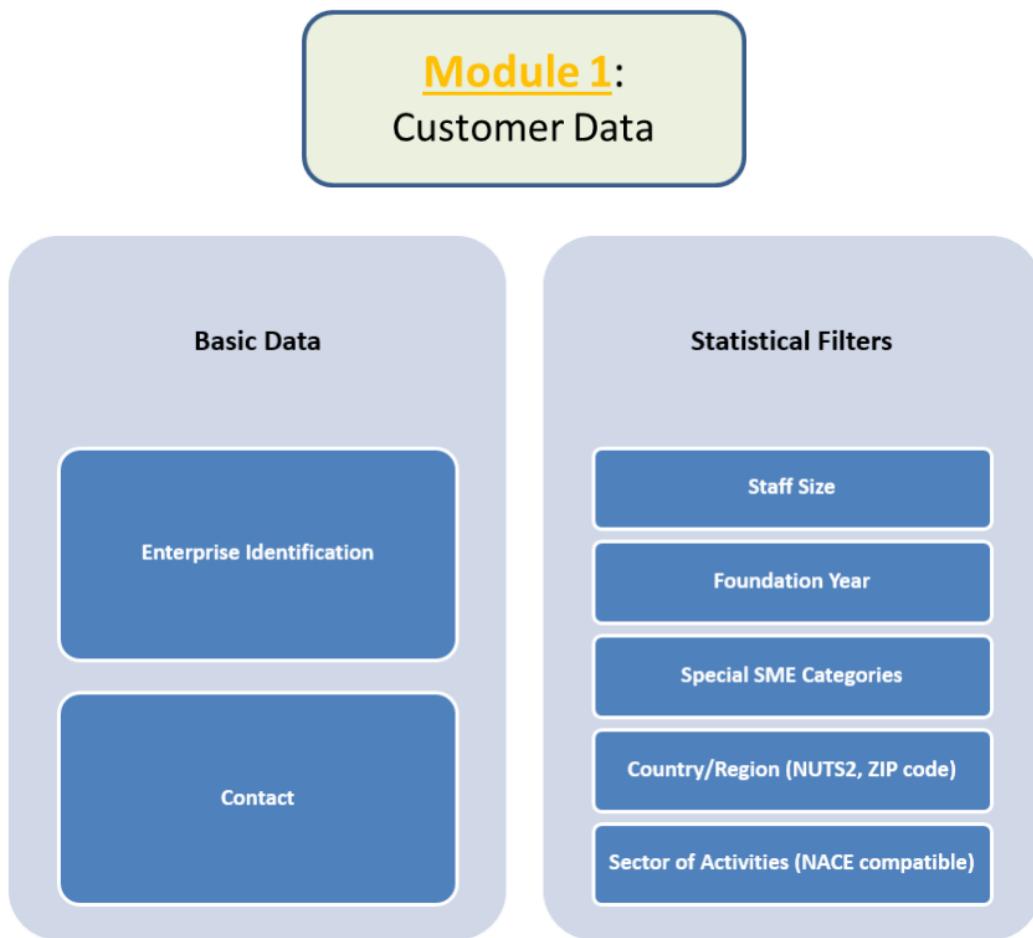


Figure 4.2: Structure of Module 1. Source: De Nigris Sarah et al. (2023)

The second module is the core part of the DMA questionnaire. It consists of questions assessing different aspects of the digital maturity of an organization, grouped under six dimensions, which for the SMEs' survey are:

3. The classification system itself is compatible with the NACE classification. However, for the purpose of easy compatibility with the ORBIS data and to more easily compare with the control group, I opted to use the NACE codes found in ORBIS instead of the classification designed by DG CNECT.

1. **Digital Business Strategy:** the questions in this dimension aim to assess the overall status of the enterprise's digitalization strategy. They inquire whether the firm invested in it across various business areas (both completed and planned) and the company's preparedness to undertake a digital transformation.
2. **Digital Readiness:** it measures the current uptake of digital technologies (both basic and advanced ones).
3. **Human - Centric Digitalization:** this dimension evaluates how staff are trained, involved, and empowered through digital technologies, and how their working conditions are enhanced to boost productivity and wellbeing.
4. **Data Management:** it captures if and how data is digitally stored and made accessible throughout the organisation, as well as exploited for business purposes, while being protected via cybersecurity schemes.
5. **Automation & Intelligence:** it explores the level of automation and AI employed in business processes.
6. **Green Digitalization:** measures the capacity of the firm to undertake digitalization with an approach that cares about the protection and sustainability of natural resources and the environment.

Each dimension is divided into two subdimensions (apart from Automation & Intelligence, which has only one), as you can see from the image below.

Each of the subdimension constitute one question in the survey, with several items contained in the question. The actual number of options in each question varies from one subdimension to the other. Each dimension is scored on a scale from 0 to 100, with higher scores indicating higher maturity. The overall DMA score is then computed as a simple average of the six dimensions, and is also on a scale from 0 to 100.

See below an example of a question from the DMA, specifically from the Advanced Digital Technologies subdimension in the Digital Readiness dimension. As you can see,

Module 2: Digital maturity

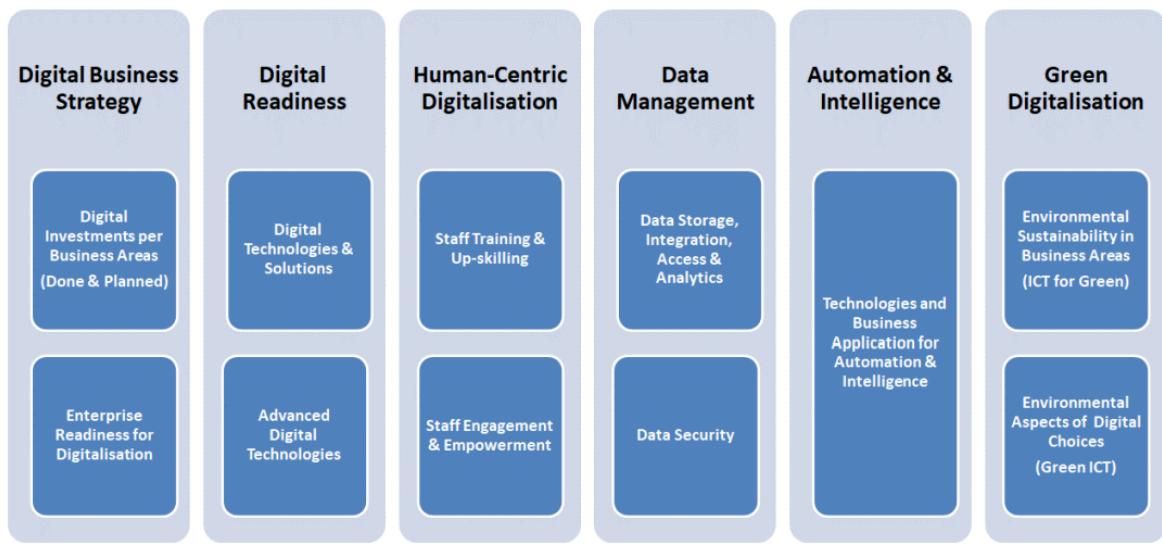


Figure 4.3: Structure of Module 2. Source: De Nigris Sarah et al. (2023)

for each advanced digital technology considered, the firms can select the level at which they use that specific technology, and get a score based on that.

As an incentive for firms to complete the survey in an accurate way the DMA, upon submission the website shows the results of the assessment and how the level of the firm's digital maturity compares with peers in their area or sector.

Q4. Which of the following advanced digital technologies are already used by your enterprise? Please choose all options that apply using the provided scale. (Options: Not used, Consider to use, Prototyping, Testing, Implementing, Operational)

	Not used	Consider to use	Prototyping	Testing	Implementing	Operational
Simulation & digital twins (i.e. real-time digital representations of physical objects/processes) *	<input type="radio"/>					
Virtual reality, augmented reality *	<input type="radio"/>					
Computer-aided design (CAD) & manufacturing (CAM) *	<input type="radio"/>					
Manufacturing execution systems *	<input type="radio"/>					
Internet of Things (IoT) and Industrial Internet of Things (I-IoT) *	<input type="radio"/>					
Blockchain technology *	<input type="radio"/>					
Additive manufacturing (e.g. 3D printers) *	<input type="radio"/>					

Figure 4.4: Example of a question from the DMA. Source: DMA website

4.2 Analysis of the survey responses

Let's start with some statistics on the number of firms involved. As of the time of writing, the sample of firms that submitted the DMA consists of about 3200 entities.

You can see a breakdown of the firms by country of origin the bar chart below.

It is immediately clear that the distribution of respondents is not uniform across countries. In fact, almost a majority of the respondents come from Lithuania, Latvia, or Spain. This is likely due to the fact that the DIH program was particularly successful in the Baltic countries (and in fact, even Estonia and Finland have a large representation in the sample compared to larger countries).

Anyway, this would mean that the sample of firms we are observing here is not in any way representative of the overall population of firms in the EU. With regards to the geographical dimension, I will account for this by adding the country of origin as

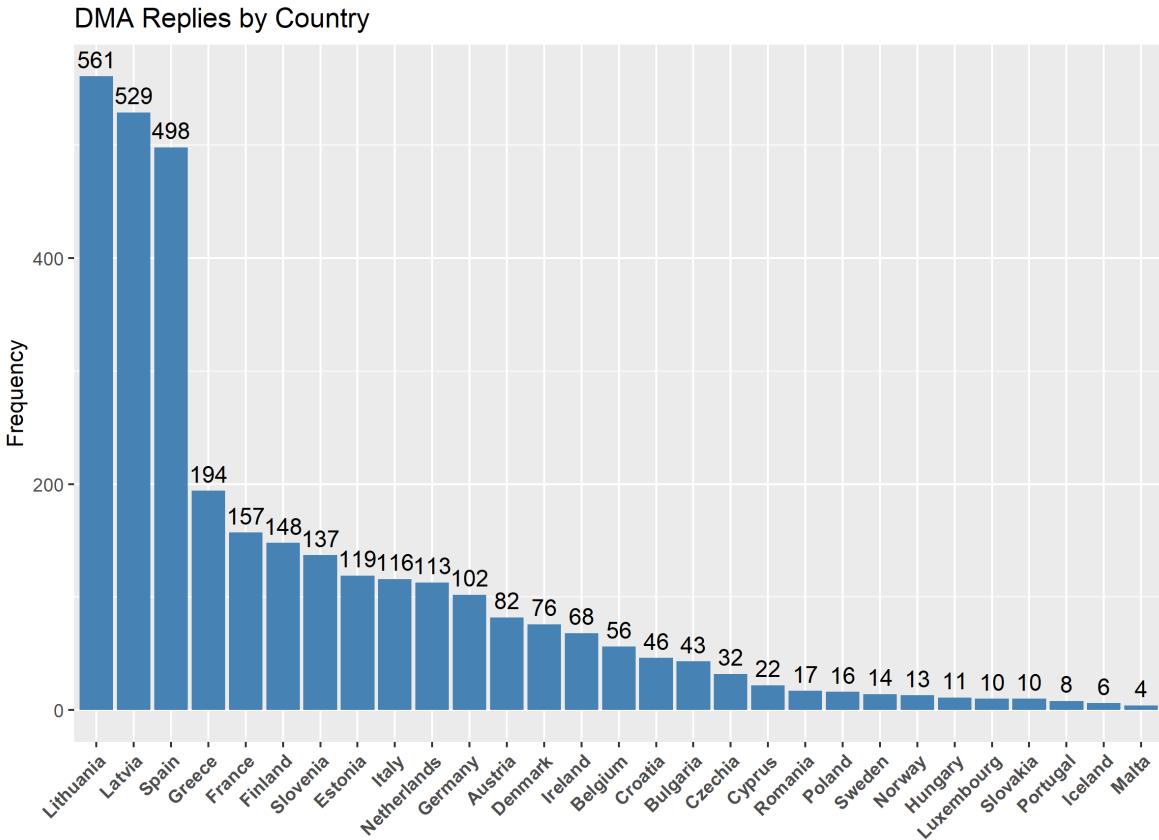


Figure 4.5: Distribution of DMA respondents by country of origin.

a control variable in the regression analysis⁴.

We can now proceed to analyze the breakdown by firm size. The first module of the DMA survey contains information on the size of firms, since it asks the respondents to declare their number of employees. This is done through a categorical variable, with the following options: Micro-size (1-9), Small-size (10-49), Medium-size (50-249), Small mid-cap (250-499), Mid-cap (500-2999) and Large company (3000+). Of course, the variable is self-reported, and thus it is possible that some respondents might have misreported their size. This is the reason why, once the matching with ORBIS data was done, I decided to use the number of employees as reported in the ORBIS database, to

4. The original idea was to utilize country-specific fixed effects. However, the sample size for some of the countries, once having accounted for the data loss related to the matching process with ORBIS, was not enough to allow for that. Thus, I decided to include group of countries as regressors in the analysis.

have a more accurate measure of firm size. Nonetheless, these responses allow us to get a first look at the composition of the sample.

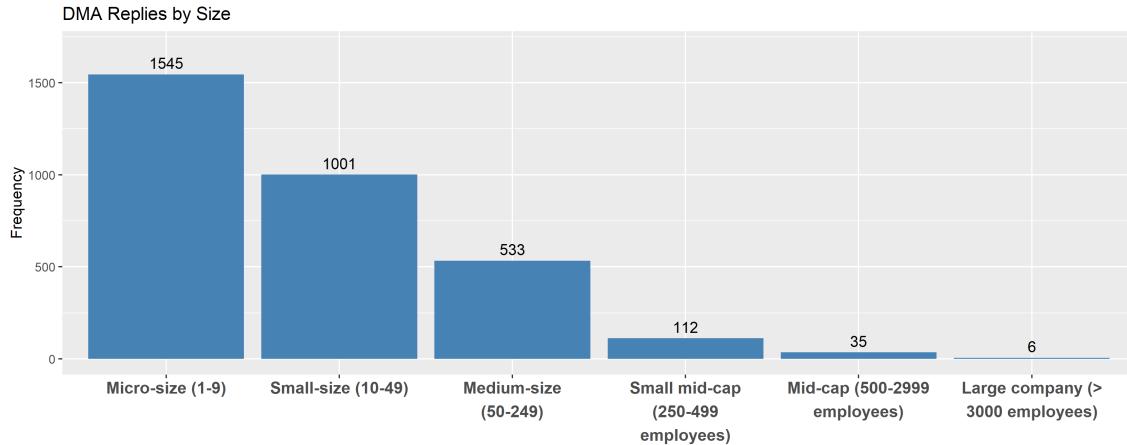


Figure 4.6: Distribution of DMA respondents by number of employees

As we would expect, we can notice right away how almost half of the firms in the sample are micro-sized, with a number of employees between 1 and 9. Another third is classified as small-sized, with 10 to 49 employees; medium-sized firms are also well represented. This seems to confirm the aim of the EDIH initiative, which is to target small and medium-sized enterprises. As a potential future research topic, since one of the objectives of the digitalization initiative is to make european firms more competitive and help them grow, it would be interesting to see if these firms were able to grow in size.

In the first module of the survey, firms also self-report their sectoral activity. This is done through a European Commission classification system, designed to be compatible with NACE classification.

As we can see, a plurality of the firms surveyed are part of the Manufacturing and Processing sector. Other sizeable portions of the sample are represented by sectors such as Telecommunication, Biotechnology, and Construction and Assembly. This is to be expected, since these sectors are already characterized by being highly technologically intensive ones; thus, it makes sense they would be among the firsts to take part in the

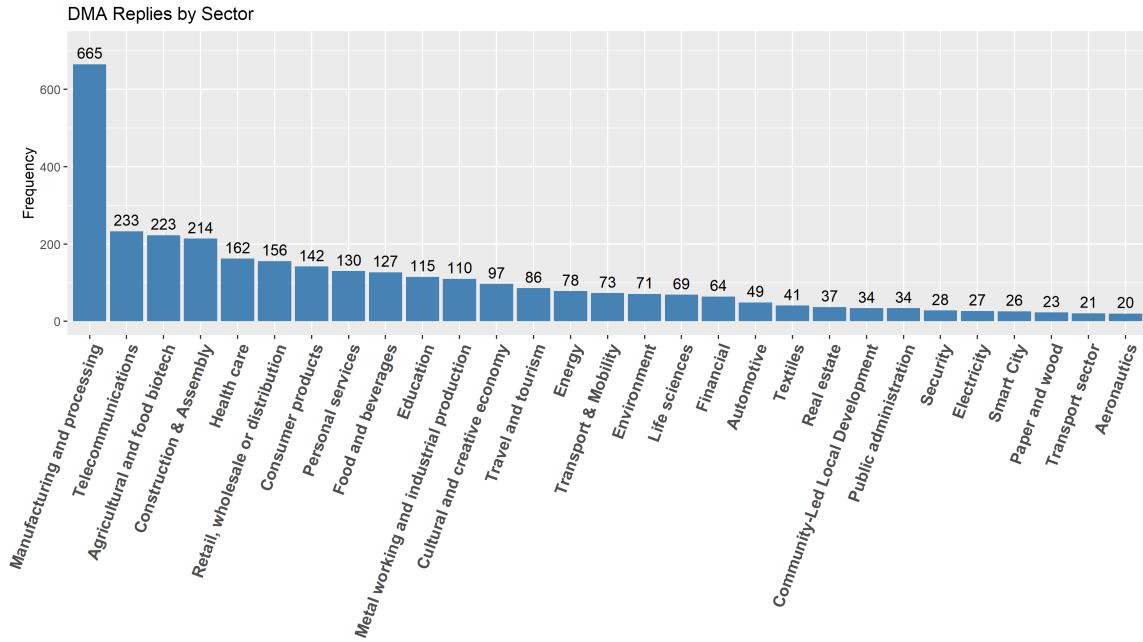


Figure 4.7: Distribution of DMA respondents by sector of activity

EDIH initiative.

Let's move to a quick overview of the replies from the DMA survey's questions. If we look at the overall score, the first thing we notice is the difference in average score and distribution between the different dimensions. We notice right away how the Automation & Intelligence dimension represents a significant drop in average score compared to all the other dimensions. In fact, we can see how more than 25% of firms present a score in that dimension that is actually equal to zero. Thus, even if the average score is around 20 points, the median is likely around 17. This suggests a serious deficiency in the uptake of innovative digital technologies, as the second-lowest average score is represented by the digital readiness dimension, which measures the uptake in specific technologies too.

We can now move over to an analysis of how scores are distributed across countries, sector, and size of the firm. Let's start with the country distribution. In the next figures you can see how the average DMA score and the average dimension score varies from country to country.

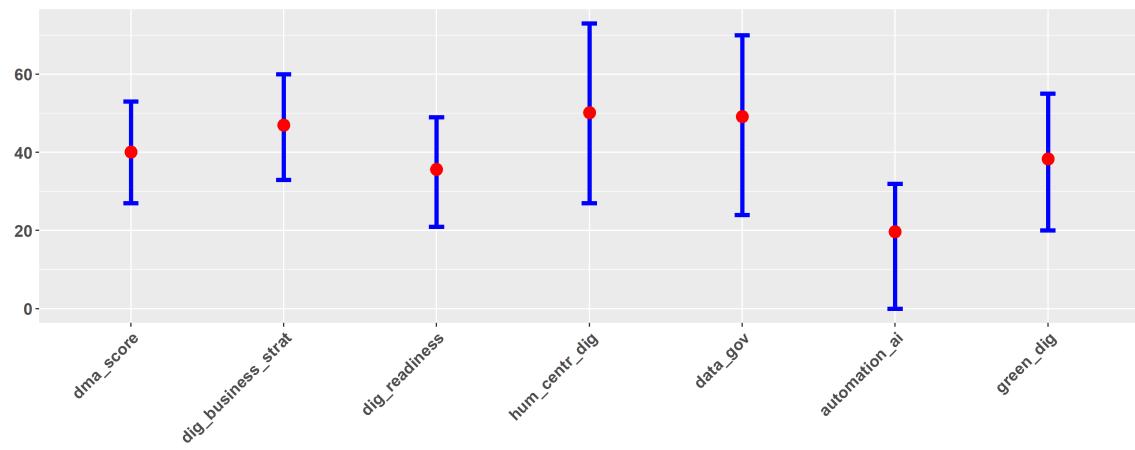


Figure 4.8: Average DMA dimension score and their 25-75 quantile intervals.

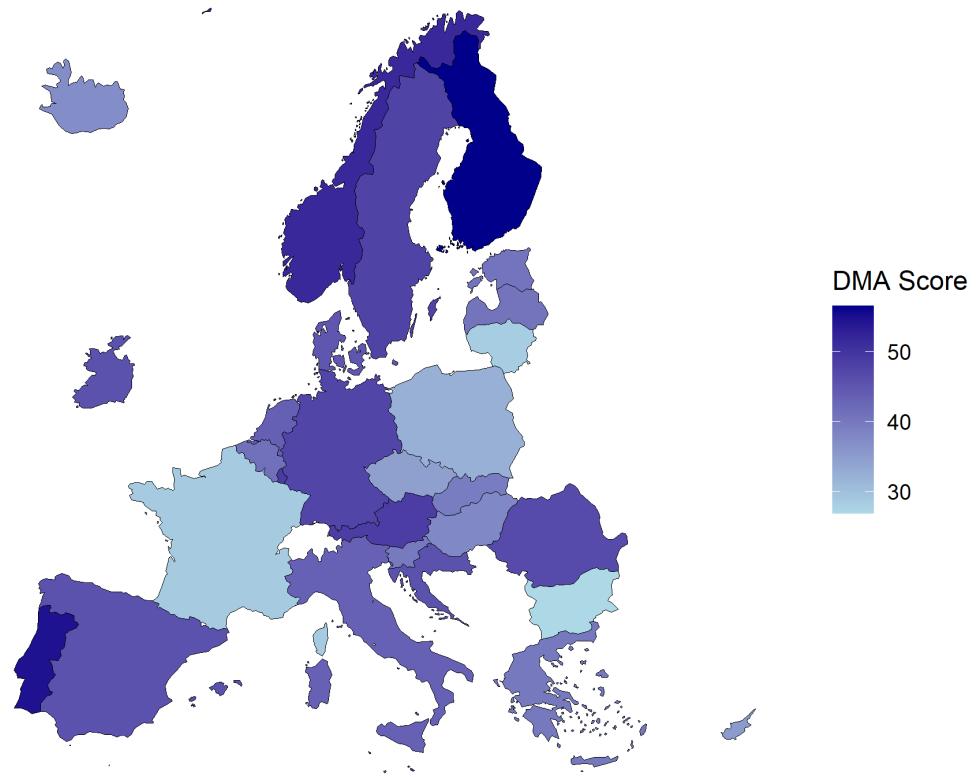
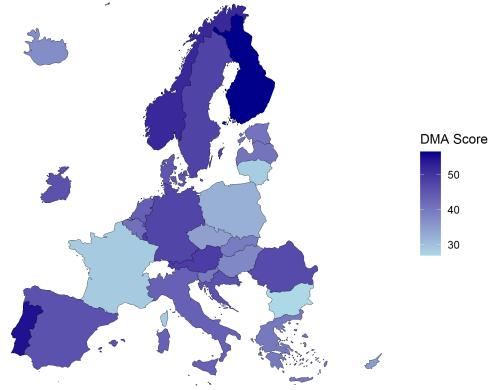
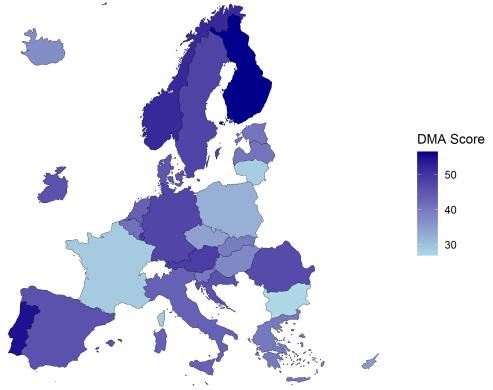


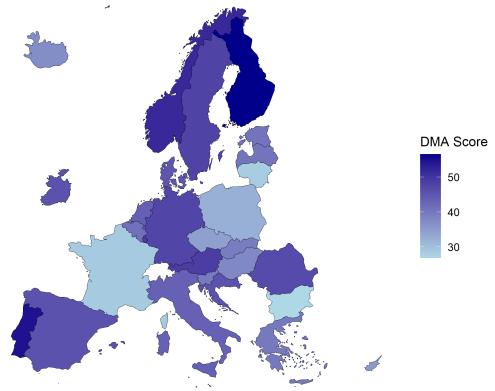
Figure 4.9: DMA score distribution by country



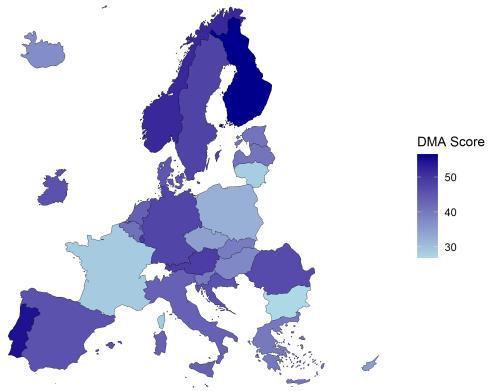
(a) Digital Business Strategy



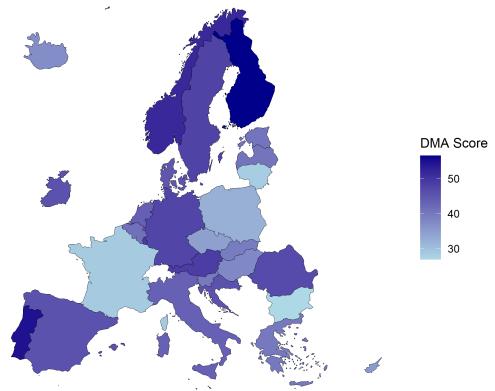
(b) Digital Readiness



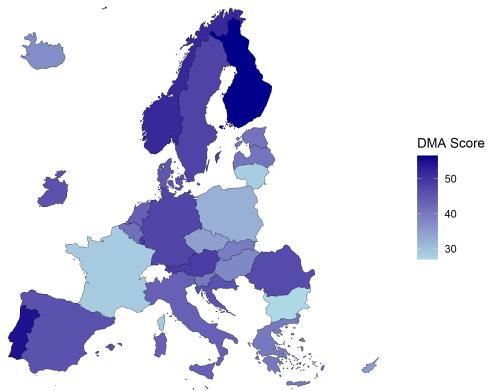
(c) Human-Centric Digitalization



(d) Data Management



(e) Automation & Intelligence



(f) Green Digitalization

Figure 4.10: Distribution of all dimensions' scores by country

Right away with the overall DMA score's geographical distribution we can notice how the score is significantly higher in the nordic countries. However, very good scores come also from southern countries such as Italy or Spain (Portugal is also very high, but the sample size is very low). Curiously, one of the worst-performing countries is France. We can see from the breakdown of components how this is not due to a specific dimension, as France performs badly in all of them, suggesting a lower level of digital maturity of the sampled firms in France.

In the Automation & Intelligence dimension, which is the one with lowest average score overall, we can see how the only countries having a slight edge are Austria and Finland, probably signalling a higher proportion of technologically intensive firms in these countries' sample.

If we analyze the distribution of scores by sector instead, we see how more technologically-intensive sectors have on average higher levels of Digital Maturity. In fact, the highest average DMA belongs to the Telecommunications sector, followed by the Environment sector, Transport & Mobility and Metal Working. Interestingly, a sector that we would expect to be exposed to digitalization presents instead a low DMA score: that is the case of the Financial services sector. Overall though, we see how the average DMA score is lower for those sectors that have had limited exposure to digitalization, as our prior knowledge would suggest.

Looking at the distribution of scores by size instead, we would expect to see a positive relationship between size and the average DMA score. In fact, this is exactly what we observe in our sample. This confirms the prior belief at the base of the EDIH initiative itself, being that the initiative should be focused on the digitalization of SMEs, since smaller firms are the ones struggling the most with the process. The lower levels of digital maturity observed in smaller firms can also explain at least part of the current gap in digitalization the EU is experiencing. Since firm size in most European member state is smaller on average than in other advanced economies (for example compared to the US), it is possible that part of the EU lag in adoption of digital technologies is due to insufficient firm size.

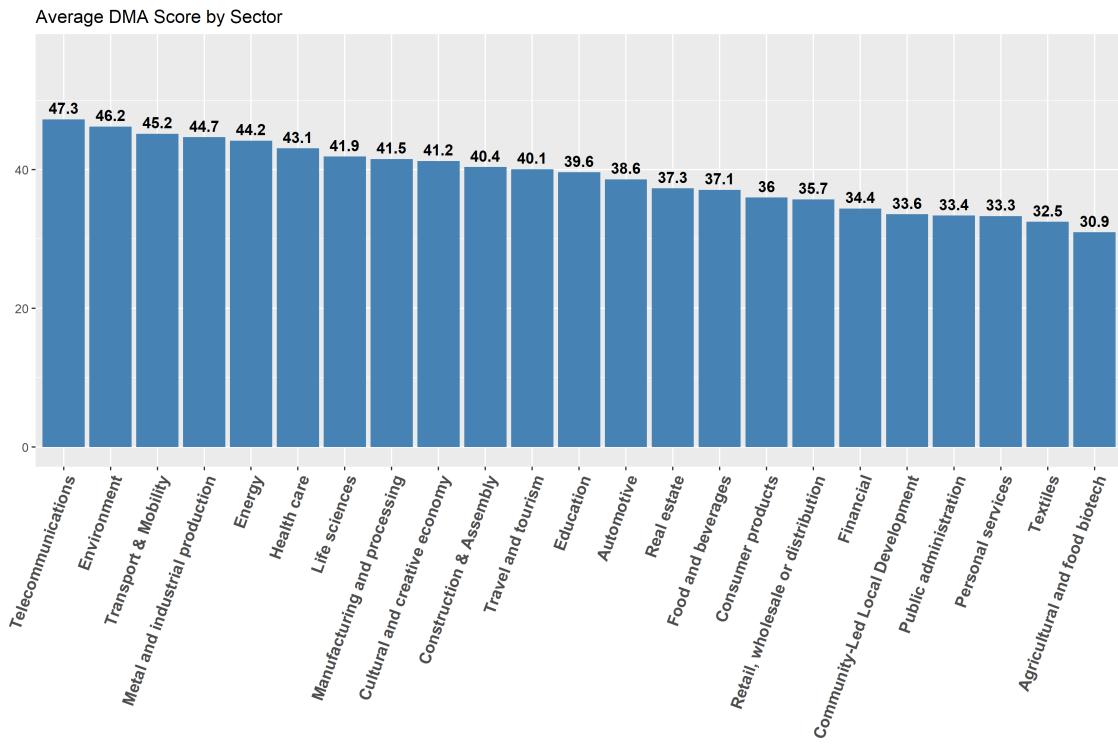


Figure 4.11: Distribution of the average DMA score across sectors of activity, as reported in the DMA, Module 1. Note that only sectors with more than 30 observations are considered.

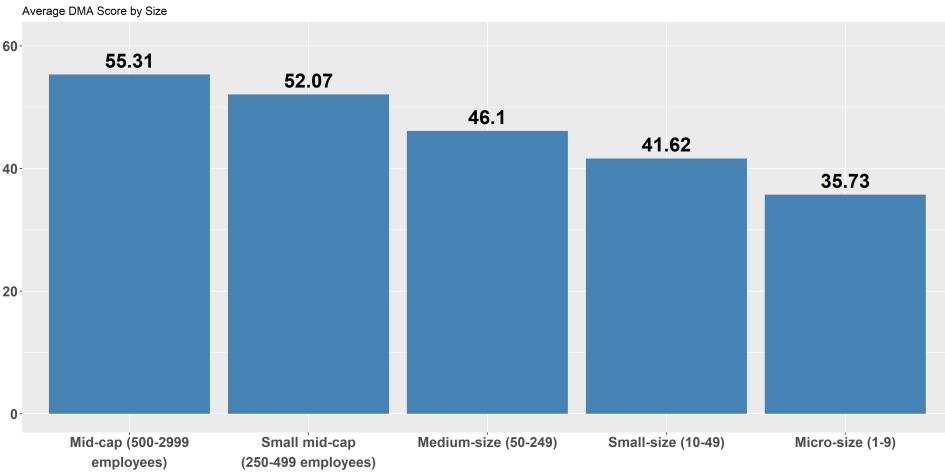


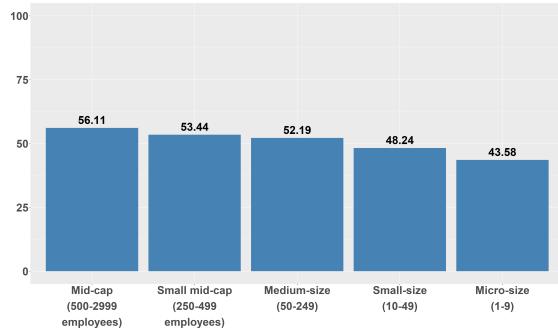
Figure 4.12: Distribution of the average DMA score across size categories.

As you can see in the breakdown by dimension below, the observation of a positive relationship between digital maturity score and size of the firm remains true for each of the six dimensions of the DMA survey. However, we can observe how the drop in average score differs from one dimension to the other. We can also appreciate how, for some of these dimensions, the drop is very steep in the smallest size category, while for others the big decline seems to come before. These observations can help identify points at which size reaches a critical mass and can lead to higher-on-average levels of digital maturity.

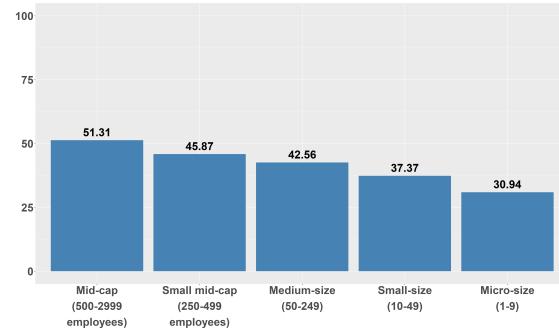
For example, if we look at the human-centric digitalization dimension, we see how there is not a huge difference in average score in the brackets going from 10 up to 249 employees, while micro-firms under 10 employees present an 8.5 point drop in average score. This suggests that the incredibly limited dimension of these firms does not allow them to invest in their staff, thus damaging the dimension's score.

Something similar can be said for the Automation & Intelligence dimension, although with a difference. In this case, every size category actually struggles with the adoption of this type of advanced technologies. The largest firms have an average score of only 35.66 in this dimension. This is probably due to the fact that most of the large firms in the EU are active in mature sectors of the industry. Thus, they are less innovative and present lower levels of advanced technology adoption with respect to counterparts belonging to newer sectors of the economy. However, here we see that the score is consistently low in the first two categories, suggesting that a certain critical mass is needed for this dimension to improve in a measurable way. If this is the case, we should not expect to see a large increase in this indicator for small and micro firms when looking at $T1$ observations, unless firms do actually grow very rapidly.

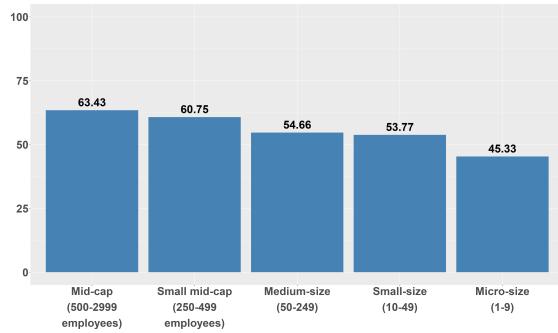
We can now move on to examine some insights obtained from the analysis of the firms' replies to the questions in the DMA survey. Specifically, one of the most interesting things we can look at is the current uptake in specific technologies. To this end, I will not present the responses to every question (you can find them in the Appendix); instead, I will present the responses from the questions belonging to the Digital



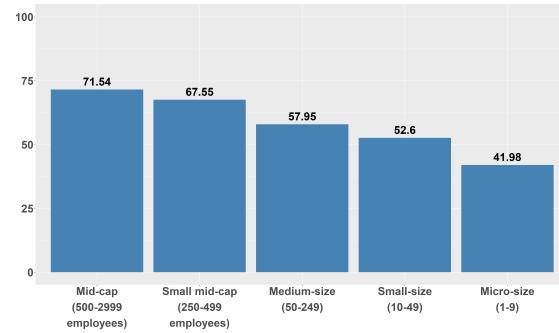
(a) Digital Business Strategy



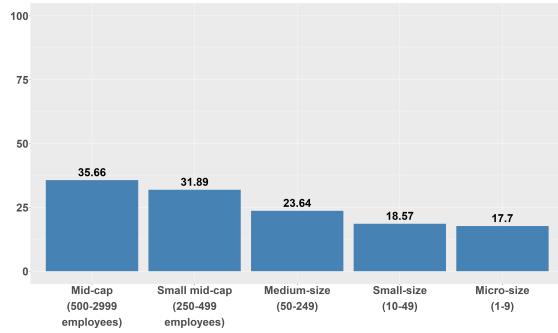
(b) Digital Readiness



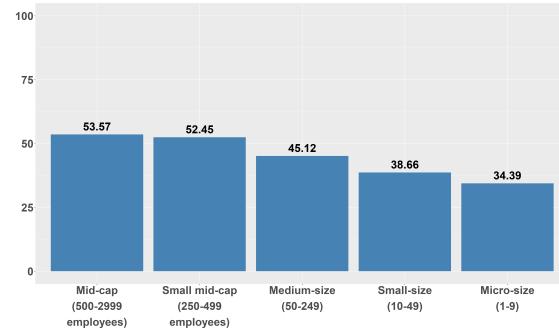
(c) Human-Centric Digitalization



(d) Data Management



(e) Automation & Intelligence



(f) Green Digitalization

Figure 4.13: Distribution of the average DMA score across size categories for each of the six DMA dimensions.

Readiness, Data Management, and Automation & Intelligence dimensions.

Starting with Digital Readiness, we can see the breakdown of survey responses from its first subdimension, related to Basic Digital Technologies, in the image below.

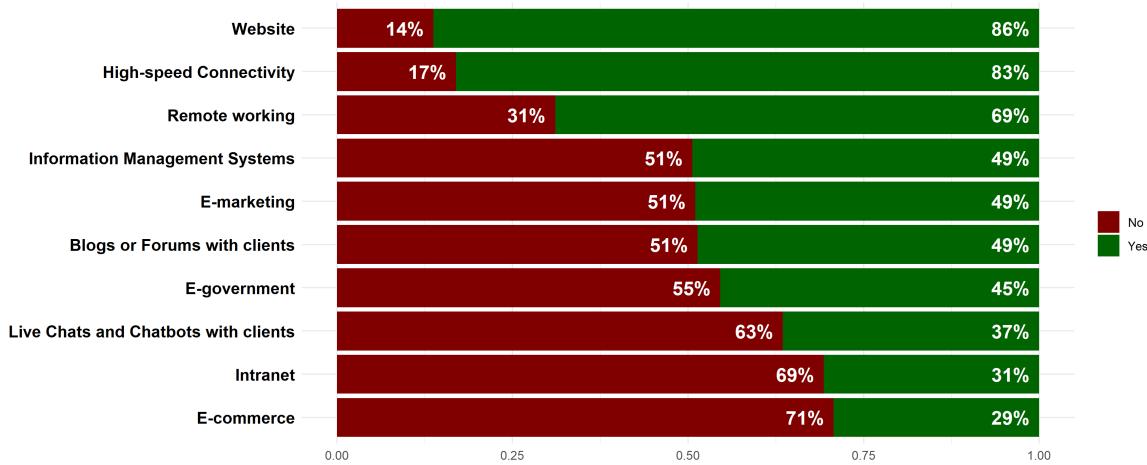


Figure 4.14: Breakdown of responses to Question 3 of the DMA survey, part of the Digital Readiness dimension.

Items are ordered from the most widely adopted to the least widely adopted. Right away, we can see how almost all of the surveyed firms have an online website, and high-speed internet connectivity. Almost 70% of firms report to utilize remote working for their operations: the significant amount of firms giving remote working as an option is probably due to the teleworking policies put in place during the Covid-19 pandemic. Another interesting item is Intranet; only 31% of firms report to have an intranet; however, this is probably due to the fact that most firms in the sample are small or even micro enterprises. Lastly, we can notice how few firms utilize e-commerce: however, this is probably a consequence of the sample being made mostly by manufacturing firms, and not by firms in the retail sector.

Let's move on to the subdimension on Advanced Digital Technologies. In this case, respondents are not restricted to a Yes/No response. In fact, they can choose the option that most accurately represents the level at which that technology is implemented in the firm. The levels from which respondents can choose are: Not Used, Consider to

use, Prototyping, Testing, Implementing, and Operational.

The added granularity of the information in this case can help us identify which technologies are deemed interesting, even if not yet operational. Also, this subdimension in the future can be useful to determine if the declared path on which the firm's technology adoption process was set has become reality or not.

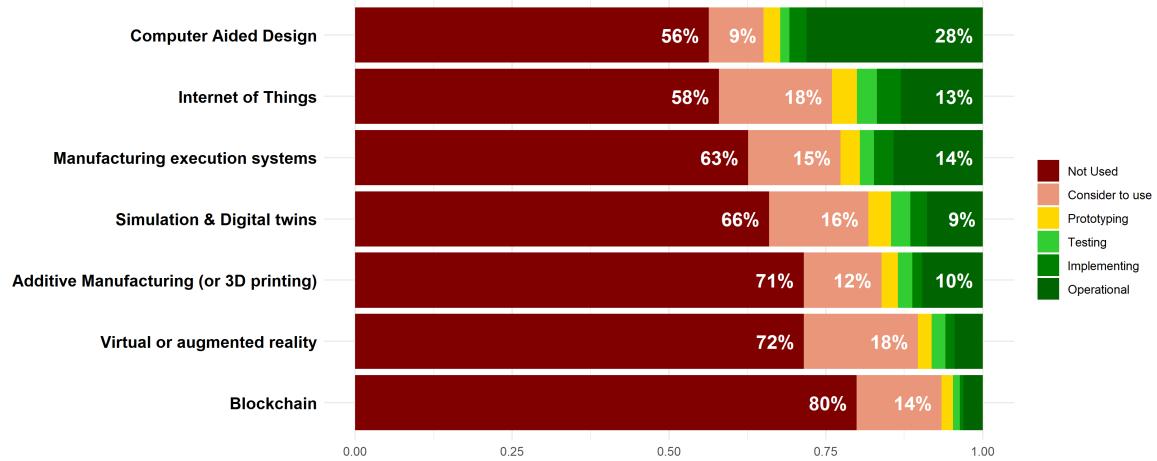


Figure 4.15: Breakdown of responses to Question 4 of the DMA survey, part of the Digital Readiness dimension.

What we can see for now in the breakdown of survey responses is how most of the advanced digital technologies surveyed are not used at all. Actually, none of the technologies presented would be adopted by a majority of firms even if all declared plans to adopt them were to succeed. This is probably an indication of how few firms actually appreciate the potential positive impact these technologies can have on their operations. This should be an aspect to be checked when *T1* observation start coming in relevant numbers. At that point, we would see if any firm has changed their views after the EDIH intervention. If for example we see that on average more firms are taking into consideration the idea of adopting these advanced digital technologies, then we would have indication of the program working as intended. Even more positive would be if we got data confirming that technology adoption plans are proceeding, thus getting higher proportions for the "Implementing" and "Operational" response options.

Let's now consider survey responses coming from the Data Management dimension. This dimension of the survey is important to understand how firms deal with internal data as well as data coming from their customers. The first question in the Data Management dimensions asks firms "How is your enterprise data managed (i.e. stored, organised, accessed and exploited)?"

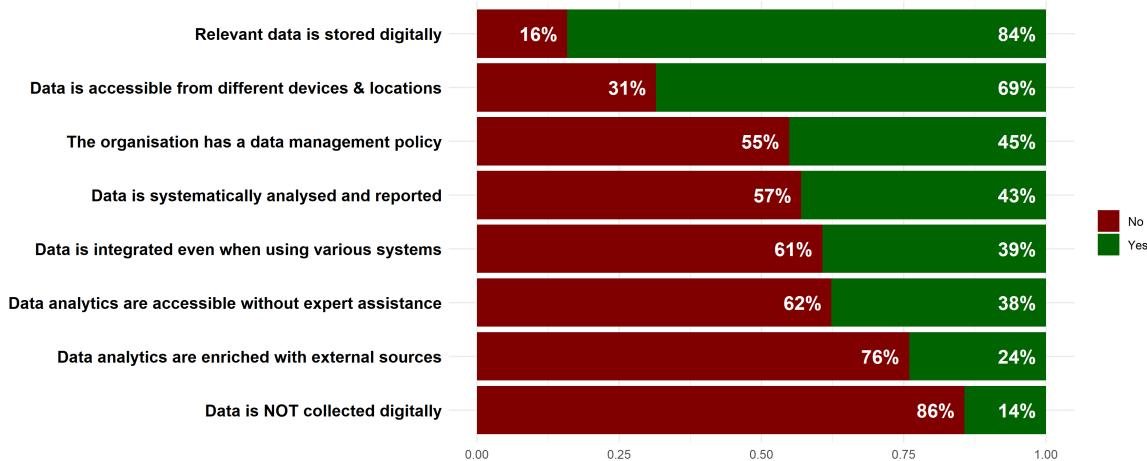


Figure 4.16: Breakdown of responses to Question 7 of the DMA survey, part of the Data Management dimension.

From the responses, we can see how basic data management seems to be in place in the vast majority of firms. In fact, 84% of respondents declare that they store digitally their data. However, this does not mean that the data is actually utilized. In fact, we can see that only 43% of firms in the sample systematically analyze their data, and only 38% make data analytics accessible without expert assistance (meaning they have data analytics dashboards). Even fewer firms enrich their data with external sources, making data analysis one of the most promising areas of improvement, since it appears that most firms have plenty of data, but they do not leverage it to increase their profitability.

The next question in the Data Management dimension deals with cybersecurity. As we can see, this is a point of particular weakness for firms in the sample. In fact, almost a third of firms respond saying that there is not even a full backup of critical data maintained. In less than half of firms, staff is trained on cybersecurity, and only

in a quarter of the firms surveyed there is a business continuity plan in case of cyber threats. Overall, these responses paint a grim picture of the state of cybersecurity for firms that replied to the DMA survey. This can be a consequence of a lack of literacy on the subject, or rather a reflection of the small size of most firms in the sample, making it infeasible for them to invest in cybersecurity measures. If the former explanation is the reason, we should expect the EDIH intervention to improve the responses to this specific question, but if the latter explanation is the real cause, it is unlikely we will see any substantial change come once T_1 observation are available.

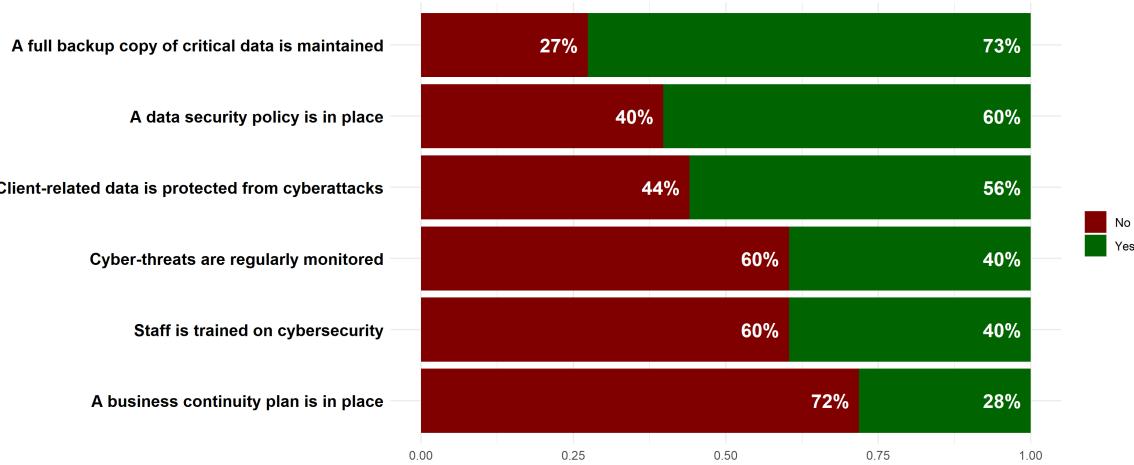


Figure 4.17: Breakdown of responses to Question 8 of the DMA survey, part of the Data Management dimension.

Now let's see how the firms responded to the only question present in the dimension concerning Automation & Intelligence. In this case, the question asked: "Which of the following technologies and business applications are your enterprise already using?". As per the question on Advanced Digital Technologies in the Digital Readiness dimension, here too there was the option to select different levels at which the implementation of the technologies was at the time of the survey's submission.

Thus, here too we can see which of these cutting-edge technologies our sample of firms finds the most interesting, by seeing which ones are considered the most, for example.

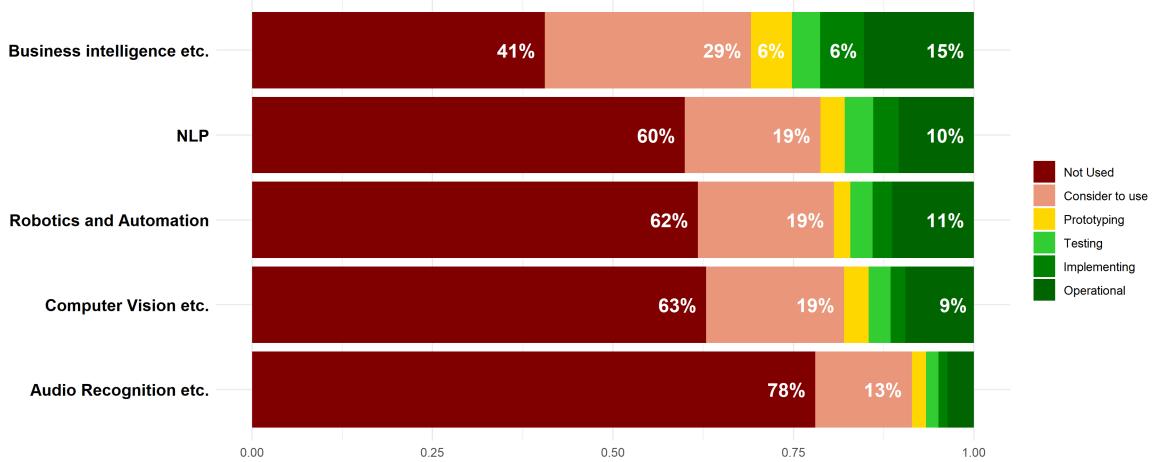


Figure 4.18: Breakdown of responses to Question 9 of the DMA survey, part of the Automation & AI dimension.

We can see right away that Business intelligence is by far the most sought-after technology for firms. In fact, almost 30% of firms are considering the use of this technology. Business intelligence refers to the process of using technology to collect analyze and transform data into actionable insights for the support of decision-making. Thus, it is reasonable that firms are interested in this, and it is actually also the most widely implemented technology.

Moving down, Natural Language Processing is the second-most considered technology, together with Robotics and Automation and Computer Vision, with 19% of firms considering to implement these technologies. Audio recognition is the least implemented by far and also the least interesting to firms, perhaps also due to a smaller set of applications in some sectors, particularly manufacturing.

As a last section in our analysis, it is useful to look at how the different dimensions in the DMA relate to each other. This would serve as a check on the usefulness of the DMA score as a composite indicator. In fact, one general rule of thumb when dealing with composite indicators is that all dimensions should be related with one another, at least to some degree⁵. The analysis of correlations between the different dimensions is

5. For more details on rules and criteria for the construction of Composite Indicator Frameworks, see: OECD and European Commission (2008)

going to also help us understand which ones are likely to be related, and can be helpful for future research on the subject of the DMA and the EDIH initiative more broadly.

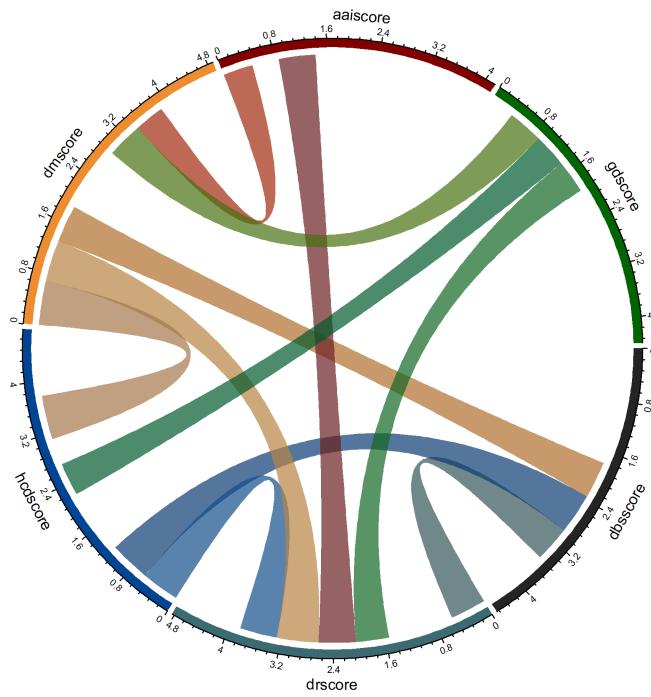


Figure 4.19: Circle graph of the six dimensions making up the DMA score and the correlation between them. Only correlations above 0.45 are shown in the graph; strength of the correlation is given by the thickness of the band connecting the two dimensions together. Only correlations above 0.3 are shown as lines.

Let's start with a high level view. In this plot you can see a circular representation of the six dimension that make up the overall DMA score. From this it is immediately apparent how some of the dimensions are more correlated with the whole compared to others. For example, digital readiness, human centric digitalization and data management are correlated with at least 0.45 correlation with most of the other dimensions in the DMA. Green digitalization and automation & intelligence instead stand out, due to their correlations with a small subset of dimensions in the DMA.

As you would expect, green digitalization is most correlated with human-centric digitalization, but more interestingly also relatively strongly with data management and

digital readiness. Automation & intelligence, on the other hand, appears to be correlated only with digital readiness and data management, the only other two dimensions where uptake in digital technologies is measured directly.

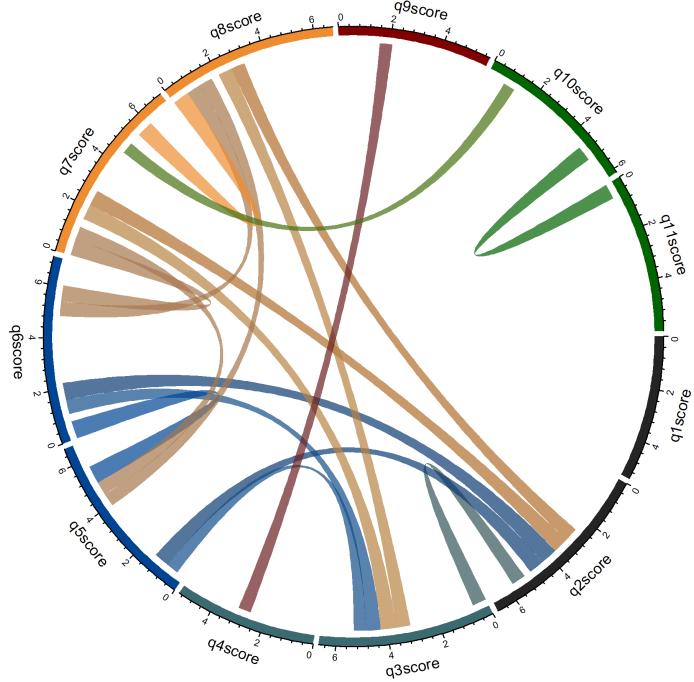


Figure 4.20: Circle graph of all the subdimensions making up the DMA score and the correlation between them. Only correlations above 0.45 are shown in the graph; strength of the correlation is given by the thickness of the band connecting the two dimensions together.

We can go one level down, to look at the correlation between subdimensions, meaning individual questions in the survey. In this case, we can see how the Automation & Intelligence dimension is actually correlated only with question 4, which deals with the uptake in Advanced Digital Technologies; furthermore, these two subdimensions are linked only between each other. This suggest that instead the Basic Digital Technologies' adoption measured in question 3 are not directly related with those technologies considered in the Advanced Digital Technologies subdimension or the Automation & Intelligence dimension as a whole. It is also interesting to see how, when disaggregating

green digitalisation into its two subdimensions, they are not linked anymore to the subdimensions making up human-centric digitalization. They are only correlated amongst themselves, and for question 10, which deals with Environmental Sustainability of business processes, with Question 7 of Data Management⁶.

We can still zoom a little bit further down, by looking at the correlation between the individual items that make up the DMA survey. This can give us useful information on which technologies are correlated the most, and which ones instead seem to be totally unrelated with each other. To do this, I created a network graph showing each individual item in the DMA survey, for a total of 96 points in the graph.

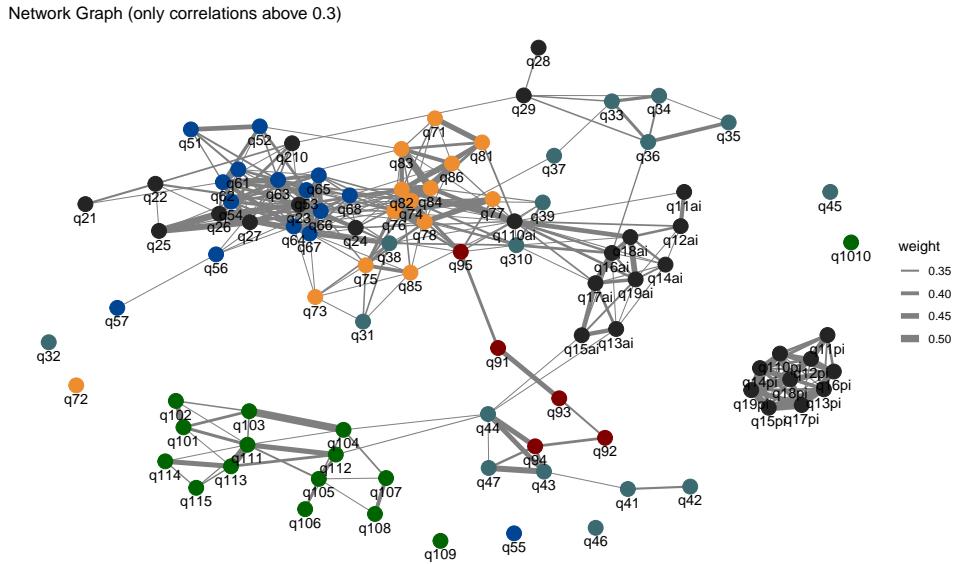


Figure 4.21: Network graph showing the correlation between every item of the DMA. Correlation computation are performed using the `hetcor` function in R.

Here we can notice most of the same patterns we identified in the circular correlation plots. For some dimensions in fact, most of the individual items are quite strongly correlated not just with each other, but also with items in other dimensions. This is the case for the human-centric digitalization and the data management dimension, as well as the second question in the digital business strategy dimension. Regarding the

6. This is probably due to the fact that both questions 10 and question 7 contain information on the digitized storage of data, one under the guise of paperless processes and the other while considering data management practices itself. We can see this more in detail by examining the network graph reported in the next paragraphs.

digital readiness dimension, and the automation & intelligence dimension instead, there are interesting details that ought to be explored.

We can see right away for example that items from question 3, related to Basic Digital Technologies, are much closer to the centre of mass of our network. For example, you can see how item q38 is correlated with q67. This is because item q38 measures the adoption of a teleworking digital platform, and q67 measures if flexible working arrangements are in place in the enterprise. In this case the correlation is due to the same underlying phenomena being investigated.

However, if you look at the correlation between q39 and q95, the relationship seems more interesting. In fact, q39 measures through a 0-1 response if the firm has in place any kind of organizational intranet, while q95 measures the degree of adoption of business intelligence (e.g. data analytics). In this case, the technology measured is not the same, so it is interesting that at least a correlation exists⁷.

Other examples can be pointed out, such as the constellation of correlated items from question 9 and question 4, made of q94, q44, q47 and q43. These are all items measuring the adoption of technologies related to manufacturing, so it is natural that they go hand-in-hand. You can see a similar example with the q33-q34-q35-q36 agglomeration. These are all items related to digital commerce and the digital interaction with customers in a broader sense, so it is reasonable that they move together to some degree.

Overall, these patterns of correlations between dimensions, subdimensions, and individual items provide at least some indication that the DMA survey is working properly and as intended, allowing us to utilize it as a tool for the evaluation of the EDIH initiative in the future, as well as providing a potential basis for the analysis of the adoption pattern of individual technologies.

7. Perhaps one potential explanation is that one technology is a stepstone for the other, but until we have more information with T_1 and eventually T_2 observations coming from the DMA, we cannot say. Once those are available, it will be possible to see if there is a pattern in the process through which firms "unlock" new technologies

Chapter 5

What are the drivers of participation?

As we have already discussed, the current state of the EDIH initiative doesn't allow for the analysis of the impact of the policy on firms' digital maturity. The reason leads back to the fact that the program itself is too recent (EDIH activities began in spring of 2023), thus we have no way of knowing what its impact has been, based on firm's financials or even based just on scores from future submissions of the DMA survey¹.

However, what we can do is investigate what are the driving factors that are leading firms to participate in the program itself. This can lead us to better understand if the targeting of the policy is working properly, or for example if there are some steps that can be taken in order to increase participation in the program itself. Furthermore, understanding the drivers of participation can also give us insights into the size and sign of the potential bias that can be present in the future analysis of the impact of the program on firms' digital maturity².

1. As of now, we have only a limited number of observations related to T1 issuings of the DMA survey, taken 1 year after the intervention of the EDIH. This initial data indicates a betterment of the Digital Maturity Score, at least in treated firms, even for those sections related to the adoption of specific advanced technologies. However, since the DMA was never submitted to a control group of firms not participating in the program, an analysis of this type would not be able to establish causality.

2. For instance, if the analysis were to suggest that one of the positive drivers of participation is firm size, we can hypothesize that when measuring the impact of the policy only on the treated firms, the bias on the effect of digital maturity will be positive, since we can reasonably assume that larger

To perform the analysis, I have matched the firms that participated in the EDIH initiative with the ORBIS database to retrieve firm-level financial information (such as their size, country, sector in which they operate, and indicators of financial performance).

Once having had the data on treated firms, I have constructed a synthetic control group of untreated firms. These control firms will be selected randomly from the ORBIS database. After constructing the control group, first of all I've compared the treated and control firms to identify any significant differences in their characteristics. This comparison is instrumental to understand the factors that may influence a firm's decision to participate in the EDIH initiative.

As an additional information to add on top of country information on where the firms are located, I decided to assess the effect of the distance of firms from the closest EDIH hub. The working hypothesis is that firms that are sited closer to an EDIH hub are more likely to have heard about the program and thus to participate in it. Knowing how distance affects firms' participation can be instrumental in understanding if an expansion in the number of hubs or in their geographical coverage can lead to an increase in participation. Also, to account for the possibility of the effect being confused by areas with a large concentration of firms (which are more likely to have an EDIH hub nearby, and thus more likely to participate), I have computed a measure of firm density. To compute both density and distance, I geocoded the firms' addresses, taken from the ORBIS database or from the DMA survey itself.

Finally, I have run a probit regression model to analyze the drivers of participation in the EDIH initiative. The regression model includes various firm-level characteristics as independent variables, and the dependent variable is a dummy variable indicating whether the firm participated in the EDIH initiative.

In the following sections, I am going to describe the matching process, the construction of the control group, the treated-control comparison, the geocoding process, and

firms can benefit more easily from digitalization initiatives. Conversely, if smaller firms are more likely to participate, the bias might be negative.

the results of the regression analysis in detail, as well as some robustness checks on those results.

5.1 Matching of treated firms with ORBIS

To safely match the treated firms with the ORBIS database, I have used the data from the DMA survey replies. Every firm, when submitting the DMA to the Digital Transformation Accelerator platform, has to provide basic company information. Among other things, the database includes firms' legal names, as well as VAT numbers.

These information can be used to match companies with data in ORBIS. However, the DMA tool does not automatically check for invalid VAT numbers, thus some of the reported VAT codes are incorrect. Furthermore, I was relying on WRDS as the website to download the ORBIS database, and WRDS allows for downloading big chunks of data, such as all firms in Europe (or filtering by country, in each European Member State), but only for some specific variables. Among the variables that can be downloaded, there is no VAT code. Thus, I had to rely on company names to match the firms. This is not ideal, since it is possible to have misspellings or different ways of writing the same company name.

Thus, to match the treated firms i proceeded in the following way:

1. I downloaded the whole universe of firms in ORBIS country by country (for 18 EU Member States where the number of firms participating in the EDIH initiative was the highest), saving them as `Orbis-COUNTRYCODE-Universe.csv` files.
2. To have a more efficient code, I loaded the files into memory one country at a time and then, for every country in my DMA database, I computed the Jaro-Winkler distance³ between that firm's name and that of every company in that country's ORBIS universe, substituting the match every time a smaller distance was found until a threshold of 0.01 was reached⁴. After having done this for every
3. The Jaro-Winkler distance is defined as being equal to 1 minus the Jaro-Winkler similarity:

$$dist(JW) = 1 - sim(JW)$$

4. The threshold of 0.01 was chosen after some trial and error, to minimize the number of false positives while still speeding up the matching process.

country's ORBIS universe, I saved the results in a .csv file, containing the matched DMA company name, the ORBIS company name, and the corresponding BvD ID number as well as the VAT code from the DMA database.

3. This allowed me to then download the financial data for the matched firms by searching them in ORBIS through the BvD ID number. I then checked if the VAT code corresponded to the one in the DMA database, by again computing the Jaro-Winkler distance between the 2 strings.⁵ If the distance was above the threshold, I inspected other variables to check if the match was correct and the VAT code was indeed incorrectly specified.

As I was mentioning in the previous chapter on the DMA survey, at the time the analysis was performed the firms in the database were 3204, out of which I was able to match through this procedure 2886 of them. Thus, the percentage of matched firms through this procedure was roughly 90%. The remaining 10% of firms were not matched for a number of potential reasons, that can have different implications on the analysis I am going to perform. Potential reasons include:

- The firm was not present in the ORBIS database, either due to the firm being too small, too young, or perhaps being located in a country where ORBIS coverage is not perfect.
- The firm was present in the ORBIS database, but the name was misspelled or written in a different way, and the Jaro-Winkler distance was too high to match the firm, and instead it matched to a completely different one. This can be the case in particular for firms with shorter names, or sometimes cited with their acronyms.

I will discuss this more once I present the design of the control group, but for now it is important to note that if the second case is the one driving the majority of the

5. This was done to make sure that simple misspellings in the input of the VAT code would not cause the matching to be negative.

non-matches, since having shorter names or acronyms is not likely to be significantly correlated with any other firm-level characteristics, the bias introduced by the non-matches is likely to be small and negligible.

However, if the non-matched firms are mostly belonging to the first case, then we have to keep in mind that it is possible that the matched firms are systematically larger in size, possibly older, and located in country with better ORBIS coverage with respect to the overall sample of firms that participated to the EDIH initiative.

5.2 Creating the Control Group

When designing the control group, one has to consider the aim of the analysis it is going to be performed with it.

For example, if the aim is to estimate the impact of the EDIH initiative on firms' digital maturity, the control group should be constructed in a way that it is as similar as possible to the treated group, except for the fact that it did not participate in the program. This is because the aim of the analysis is to estimate the counterfactual: what would have happened to the treated firms if they had not participated in the program.

The case of my analysis however, is a bit different. Since we are trying to estimate the drivers of participation in the program, the control group should be as "agnostic" as possible, the reason being that the best control group for our analysis would be the one that is most representative of the population of firms that could have participated in the program (so, in theory, the whole population of firms in the EU).

Since I cannot have access to the whole population of firms, but only to a subset of them who are reported in the ORBIS database, the best control group is the one that is most representative of the population of firms that are in the ORBIS database. This choice of course has implications, since ORBIS coverage is not the same across all countries, and it is not independent of firm-level characteristics.

However, since the treated firms included in the analysis are also only the ones that are present in ORBIS, the bias introduced is likely to be attenuated, since they will

share the same selection bias.

It should also be noted that if the characteristics leading to better coverage in ORBIS are also the ones leading to participation in the EDIH initiative, the bias introduced by the control group can be positive.

All this taken into account, I have constructed the control group in the following way: I have taken all the firms in the ORBIS database for the 18 selected countries, and picked random BvD IDs with the only constraint that the proportion of firms from a specific country in the synthetic control group should be the same as the proportion of firms from that country in the overall ORBIS database. This was done to ensure that the control group is representative of the population of firms in the ORBIS database.⁶

After running this procedure, I had a list of 100000 firms, for which I then downloaded firm-level characteristics from the ORBIS database.

After doing this, everything was ready to proceed with the initial comparison between treated and controls.

5.3 Treated - Control Comparison

In the table below, I present the results of an initial comparison between treated and control firms. The table shows the mean and median values of the main firm-level characteristics for the two groups, as well as the results of a t-test for the equality of means between the two groups. The t-test is performed for each variable, and the p-value is reported in the last column.

What is immediately noticeable from the first line, where the number of observations for each group is reported, is that both groups shrank in number. This is because not all firms in the ORBIS database have complete financial information, and thus some of them were dropped from the analysis.

To limit this issue, I restricted the analysis to variables that had at most 25% of NAs

6. Representativeness with respect to other firm-level characteristics was not considered as a constraint. However, the large number of the sampled firms making up the control group should allow for representativeness due to the law of large numbers.

Statistic	Treated_0	Treated_1	p_value
N	3643.00	1249.00	
Mean_Employees	0.03	0.09	0.00
SD_Employees	0.17	0.37	
Mean_TotalAssets	14.09	15.18	0.80
SD_TotalAssets	185.97	108.15	
Mean_LiquidityRatio	4.15	2.33	0.00
SD_LiquidityRatio	8.41	2.90	
Mean_Solvency	37.65	50.51	0.00
SD_Solvency	35.45	25.62	
Mean_turnover	9.15	16.02	0.00
SD_turnover	70.39	74.00	

Table 5.1: Comparison between Control Firms (Treated = 0) and Treated ones (Treated = 1), with p-values for the t-test of equal means

observations in each of the two datasets. However, since the observations were retained for the analysis only if all variables had non-NA values, the drop in observations was still significant. The number of treated firms dropped from 2886 to 1249, while the number of control firms dropped from 100000 to 3643 firms⁷.

Moving on to the initial comparison of means in the 2 groups, we can see how for the variables showed in the table the means are significantly different. This is confirmed by the t-test, which reports a p-value of 0 for all variables except for total assets, where the p-value is 0.80, probably to the relatively high standard deviation making it difficult to detect a significant difference in means.

The comparison highlights something that could be already expected: the firms participating in the EDIH initiative differ significantly from the overall population of firms in ORBIS. However, the sign is what is most interesting: treated firms seem to be having more employees, a better value of the liquidity ratio, higher solvency ratios and turnover (while for total assets the difference from the simple t-test is not significant).

We can investigate a little bit further by comparing the distributions of the variables in the treated and control group, variable by variable. this is done in the following

7. It should be noted here that the very high sample size for the control group was chosen specifically to make sure that after the drop in observations, the size of the database was still considerably large

figures. There you can see how the treated firms' frequency distribution (in blue) compares to the control group (in red), as well as a kernel density function estimating the distribution of the variable based on the data. In every graph, above you can also see reported the p-value of the KS-test, which tests the null hypothesis that the two distributions are the same. Since the KS-test is a non-parametric test, it allows us to confirm that also in the case of total assets, the two distributions are significantly different.

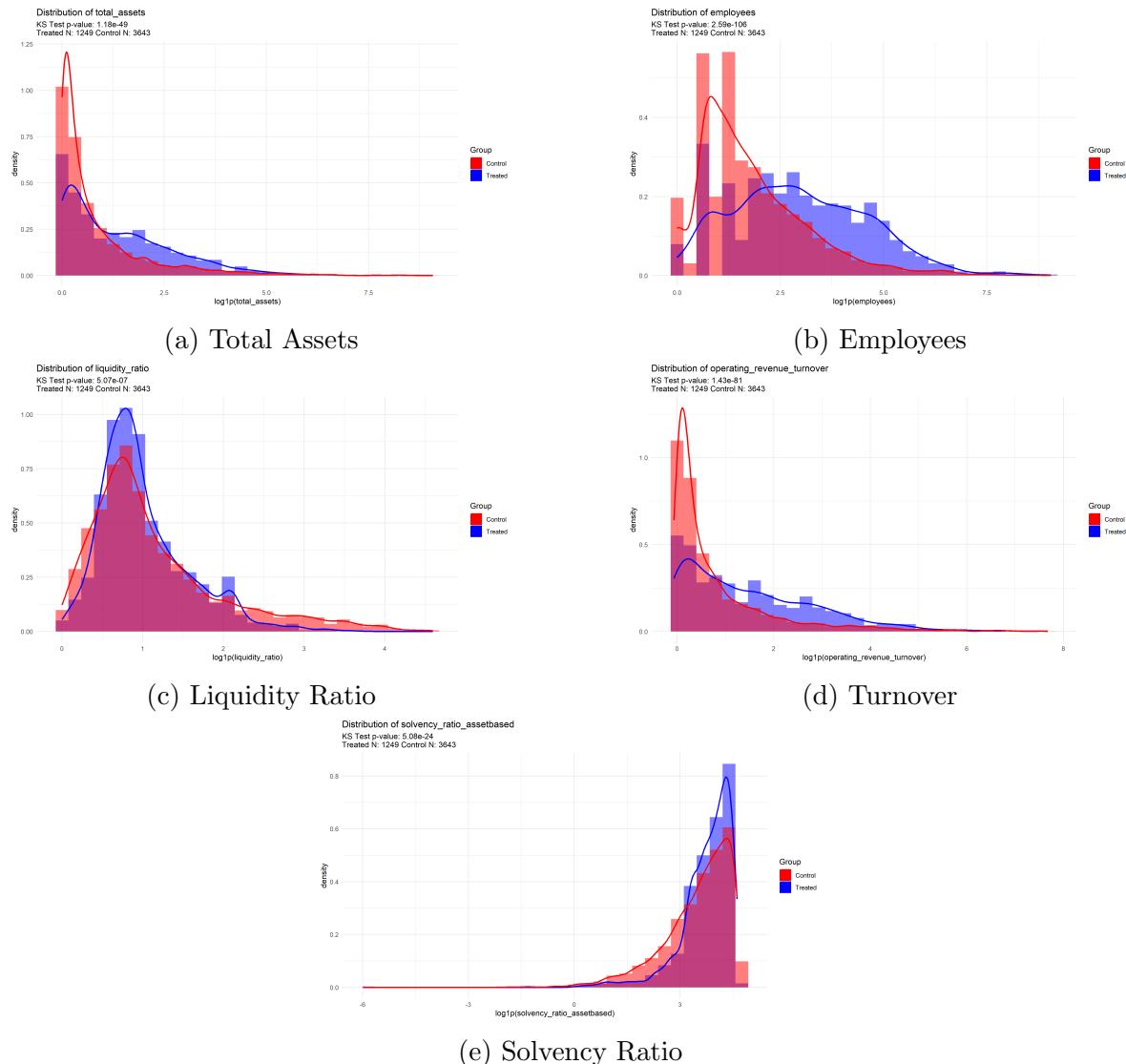


Figure 5.1: Distribution Comparisons Across All Countries

I also wanted to compare the sectors in which the firms operate, to see if we can find significant differences there as well. To do that, I divided the firms in the treated and control group using the NACE classification. Industry codes from NACE come as 4-digit codes, but I aggregated them to the 2-digit level to have a more general classification, as well as a more readable graph. Results can be seen in the following figure (as usual, treated firms are in blue and controls in red).

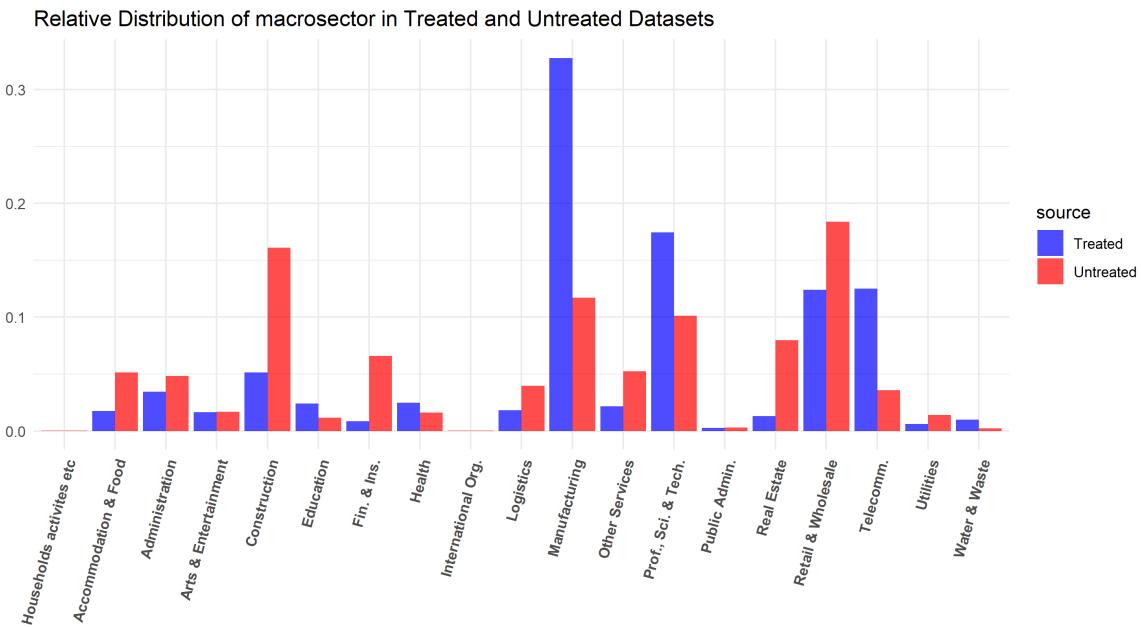


Figure 5.2: Comparison of Macro-Sectors in Treated and Control Group

What is immediately noticeable from the graph is how much the sectoral distribution of firms differs between treated and controls. In particular, the macrosectors relating to "Manufacturing", "Professional, scientific and technical activities" and "Information and communication" are overrepresented in the treated group, while "Real estate", "Construction" and "Wholesale and retail trade" are underrepresented.

This is probably due to the declared focus of the EDIH initiative on digitalization, and especially on manufacturing technologies. After all, a manufacturing firm can be more easily interested in digitalization since it can profit more from the adoption of these new technologies, while a construction firm may not be as interested. This is an

argument for including the macrosector as an independent variable in the regression.

5.4 Geocoding, distance and density computation

As I was mentioning in the introductory section of this chapter, I wanted to investigate the effect of the distance of firms from the closest EDIH hub on the probability of participation in the program.

At the start of the process, what I had were addresses, either taken from the DMA survey (for the treated sample) or from the ORBIS database (for the control sample)⁸.

To get the correct coordinates of the firms, I used the `geocode` function (part of the `tidygeocoder` package) in R. This function takes the address string as input and runs the Nominatim API to get coordinate information from the OpenStreetMap database. The function is efficient, but the Nominatim API limits requests to one per second, so I had to time 1 second of sleep for the code for it to work properly. Thus, the process was time-consuming⁹.

Once the geocoding loop was done, I had the coordinates of all the firms in the treated and control groups. You can see their geographical distribution in the following map, where treated firms are represented by the red dots, control firms by the light blue dots, and EDIH hubs by the dark blue and dark green dots (the former for EU-financed hubs, the latter for Seal of Excellence hubs).

Already from this map of the whole EU, we can notice some aspects of the analysis we've already discussed. For example, it is clear how the ORBIS coverage is not equally good in each EU country. As a reference, look at the difference in coverage between France and Belgium, or between Italy and Spain. This is important to keep in mind when designing the regression model to use for the analysis, and this is on of the

8. In a small number of cases, ORBIS reported coordinate information as well as the address; in those cases, I took those as being correct, and did not run those observations through the geocoding process

9. Note that the process was done for roughly 18000 firms in the control group, as well as all the firms in the DMA database, and every EDIH in the database as well. Thus, the geocoding part of the process took some hours to complete

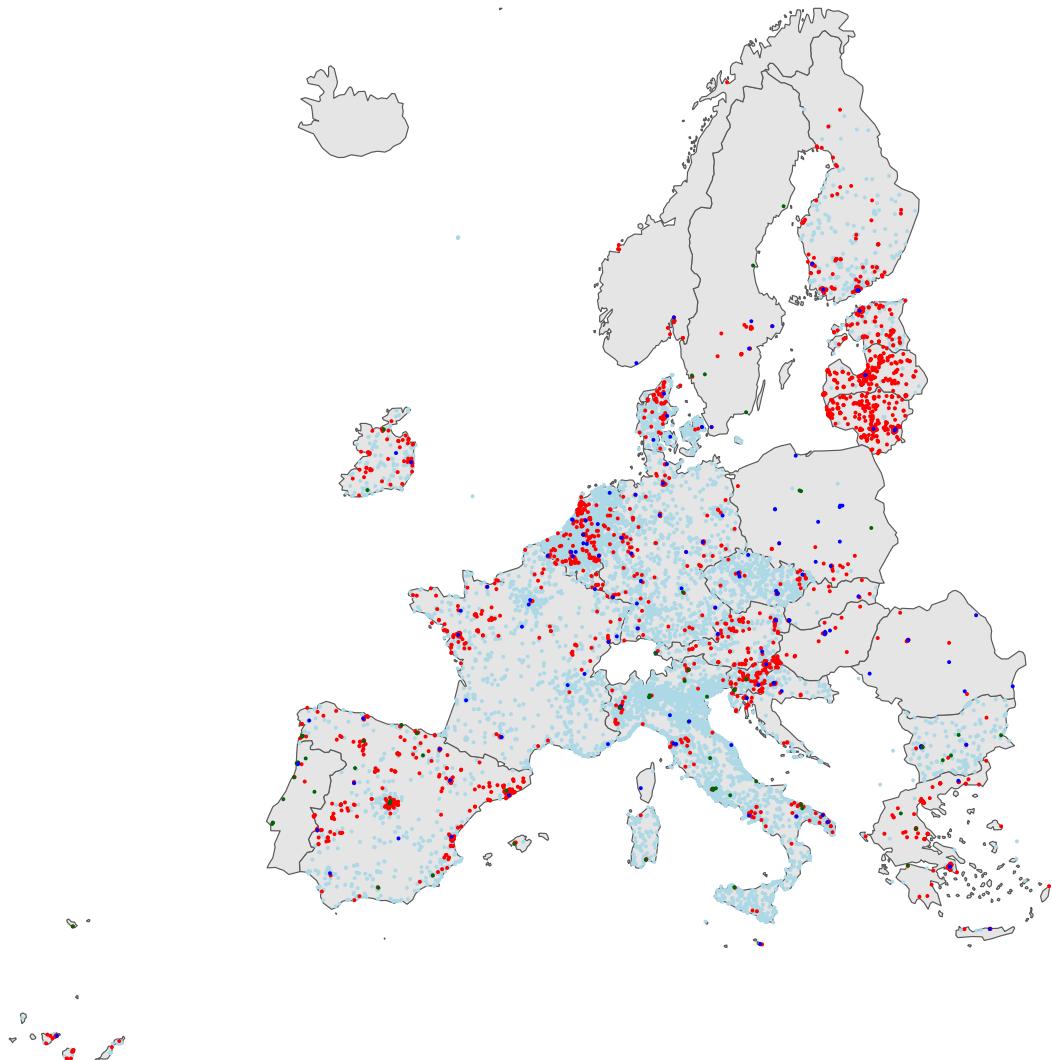


Figure 5.3: Map showing the geographical distribution of treated firms (in red), control firms (in light blue) and EDIH hubs (with EU-funded ones in dark blue and regionally/nationally funded ones in dark green) in the whole EU

reasons why we will include the country-groups' dummies as an independent variable in the regression.

One other thing that we can notice is how the EDIH coverage is not uniform across EU countries, and not even across regions in the same country. For example, let's take a look in detail at Italy, reported individually in the map below.

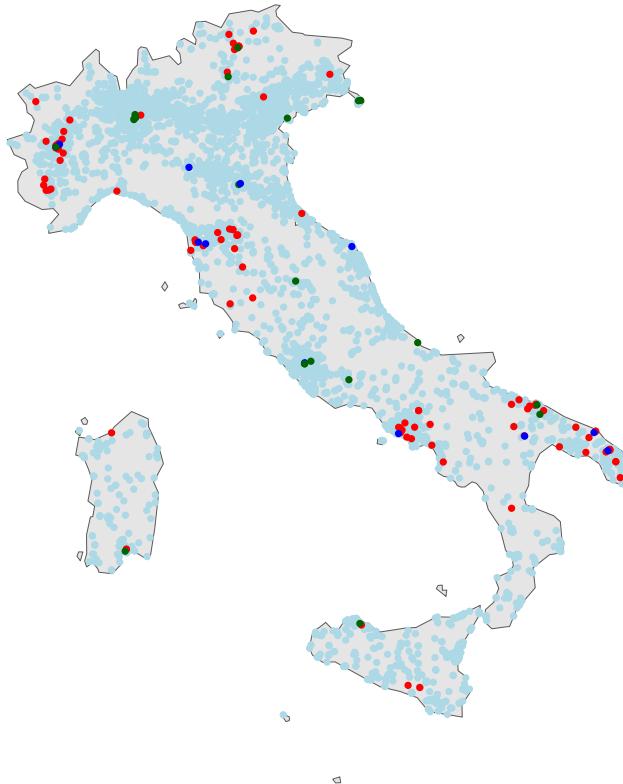


Figure 5.4: Map showing the geographical distribution of treated firms (in red), control firms (in light blue) and EDIH hubs (with EU-funded ones in dark blue and regionally/nationally funded ones in dark green) in Italy. All the other individual countries' maps are reported in the Appendix.

From this map of firms in Italy, we can see how treated firms are concentrated close to the location of EDIH offices¹⁰. This is an initial indication of the negative effect of distance on the probability of accessing the initiative.

However, it can also be noted that in many cases treated firms are concentrated in clusters around densely "populated" areas, where also hubs tend to be located. Thus,

10. One additional note can be made about the fact that EU-financed EDIHs seem to have more treated firms around them w.r.t Seal of Excellence Hubs, which are instead financed through national or regional funds. This fact is not valid only in Italy, and can be an avenue for future research on the effectiveness of the initiative itself. Initial data from the DMA seems in fact to suggest that EU-EDIHs are more active than SoEs. However, please note that this is more likely to be due to a lack of certainty about funding, which is likely delaying start of operations for SoEs. Time will tell if their performance will be better in the future.

without including a measure of the "firm density" of the territory in which the firms operate, the effect of proximity to the hubs is likely to get confounded by the effect of being in an area of high firm density¹¹.

Now that we have coordinates for every firm and EDIH, computing distances is a straightforward (albeit lengthy) process. To get distances, for every firm in the database of treated and controls, I computed its distance with each of the EDIHs in the database, and then chose the closest one. This was done using the `distHaversine` function from the `geosphere` package in R. This function takes the 2 sets of coordinates as input and computes the distance between them using the Haversine formula, which is the most accurate way to compute distances on a sphere.

After having computed the distances, I then computed the firm density for each firm. To do this, I utilized the following procedure. For every firm, I first computed the distance between it and every other firm in the database (both treated and controls). Then, I filtered out all firms more distant than a certain threshold¹², and counted the number of firms left. You can see a representation of the process in the following figure.

In the diagram you can see how Firm A has a firm density of 3, as firms B, D, and E all fall inside of the 5km-radius circle around Firm A (while EDIH offices are not counted as firms). On the other hand, Firm C has a firm density of only 1, since Firm B is the only one inside the 5km-radius circle around Firm C (and again, EDIH offices are not counted as firms). From the diagram you can also see how in the computation of firm density, both treated and control firms are taken into account.

Thus, the designed measure of firm density can be considered as an accurate proxy of the density of firms on the EU territory, at least when the ORBIS bias is taken into account.

Now that we have computed both firm distance and density measures, we can add

11. We can think of firm density as impacting the decision of the firm to participate in the EDIH initiative in many ways. For example, it can be that being surrounded by many firms exposes the firm more to competition, thus making it more imperative to digitalize, and thus seeking help to do so.

12. After a few trials, I settled on a threshold of 5 km, deemed as a good compromise to have an acceptable level of variance in firm density. In fact, a threshold too high would have led to a very high density for every firm, while a threshold too low would have led to a very low density for every firm.

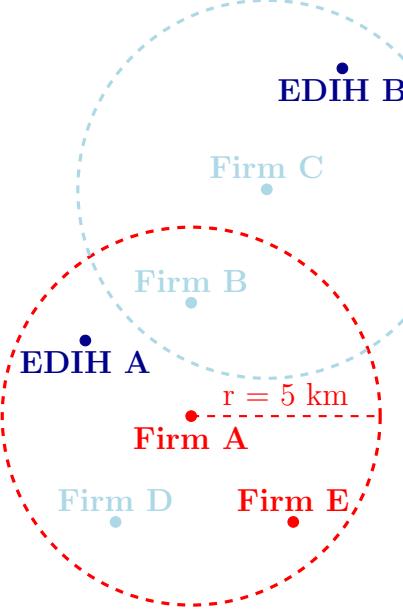


Figure 5.5: Diagram showing the criteria for computing firm density. In the image, Firm A has a firm density of 3, while Firm C has a firm density of 1.

them to the initial treated-control comparison, to see how they differ between the two groups, as seen in the following tables.

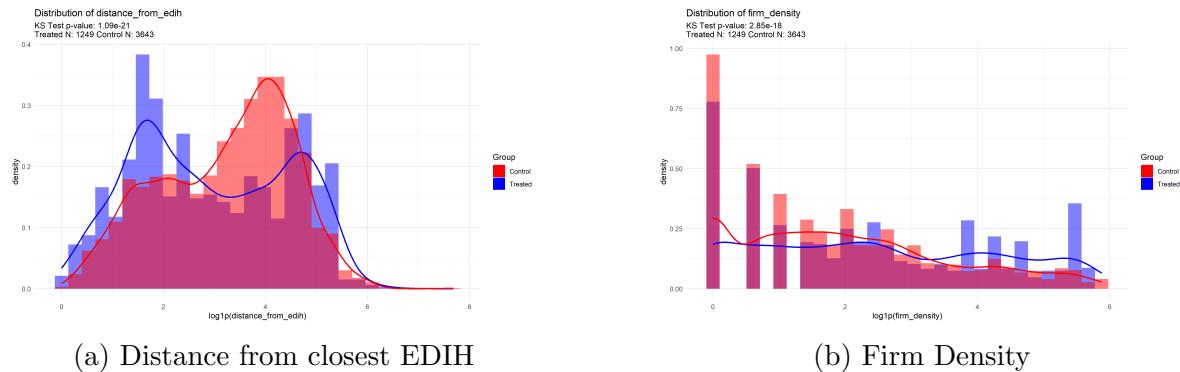


Figure 5.6: Distribution Comparisons of distance and firm density across all Countries

In both cases, we can see from the results of the KS-tests that the treated and control groups hail from different distributions. In addition, it seems that treated firms are on average closer to the EDIH hubs, are also located in areas with higher firm density. This is an additional indication that the spatial dimension is significant in understanding the drivers of firms' participation to the EDIH initiative.

5.5 Regression Model and Results

As previously mentioned in the sections above, to analyze the drivers of participation in the EDIH initiative, I have run a probit regression model. The model includes various firm-level characteristics as independent variables, and the dependent variable is a dummy variable indicating whether the firm participated in the EDIH initiative. The model is specified as follows:

$$\begin{aligned} D_i(\text{Treated}_i = 1) = & \beta_0 + \beta_1 \text{Distance}_i + \beta_2 \text{Firm Density}_i \\ & + \beta_3 \text{Sector}_i + \beta_4 \text{Tech Level}_i \\ & + \beta_5 \text{Turnover}_i + \beta_6 \text{Employees}_i \\ & + \beta_7 \text{Liquidity Ratio}_i + \beta_8 \text{Solvency Ratio}_i \\ & + \beta_9 \text{Country Group}_i + \varepsilon_i \end{aligned} \quad (5.1)$$

Where the dependent variable is a dummy indicating participation to the program, Distance and Firm Density are defined as above, Sector is a variable indicating the type of sector to which the firm belongs (either Manufacturing, Services or Other, with the baseline being Other). Tech Level is a variable indicating the level of technological intensity of the sector in which the firm is operating, according to the OECD taxonomy defined in Galindo-Rueda and Verger (2016); in this case, the baseline is represented by the lowest level of the OECD taxonomy. Turnover, Employees, Liquidity Ratio and Solvency Ratio are all firm-level variables coming from the ORBIS database. More specifically, turnover serves as a proxy for firm profitability, employees proxies for firm size (total assets was excluded to being too collinear with employees), while liquidity ratio and solvency ratio are proxies for the financial wellbeing of the firm. Country Group is a dummy indicating the group to which the country in which the firm is located belongs; the groups are Central Europe, Southern Europe, Baltics, Benelux, Balkans, and Western Europe, with the baseline being Central Europe.

In addition to running this regression on the whole sample, I also ran it on the subsamples of firms in the Manufacturing sector, in the Services sector, and in the Other sector, to see if results were particularly driven by one of these macrosectors of the economy.

In the following table, you can first see the results of the regression analysis on the whole sample, first without the spatial dimensions (distance and firm density), and then with them included. The results are reported as coefficients, with the standard errors in parentheses.

From the results of the regression¹³, we can see how the spatial dimensions are significant in explaining the probability of a firm to participate in the EDIH initiative. In particular, the coefficient of Distance is negative and significant, while the coefficient of Firm Density is positive and significant. This is an indication that firms that are closer to the EDIH hubs are more likely to participate in the program, while firms that are located in areas with higher firm density are also more likely to participate.

The other variables in the regression are almost all significant. Starting with the sectoral type variable, we see confirmed what we have already noticed in the comparison of macrosectors: firms in the Manufacturing sector are more likely to participate in the program, compared to the baseline of firms in the Other sector. This is true also for firms in the Services sector, indicating that it is not Manufacturing per se that is driving the effect, but perhaps a particular focus on digitalization that is more present in segments of both of these macrosectors.

Proceeding to analyze the effect of the tech level of the sector in which the firm operates (based on the OECD classification and NACE 2-digit codes), we see that firms in sectors with higher technological intensity are more likely to participate in the program. The baseline here is represented by those sectors with the lowest level of technological intensity according to the OECD. This indicates that the EDIH initiative is indeed correctly targeting firms that are in technologically intensive sectors, and thus

13. Be aware that results from this regression cannot be interpreted in any way apart from their sign: to have coefficient that can be interpreted we have to compute Average Marginal Effects

Table 5.2

	w/o Dist	w/ Dist	treated w/ Dist and Dens
distance_from_edih		-0.002*** (0.0004)	-0.002*** (0.0004)
firm_density			0.001*** (0.0005)
sector_typeManufacturing	0.328*** (0.103)	0.337*** (0.103)	0.361*** (0.103)
sector_typeServices	0.269*** (0.074)	0.245*** (0.074)	0.222*** (0.074)
tech_levelHigh	1.260*** (0.181)	1.233*** (0.182)	1.216*** (0.183)
tech_levelMedium	0.783*** (0.143)	0.766*** (0.144)	0.749*** (0.144)
tech_levelMedium-High	1.592*** (0.088)	1.554*** (0.089)	1.539*** (0.089)
tech_levelMedium-Low	0.774*** (0.070)	0.745*** (0.071)	0.723*** (0.072)
operating_revenue_turnover	-0.0004 (0.0005)	-0.0004 (0.0005)	-0.0004 (0.0005)
employees	0.768*** (0.165)	0.749*** (0.165)	0.741*** (0.164)
liquidity_ratio	-0.092*** (0.009)	-0.094*** (0.009)	-0.095*** (0.009)
solvency_ratio_assetbased	0.010*** (0.001)	0.010*** (0.001)	0.010*** (0.001)
Observations	4,864	4,864	4,864
Akaike Inf. Crit.	3,421.334	3,395.981	3,388.624

Note:

65

*p<0.1; **p<0.05; ***p<0.01

more likely to benefit more from digitalization initiatives.

Passing on to firm-level characteristics from ORBIS, we notice how the coefficient for turnover is the only one being not statistically significant, while all the others are. This could indicate that profitability is not a significant driver of participation in the program¹⁴. However, the results of the regression also denote how larger firms are more likely to participate in the program. The coefficient related to liquidity ratio is instead negative, suggesting that firms that have more cash on hand perhaps do not need to participate in the program, as they can finance their digitalization initiatives without accessing the EDIH network. Finally, the coefficient for the solvency ratio is positive, indicating that firms that are more financially stable are more likely to participate in the program.

Note that coefficients related to the country group variable were significant, but I have cut them from the table to save space. The results however confirm what was already seen in the treated-control comparison of numerosity by country. Firms in the Baltics, Benelux, Southern Europe and Balkans are all significantly more likely to participate in the program, compared to firms in Central Europe.

In the following table, I report the computed Average Marginal Effects for the full regression model, which can be interpreted as the effect of a small change in the independent variable on the probability of a firm to participate in the EDIH initiative.

The Average Marginal Effects (AMEs) reported in Table 5.3 provide a more intuitive understanding of the impact of each independent variable on the probability of a firm participating in the EDIH initiative.

Starting with the spatial dimensions, the AME for Distance is -0.0003, indicating that for every additional kilometer a firm is located away from the nearest EDIH hub, the probability of participating in the initiative decreases by 0.03%. This effect, although small, is statistically significant and highlights the importance of proximity to

14. This coefficient can be interpreted as a good sign w.r.t. the targeting of the initiative. If we were to find that, on average, for example, the EDIH initiative was participated by less profitable firms, then we could consider it akin to a government subsidy to inefficient firms, instead of an effective digitalization policy.

EDIH hubs in driving participation. On the other hand, the AME for Firm Density is 0.0003, suggesting that an increase in the number of firms within a 5 km radius by one unit increases the probability of participation by 0.03%.

For the sectoral type variables, the AMEs for Manufacturing and Services are 0.071 and 0.042 respectively. This means that firms in the Manufacturing sector are 7.1% more likely to participate in the EDIH initiative compared to firms in the Other sector, while firms in the Services sector are 9.8% more likely to participate.

The AMEs for the technological intensity levels show a clear gradient. Firms in sectors with medium-high and high technological intensity are 38% and 29% more likely to participate in the initiative, respectively, compared to firms in low-tech sectors. This gradient indicates that the EDIH initiative is effectively targeting firms in more technologically advanced sectors, which are likely to benefit more from digitalization support. Furthermore, it indicates that most of the firms participating in the initiatives are in cutting-edge sectors.

Among the firm-level characteristics, the AME for Employees is 0.143, indicating that each additional employee increases the probability of participation by 14.3%. This positive effect suggests that larger firms are more likely to engage with the EDIH initiative, possibly due to their greater capacity to undertake digital transformation projects.

Finally, the country group variables (which were not shown in the previous regression table) also have significant effects. Particularly worthy of note is the coefficient related to the Baltic States, indicating that firms in that area have a 45.9% higher probability of participating in the initiative compared to firms in Central Europe. This is in line with what we discussed in previous chapters, when we noted that the Baltics had a head start in the EDIH initiative due to having hosted multiple successful DIH in the past.

As previously mentioned, I also ran the regression on the subsamples of firms in the Manufacturing, Services, and Other sectors. The results of these regressions are reported in the following table.

Table 5.3: Average Marginal Effects for Full Model

	factor	AME	SE	z	p	lower	upper
1	country_groupbalkans	0.161	0.020	8.178	0	0.122	0.199
2	country_groupbaltics	0.459	0.019	23.757	0	0.421	0.497
3	country_groupbenelux	0.062	0.029	2.161	0.031	0.006	0.118
4	country_groupeastern	-0.048	0.014	-3.503	0.0005	-0.075	-0.021
5	country_groupsouth	0.138	0.015	9.506	0	0.110	0.166
6	country_groupwestern	-0.020	0.016	-1.264	0.206	-0.052	0.011
7	distance_from_edihs	-0.0003	0.0001	-3.642	0.0003	-0.0005	-0.0001
8	employees	0.143	0.032	4.545	0.00001	0.082	0.205
9	firm_density	0.0003	0.0001	3.126	0.002	0.0001	0.0005
10	liquidity_ratio	-0.018	0.002	-10.657	0	-0.022	-0.015
11	operating_revenue_turnover	-0.0001	0.0001	-0.755	0.450	-0.0003	0.0001
12	sector_typeManufacturing	0.071	0.021	3.425	0.001	0.030	0.111
13	sector_typeServices	0.042	0.014	3.086	0.002	0.015	0.069
14	solvency_ratio_assetbased	0.002	0.0002	11.253	0	0.002	0.002
15	tech_levelHigh	0.290	0.050	5.749	0	0.191	0.389
16	tech_levelMedium	0.164	0.036	4.555	0.00001	0.093	0.235
17	tech_levelMedium-High	0.380	0.023	16.253	0	0.334	0.426
18	tech_levelMedium-Low	0.157	0.017	9.264	0	0.124	0.191

Table 5.4

	<i>Dependent variable:</i>		
	Manufacturing	Services	Other
distance_from_edih	-0.004*** (0.001)	-0.001** (0.001)	-0.001 (0.001)
firm_density	-0.004** (0.002)	0.002*** (0.001)	0.001 (0.002)
tech_levelHigh	0.603 (0.698)	1.769*** (0.255)	
tech_levelMedium	0.605 (0.655)		0.784** (0.383)
tech_levelMedium-High	1.333** (0.644)	1.669*** (0.106)	
tech_levelMedium-Low	0.781 (0.640)	0.554*** (0.080)	0.733 (0.615)
operating_revenue_turnover	-0.003*** (0.001)	0.001** (0.001)	-0.004*** (0.001)
employees	1.664*** (0.484)	0.616*** (0.170)	9.114*** (1.615)
liquidity_ratio	-0.118*** (0.023)	-0.086*** (0.010)	-0.100*** (0.035)
solvency_ratio_assetbased	0.015*** (0.003)	0.010*** (0.001)	0.009*** (0.003)
Constant	-2.450*** (0.679)	-2.324*** (0.191)	-2.601*** (0.476)
Observations	842	3,084	938
Log Likelihood	-351.070	-1,040.701	-203.582
Akaike Inf. Crit.	736.141	2,113.401	437.163

Note:

*p<0.1; **p<0.05; ***p<0.01

What we can see from this table is that the effect of the spatial dimensions is not driven by one of the macrosectors. In fact, the coefficients for Distance is significant and negative in the first 2 regressions¹⁵ (albeit larger for the Manufacturing subsample), while the coefficient for firm density seems to be entirely driven by the Services sector, where it is significant and positive. If we restrict the analysis to Manufacturing only though, the coefficient is still significant, but negative.

Most of the other coefficients tell the same story as in the full regression. However, we should notice how the effect of technological intensity for very high-tech sectors seems to be driven largely by the Services sector. Also, we should note that when we do not take sector into account, the effect of turnover is significant, negative for Manufacturing and Other, while positive for the Services sector. This is an indication that the effect of turnover is likely to be confounded by the sector in which the firm operates.

As a check on the results, you can see below a graph where I plot the change in coefficients caused by the exclusion of each country, to see if any particular country is driving the results.

15. Unfortunately, the small representation of the Other macrosector in the sample does not allow us to measure the coefficient precisely

We can see here how for most of the coefficients the results are stable when removing one country at a time from the sample. However, if we focus on the coefficient for the Distance variable, we can see how the exclusion of Italy from the sample would lead to a slightly significant positive change of the coefficient. This is an indication that the negative effect of distance is stronger in the Italian subsample of firms. However, even when accounting for the potential change, the coefficient for distance would remain negative and significant.

Moving on to Firm Density, it is clear that the exclusion of the Baltics, specifically Finland and Latvia, leads to a significant positive change in the coefficient. This suggests that the positive effect of firm density is actually weakened by the presence of firms in the Baltic States. This could be due to the specific geographical characteristics of the area, which is more sparsely populated than other regions of the EU. Consequently, most firms in the sample are concentrated in small pockets of high firm density, reducing the variability needed to accurately estimate the coefficient and thus pushing it towards zero. Other countries which when excluded move the coefficient are Italy and (to a lesser degree) Spain, both moving the coefficient in the negative direction when excluded, meaning their inclusion is having a positive effect on the coefficient. This can be due to a multiplicity of reasons, however one probable cause is the fact that both Italy and Spain have a very good coverage of the population of firms in ORBIS. Thus, even "rural" firms are included in the control group, and since instead most of the treated firms come from high-firm-density areas, the coefficient is pushed upwards.

Focusing on the firm-level characteristics, we can see how the coefficient of the Employees variable seems to be driven by Belgium. When excluding the small country, the coefficient drops significantly. One potential reason for this could be that most of the very large firms in the treated sample come from Belgium, thus increasing the effect of firms' size in the subsample. For the Liquidity Ratio, the coefficient is driven by the exclusion of Italy and Latvia: Italy when excluded drops the coefficient, while Latvia increases it. However, none of the two would change the significance or sign of the coefficient. For the Turnover variable, the coefficient is once again influenced by

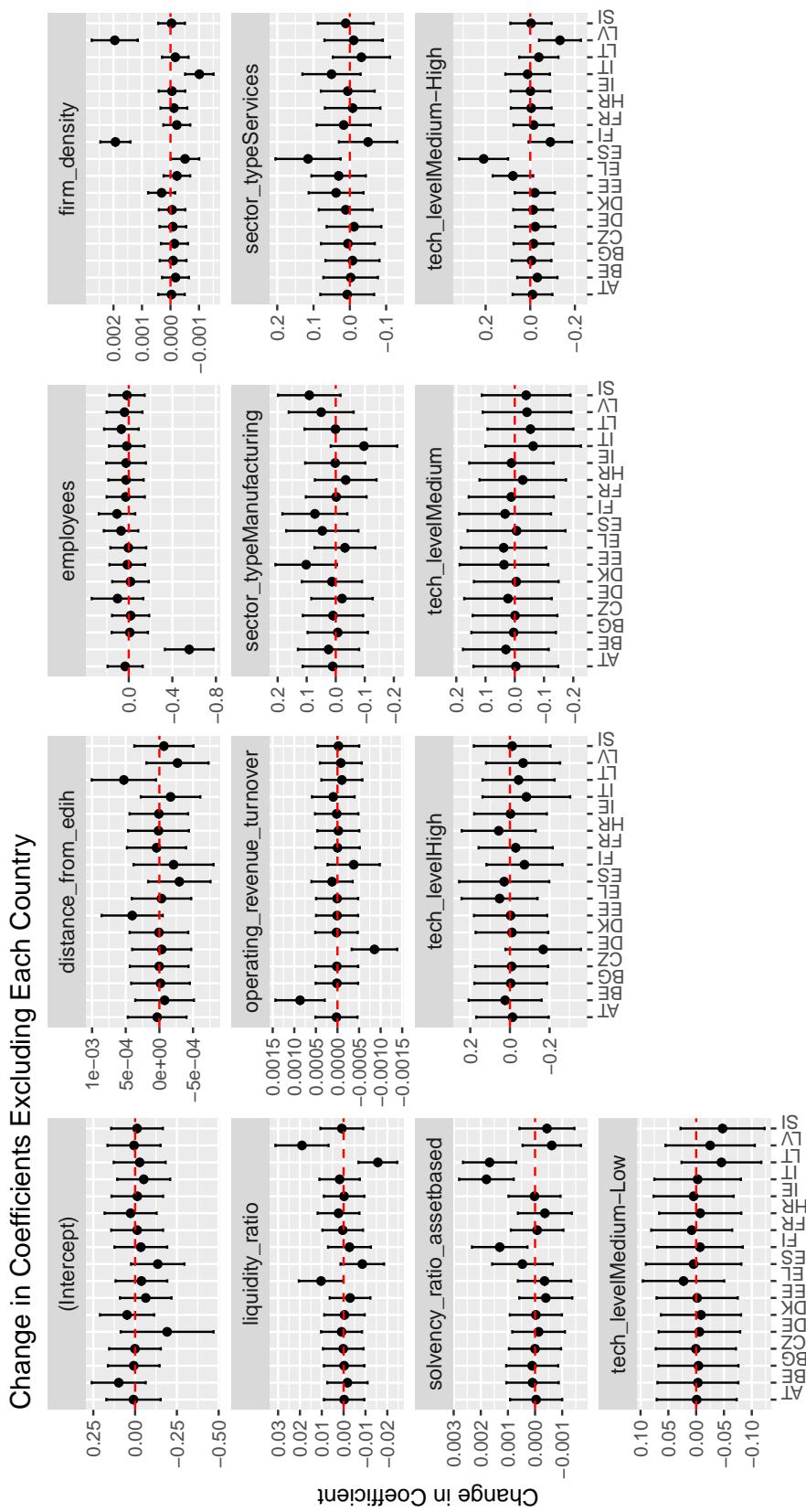


Figure 5.7: Changes in coefficients caused by the exclusion of each country. Error bars are Standard Errors from the coefficient estimation.

Belgium, as well as Germany. Since the coefficient in the analysis was not significant, this could mean that these two countries were the ones making the coefficient closer to zero. Once we exclude one or the other, the coefficient shifts enough to be considered significant. The Solvency Ratio coefficient is lifted upwards by the exclusion of Finland, Italy and Lithuania. However, this is going in the same direction of the sign of the coefficient, thus not changing the interpretation of the results.

With regards to the sectoral variables, most of their coefficients stay put, apart from Services when excluding Spain, which increases the coefficient. This is paralleled in the tech level coefficients, where we can see that they stay unchanged everywhere, apart from the Medium-High tech level when excluding Spanish firms, which increases the coefficient. This could indicate that the positive effect of being in the Services sector and in a medium-high tech level segment is significantly weaker in Spain than in the other countries present in the subsample. Again, however, the potential coefficient change would not alter the interpretation of the results from the regression.

In summary, the regression results show us what the key drivers of participation in the EDIH initiative are. Proximity to EDIH hubs and higher firm density are significant spatial factors influencing participation. Additionally, firms in the Manufacturing and Services sectors, as well as those in technologically intensive sectors, are more likely to engage with the initiative. Larger firms and those with higher solvency ratios also show a higher propensity to participate. These findings are for the most part consistent across different sectors and countries, as shown by the sectorial regression and by the analysis of changes in the coefficients when excluding each country from the sample.

Chapter 6

Discussion and Conclusions

As discussed in the previous chapter, the results of the analysis allows us to answer to our main research question, on the identification of the drivers of firms' participation to the EDIH initiative. As the regression results show, geospatial characteristics do have a role to play in the propensity of firms to participate in the initiative. In particular, firms that are located closer to EDIH hubs are more likely to take part in the initiative, and firms that are located in areas with high firm density are also more likely to participate. Moreover, sectoral characteristics also play a role, with the Manufacturing sector being more likely to participate, and more technologically intensive sectors being on average more likely to get involved in the initiative. Firm-level characteristics also play a role, especially size (number of employees) and financial constraint metrics.

How does the EDIH program fare so far then? The results of the analysis suggest that the initiative is effectively targeting firms that are in need of digitalization support, as evidenced by the fact that the coefficient related to the liquidity ratio is negative. On the other hand, the program seems to attract larger firms, and those with higher solvency ratios. This could be a sign that the initiative is not reaching the smallest firms. However, it can be argued that micro-firms may see less of an incentive in participating in the program, and probably face more barriers to participation, such as lack of resources or lack of awareness of the initiative. This should not be seen as an out-

right failure on the part of the EDIH program, since it is also possible that a minimum level of resources is needed to engage with the initiative effectively. Furthermore, the initiative is designed to kickstart the catching-up progress of the European economy, which is more likely to come from small and medium enterprises than from micro firms with 1 or 2 employees. Regarding the positivity of the coefficient for the solvency ratio instead, as we briefly touched in the previous chapter, it can be a good sign of the fact that the EDIH initiative is not an outright subsidy for inefficient firms on the brink of bankruptcy, but rather a policy attracting firms in good financial health, albeit not as rich in liquidity as to be financially unconstrained. In fact, the liquidity ratio measures the short-term ability of the firm to pay its debts, while the solvency ratio indicates the long-term financial stability of the company. Thus the fact that the initiative is attracting firms with higher solvency ratios means that the firms are likely to survive in the long term, indicating that the public investment in the initiative is likely to be well spent.

The positive effects related to sectors with higher technological intensity are also good news, meaning that the impact of the policy is likely going to be stronger, since those are specifically the sectors in which the European economy needs to catch up.

Regarding the country-group dummies instead, they show the importance of EDIH's internal expertise, since it is clear how those regions that already had a strong experience with the DIH initiative experienced faster kickstart of the EDIH initiative itself, probably because they were already well-equipped to start at full speed. If this interpretation is correct, we should see this effect attenuate in the future, as (hopefully) the other regions catch up in terms of expertise and experience.

The coefficients related to geospatial characteristics instead stress the necessity of a good coverage of the whole EU territory for the initiative. These results can serve as an indication to policymakers needing to decide where to place potential additional EDIHs. Knowing that distance from the hub serves as a potential barrier to participation, adding offices in areas where no hubs are present could be a good strategy to increase the reach of the initiative.

Let's not forget however that the analysis performed presents some limitations and drawbacks, which prevent us from having a more clear picture of the effectiveness of the initiative and providing more insights and advice on how to make it better.

For instance, the analysis is based on a sample of control-group firms that may or may not be representative of the overall population of firms in the EU. As said in the previous chapter when describing how the control group was constructed, the ORBIS database does not have equally good coverage in all EU countries, and this could have introduced a bias in the results. Furthermore, for both the treated and the control group, it is possible that availability of data for the variables that were included in the analysis is indeed correlated with the probability to access the EDIH initiative. As an example, if the ORBIS database is more likely to have data on older firms, and for some reason age of the firm has a positive impact on the propensity to participate in the EDIH initiative, then we would have a problem, since we would have a positive bias in the results. The reason for that is the way the matching process is present for treated firms but not for the control group, leading to the possibility of the matched subsample of the treated group potentially being not representative of the overall treated group. This limitation would be challenging to overcome, since even if we already know which variables are correlated with a better coverage in ORBIS, we do not know *a priori* what their effect on the probability to participate in the EDIH initiative is.

There are of course also ways to improve the analysis, as well as avenues for future research on the topic. For example, even while still keeping the same methodology, it would be interesting to check if there are any word-of-mouth effects, meaning that being close to a firm that is already participating in the initiative could increase the probability of a firm to participate as well. This could be done by adding a variable that counts the number of treated firms in a certain radius around the firm, similar to how firm density was computed. Another interesting future research question would be on the proper functioning of the network of hubs itself, meaning whether hubs do in fact operate as a network, by exchanging clients, or if they act mostly on their own. In that case the analysis would have to be done on an EDIH basis, and not on an individual

firm basis.

Of course, when more data will be available down the line, regarding T_1 observations and eventually T_2 observations, it will be possible to perform a more robust analysis, with a difference-in-differences approach, to see if the EDIH initiative is indeed effective. However, the effectiveness check cannot be done on the Digital Maturity score itself, since the DMA was never sent to a control group of firms. The DMA can eventually be used to see if the adoption of targeted technologies increases in the treated group, but having no possible comparison makes it tricky to attribute the potential increase to the EDIH initiative, as opposed to some other factor. What will be possible however is to check the effectiveness of the initiative in having a positive impact on firm-level characteristics in the medium-term. Any effect on outcomes such as profitability of the firm, its size, or its financial health, would be a good indication of the effectiveness of the initiative. Even if not directly proving efficacy in providing digitalization, an increase in those metrics in the treated group vis-a-vis the control group would indicate the success of the initiative in supporting the catching-up process the European economy needs to become competitive again on the global stage.

Overall, to conclude we can say that the initial indication is that the program is going in the right direction, seemingly targeting the right demographic group of firms to increase its probability of success in the future. The size and sign of the coefficients related to geospatial characteristics indicate the need to maintain a network of EDIHs with a good coverage of the entire EU territory. Time (and upcoming data from the DMA) will tell us if the initiative is indeed effective in supporting the digitalization of the European economy, and if it will be ultimately successful triggering the improvement in firm's outcomes that the European Commission is hoping for.

Appendix A

Additional Tables and Figures

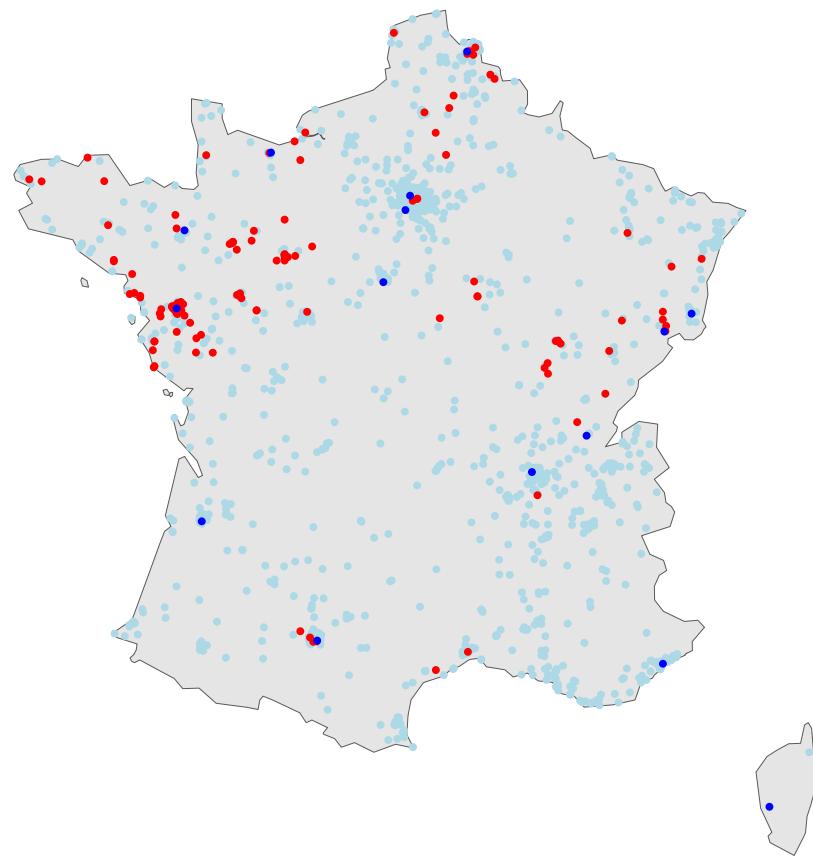


Figure A.1: France

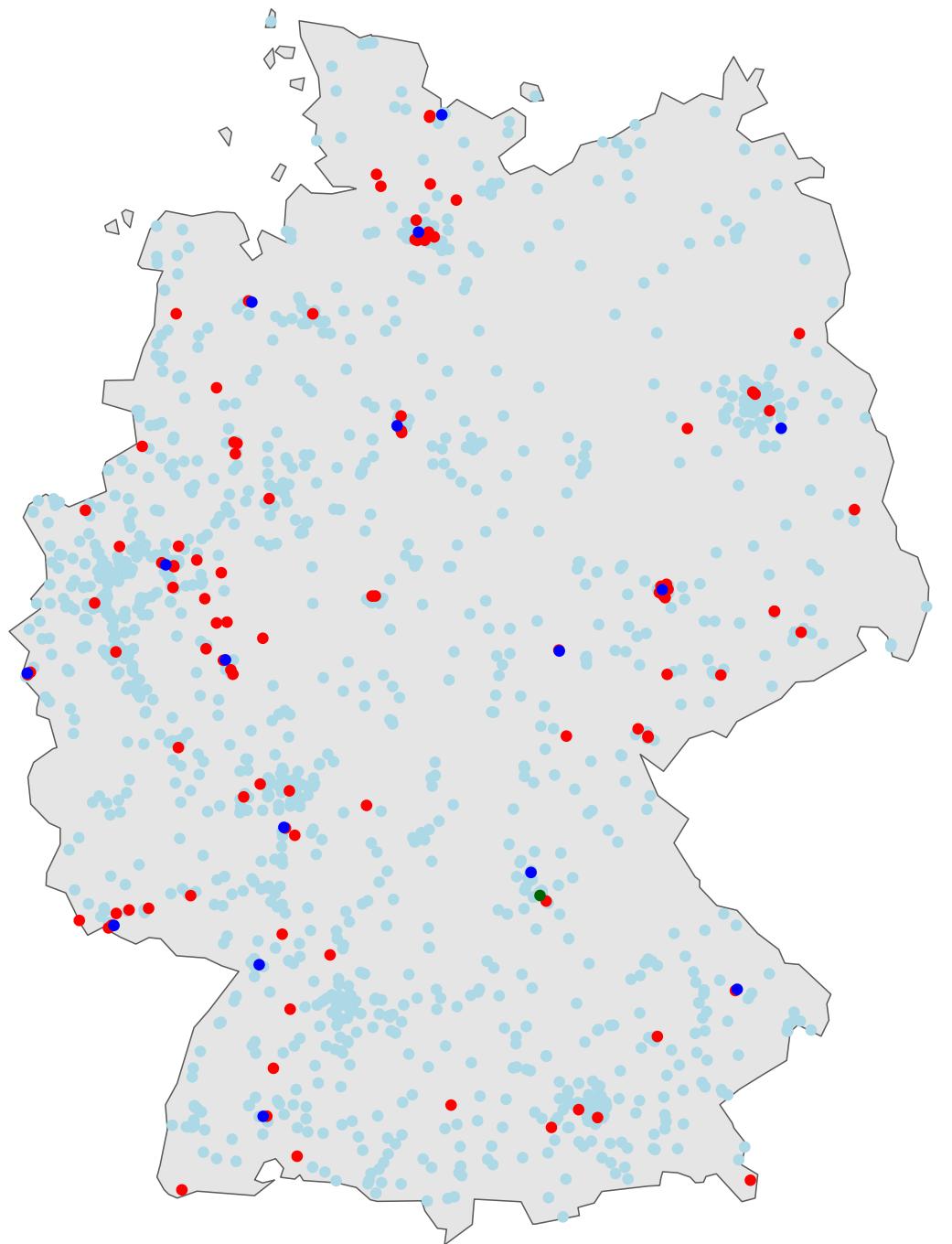


Figure A.2: Germany

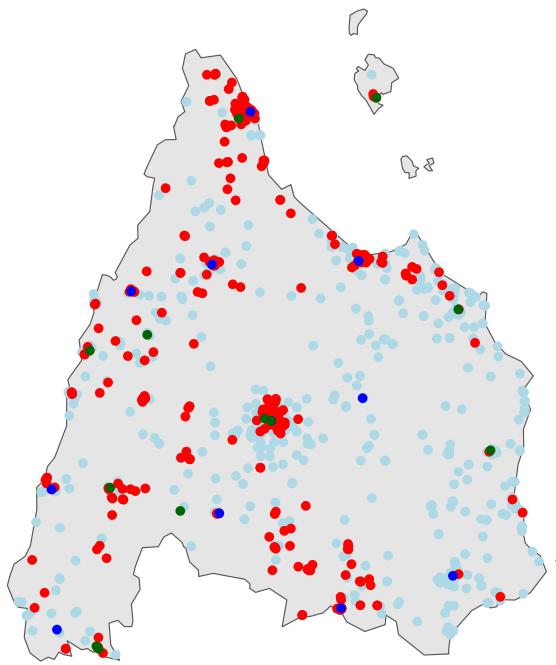
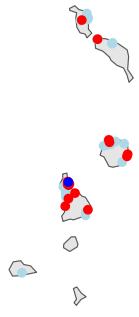


Figure A.3: Spain



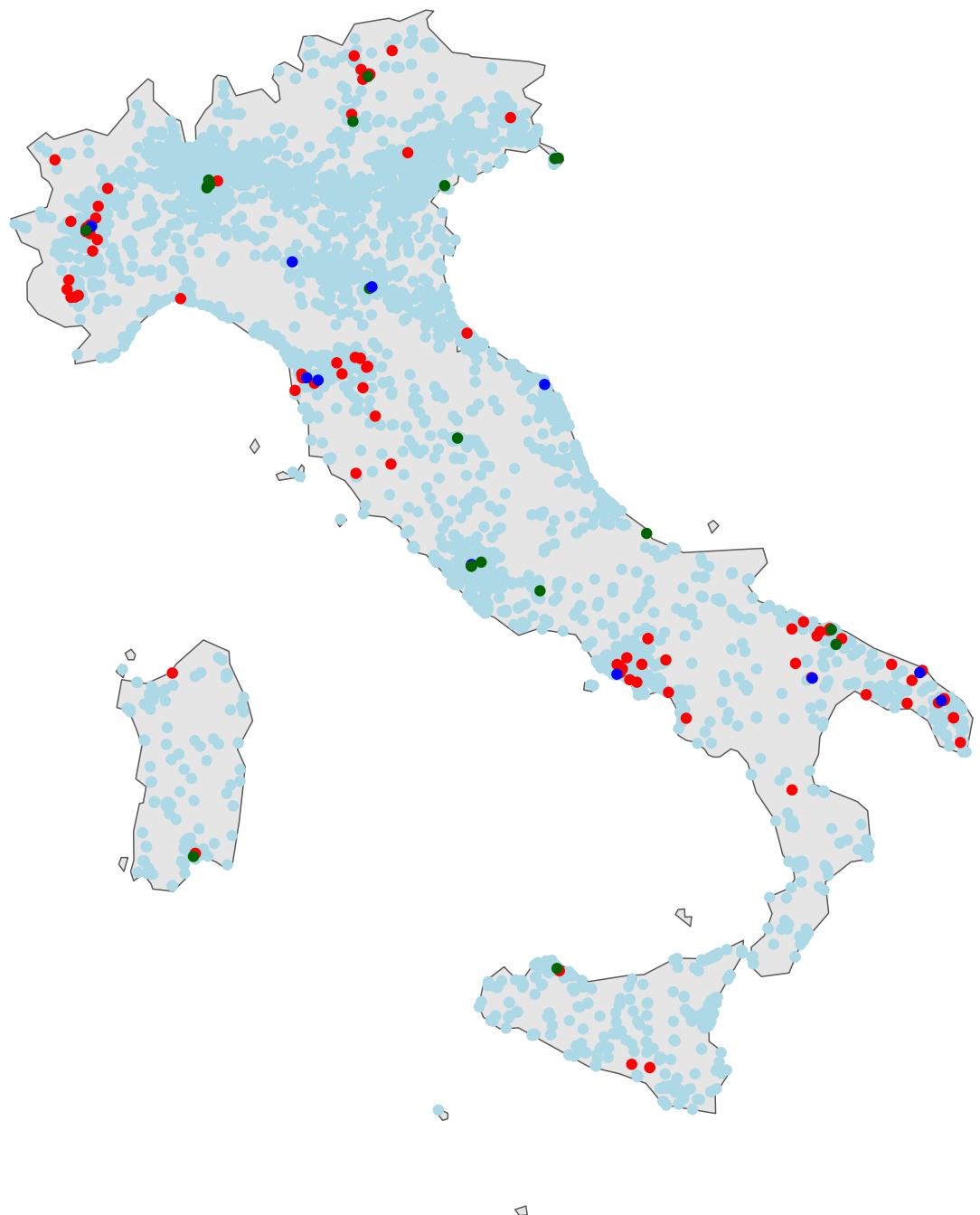
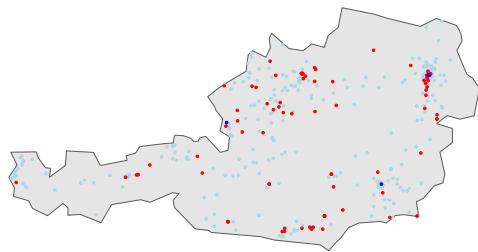
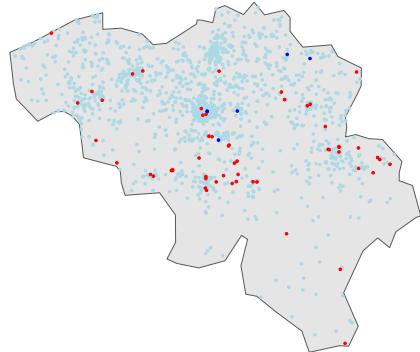


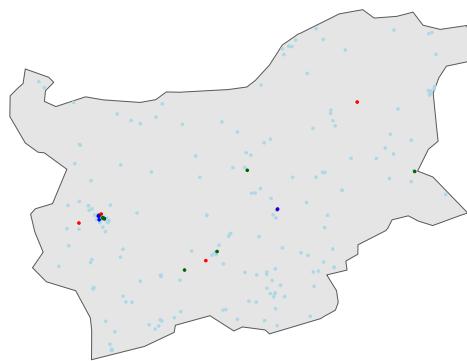
Figure A.4: Italy



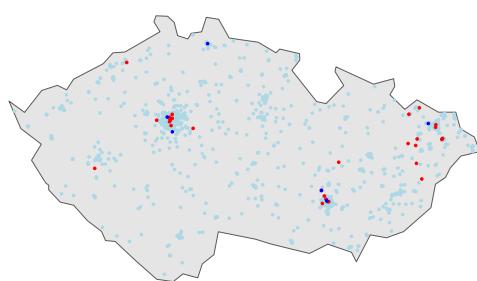
(a) Austria



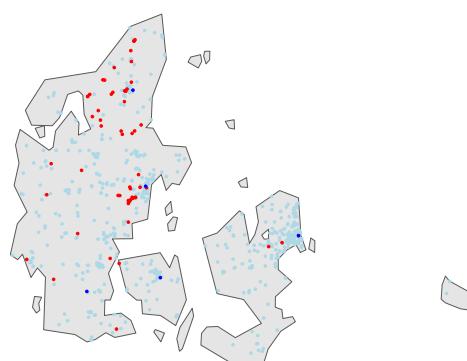
(b) Belgium



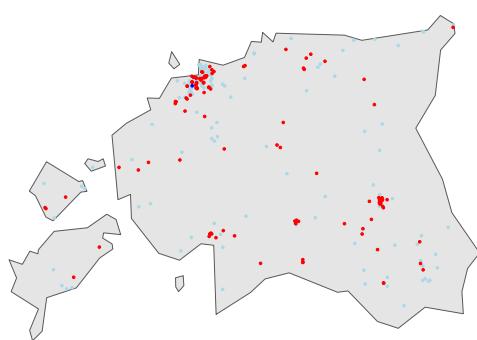
(c) Bulgaria



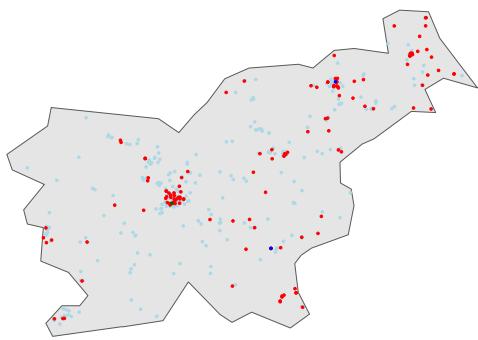
(d) Czech Republic



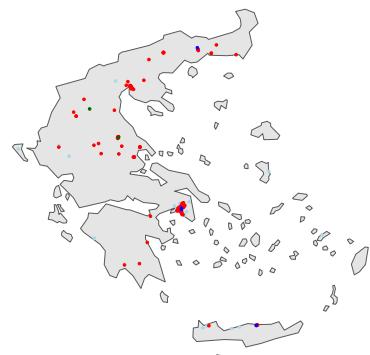
(e) Denmark



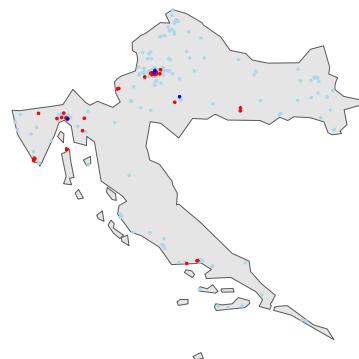
(f) Estonia



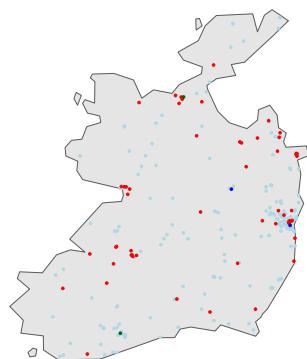
(a) Slovenia



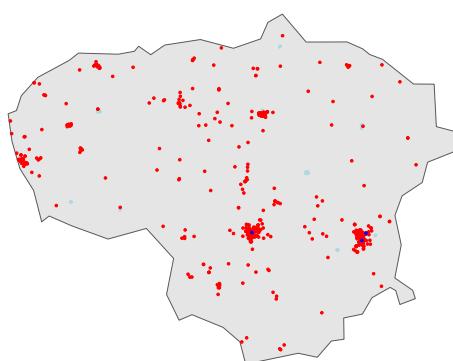
(b) Greece



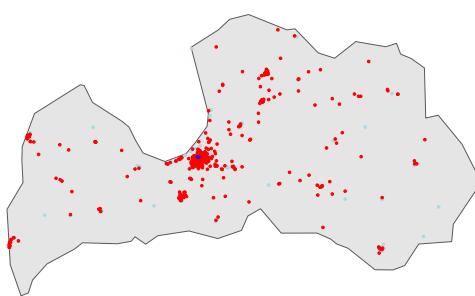
(c) Croatia



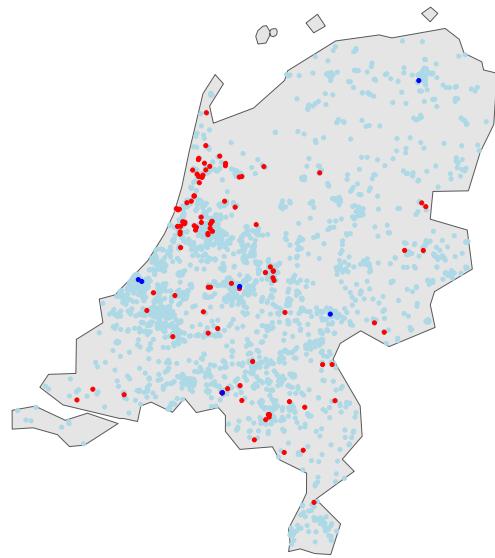
(d) Ireland



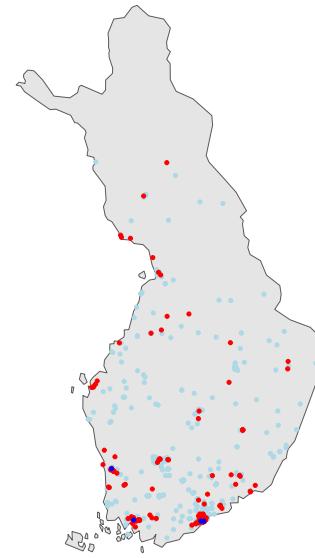
(e) Lithuania



(f) Latvia



(a) Netherlands



(b) Finland

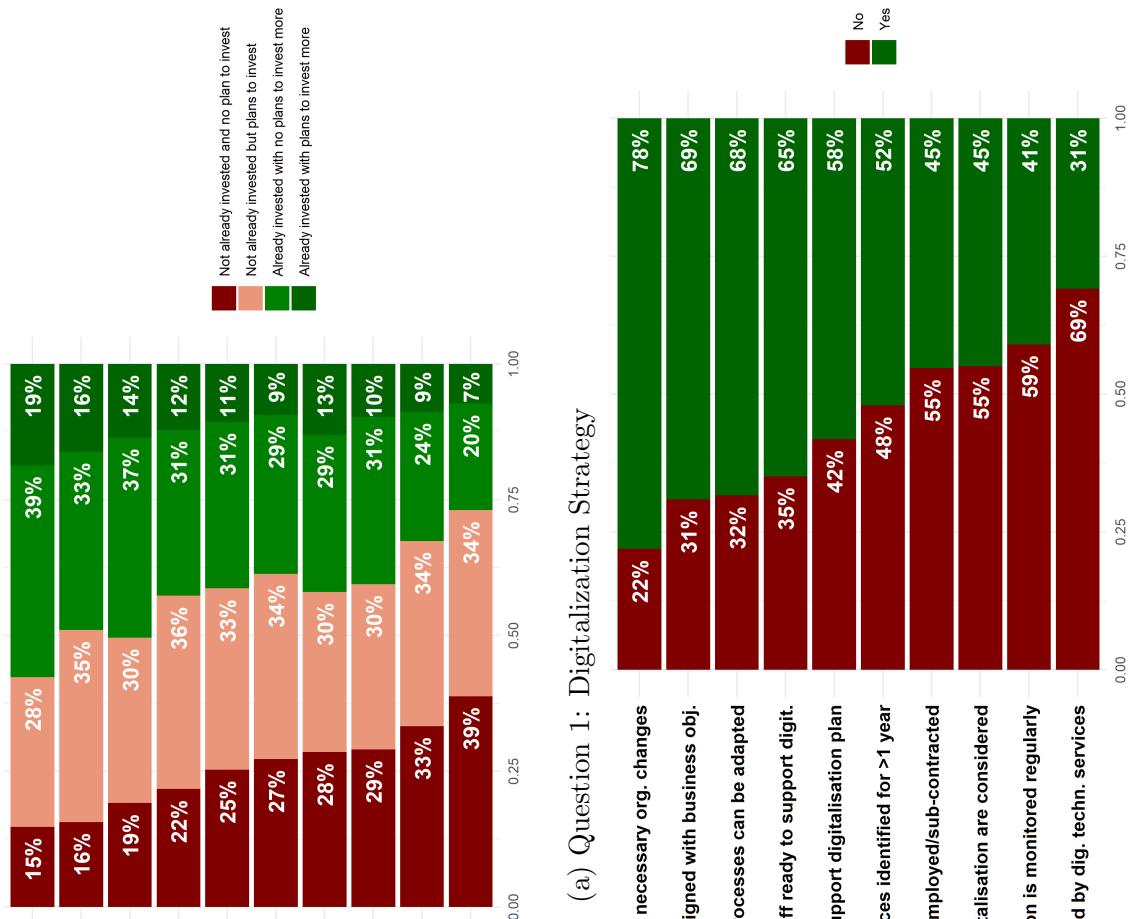
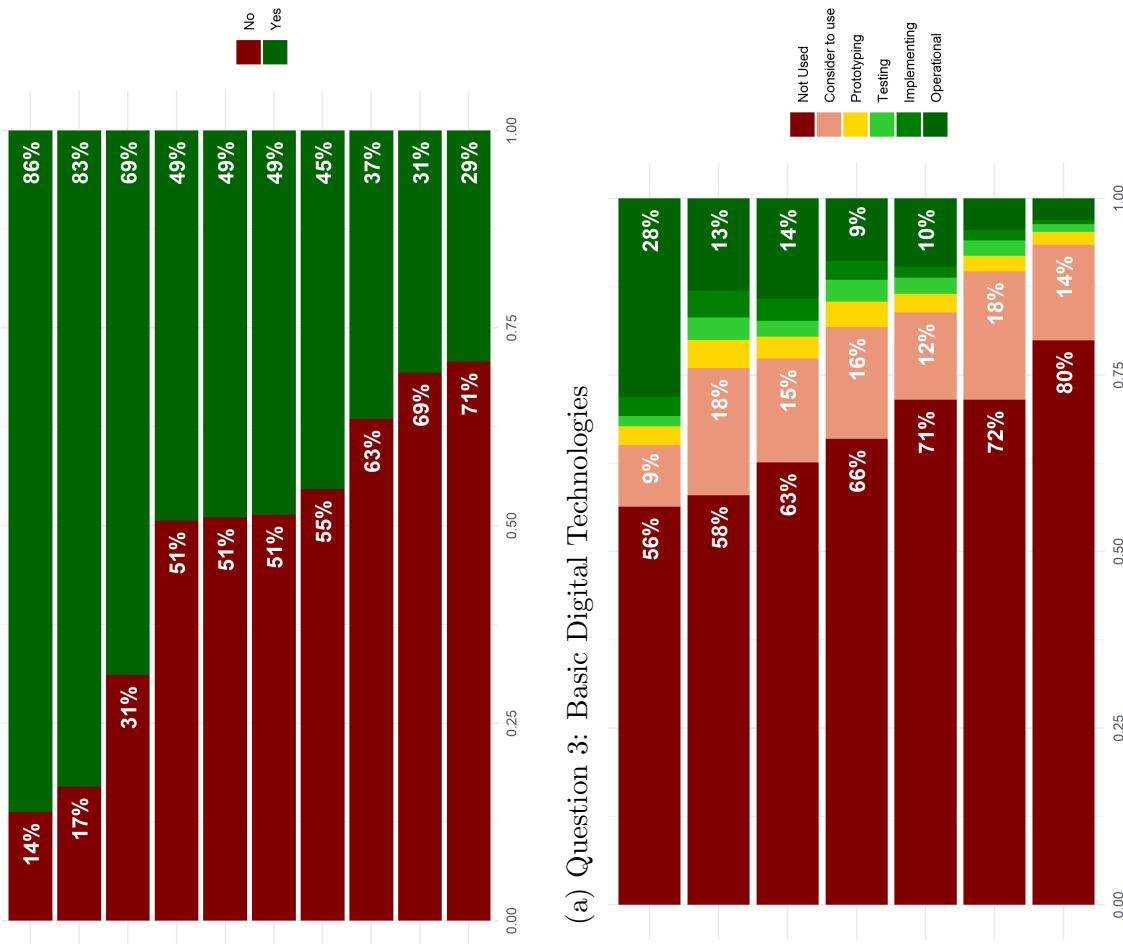
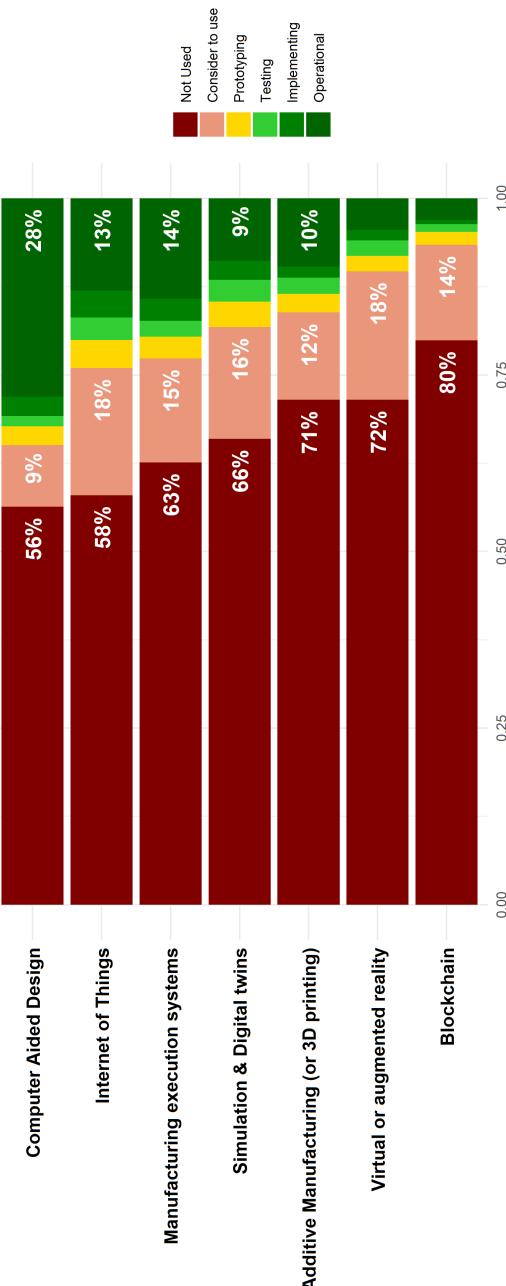


Figure A.8: Responses for Digital Business Strategy dimension



(a) Question 3: Basic Digital Technologies



(b) Question 4: Advanced Digital Technologies

Figure A.9: Responses for Digital Readiness dimension

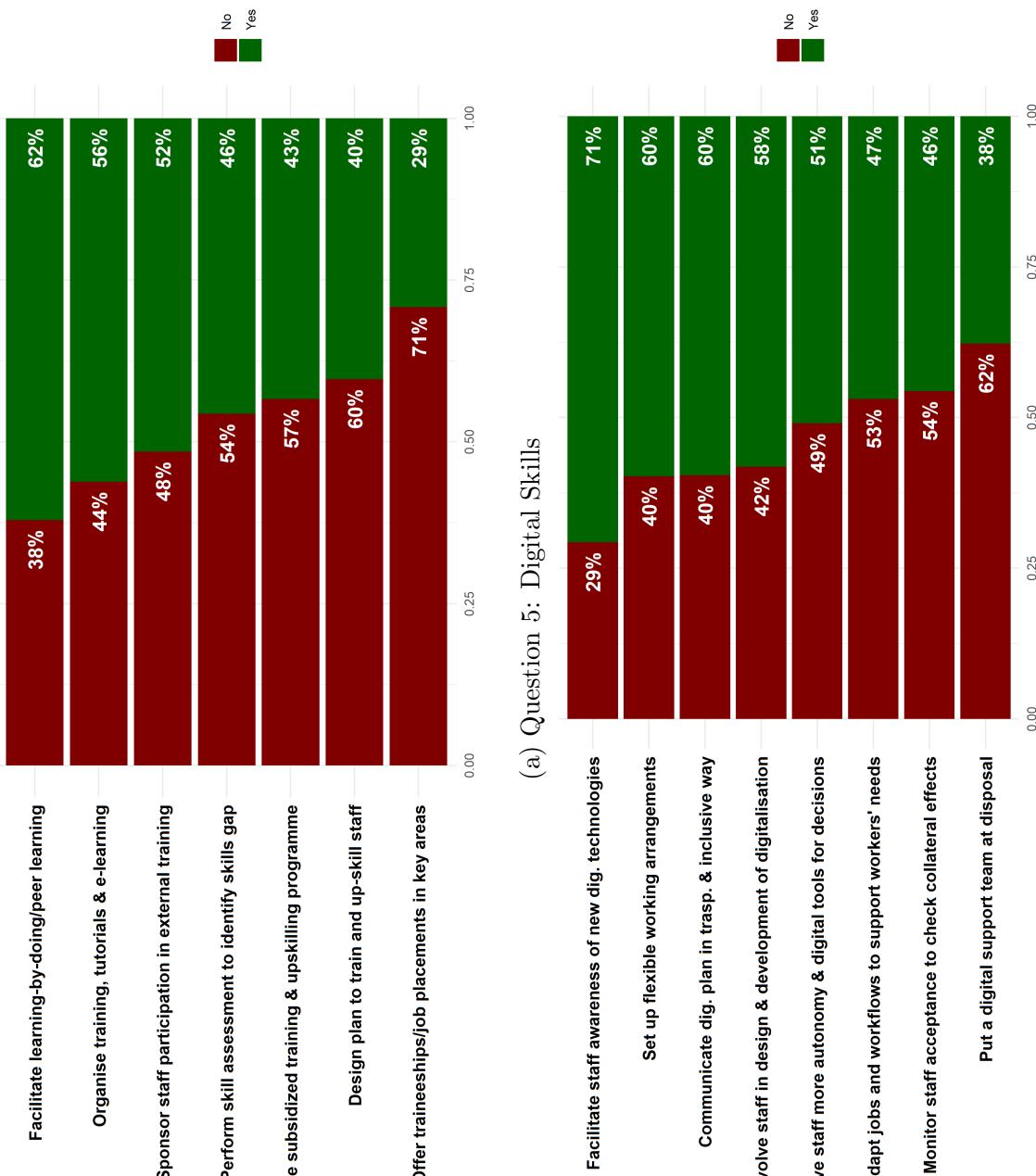


Figure A.10: Responses for Human-Centric Digitalization dimension

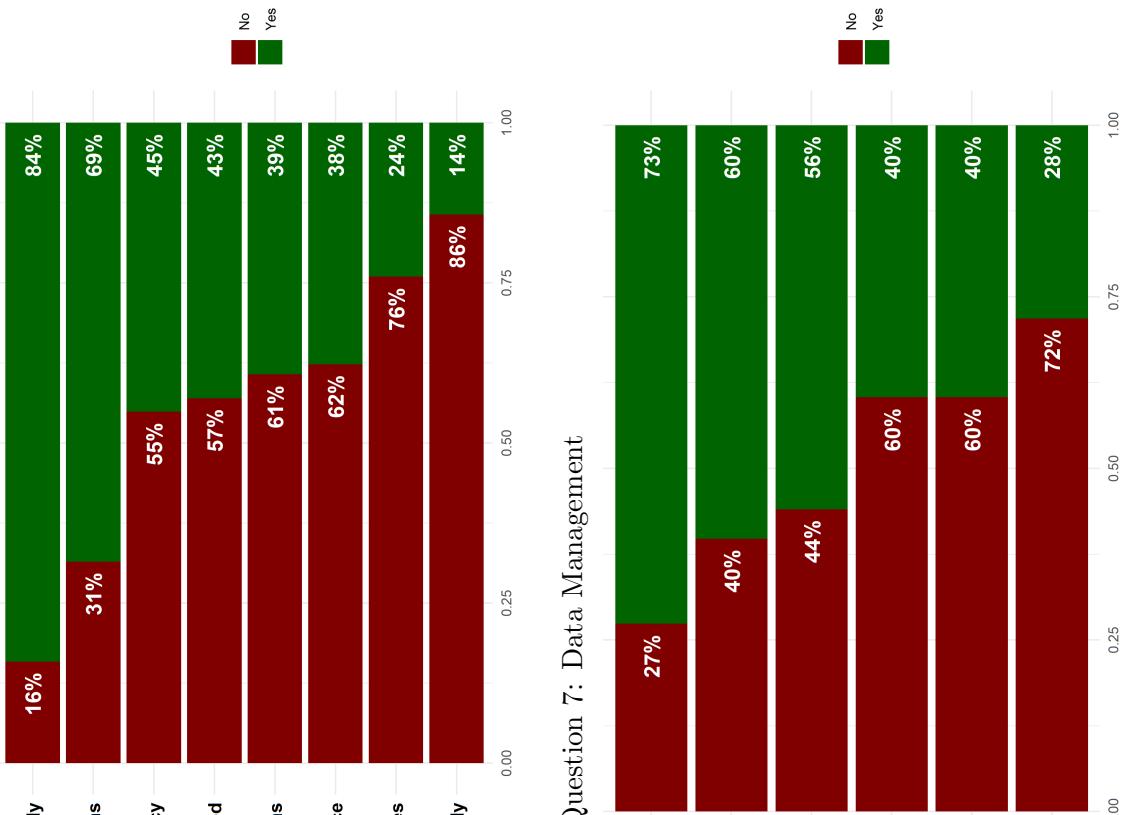


Figure A.11: Responses for Data Management dimension

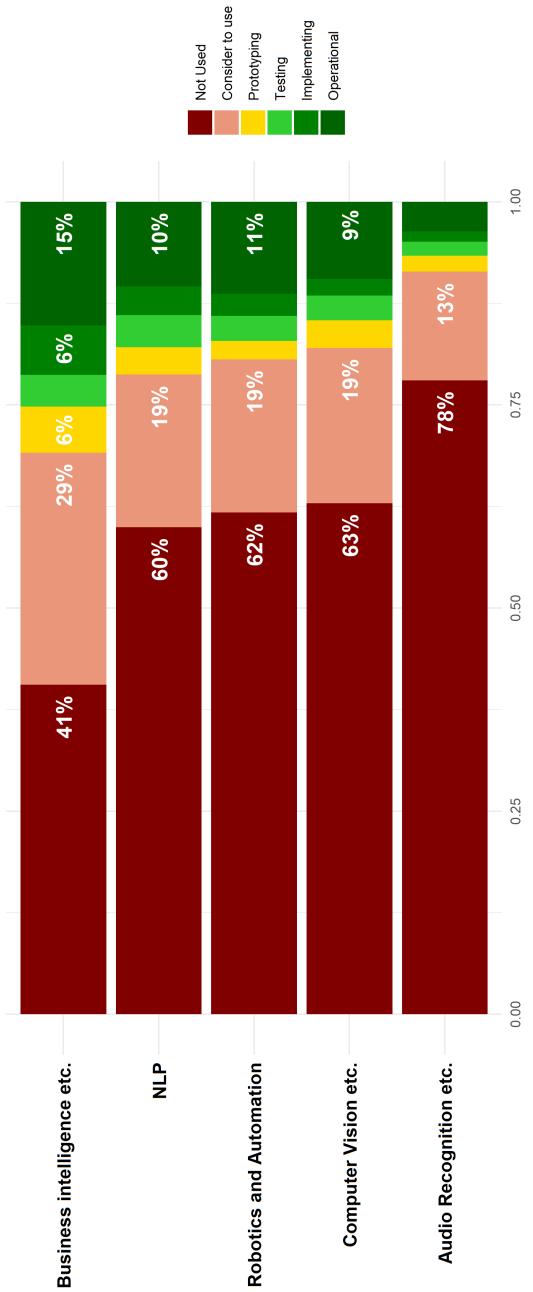


Figure A.12: Question 9: Automation & AI (Automation & Intelligence dimension)

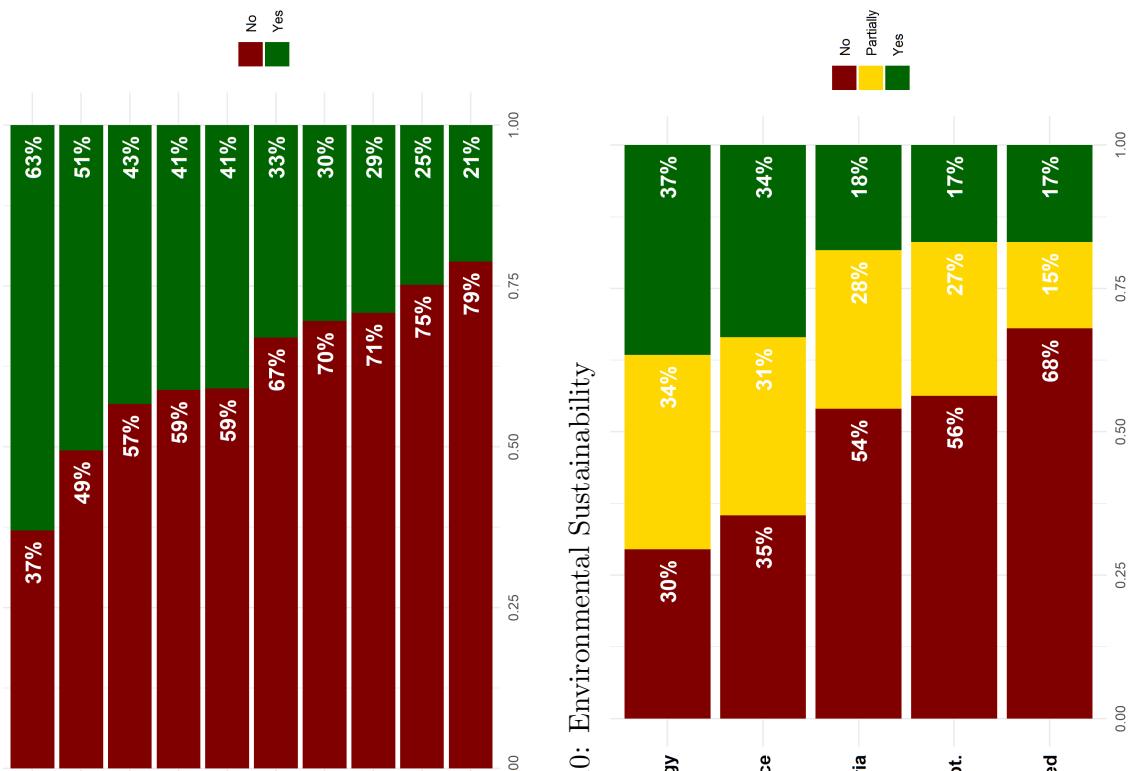


Figure A.13: Responses for Green Digitalization dimension

Bibliography

- Bloom, Nicholas, Charles I. Jones, John Van Reenen, and Michael Webb. 2020. “Are Ideas Getting Harder to Find?” *American Economic Review* 110, no. 4 (April): 1104–44.
- Carmona-Lavado Antonio, Cuevas-Rodríguez Gloria, Cabello-Medina Carmen, and Fedriani Eugenio M. 2021. “Does open innovation always work? The role of complementary assets.” *Technological Forecasting and Social Change*.
- Chesbrough, Henry. 2003. “Open innovation: The new imperative for creating and profiting from technology.” *Harvard Business School, 2003*.
- Crépon, Bruno, Emmanuel Duguet, and Jacques Mairesse. 1998. “Research, Innovation and Productivity: An Econometric Analysis at the Firm Level.” *Economics of Innovation and New Technology* 7 (2): 115–158.
- Czarnitzki, Dirk, and Katrin Hussinger. 2004. *The Link between R&D Subsidies, R&D Spending and Technological Performance*. Technical report 04-56. ZEW Discussion Paper.
- David, Paul A., Bronwyn H. Hall, and Andrew A. Toole. 2000. “Is public R&D a complement or substitute for private R&D? A review of the econometric evidence.” *Research Policy* 29 (4): 497–529.

De Nigris Sarah, Kalpaka Annita, and Nepelski Daniel. 2023. “Characteristics and regional coverage of the European Digital Innovation Hubs network.” *JRC Technical Report*.

De Nigris Sarah, Kalpaka Annita, Nepelski Daniel, and Rissola Gabriel. 2023. “Digital Maturity Assessment (DMA) Framework & Questionnaires for SMEs/PSOs - A guidance document for EDIHs.” *JRC Scientific Information Systems Databases Report*.

Draghi, Mario. 2024. *Report on the Future of European Competitiveness*. Presented to Ursula von der Leyen on September 9, 2024. European Commission.

European Commission. 2021. “DIGITAL Europe Programme.”

European Commission and DG CNECT. 2016. *Digitising European Industry*. Publications Office.

European Commission and DG CNECT. 2018. *Digitising European Industry - Progress so far, 2 years after the launch*. Publications Office.

European Commission, DG CNECT. 2021. “DIGITAL Europe – EDIH Work Programme 2021-2023.”

Galindo-Rueda, Fernando, and Fabrice Verger. 2016. “OECD Taxonomy of Economic Activities Based on R&D Intensity.” *OECD Science, Technology and Industry Working Papers* (Paris).

Idrissi Fakhreddine, Moulay Othman, and Yan Castonguay. 2024. “Scholar’s policy recommendations for open innovation in SMEs: a systematic literature review.” *Benchmarking: An International Journal* 30 (7): 2345–2368.

Juhász, Réka, Mara P. Squicciarini, and Nico Voigtländer. 2020. “Technology Adoption and Productivity Growth: Evidence from Industrialization in France.” Forthcoming in the *Journal of Political Economy*, *Journal of Political Economy*.

Mancusi, Maria Luisa, and Andrea Vezzulli. 2014. “R&D and Credit Rationing in SMEs.” *The Economic Inquiry*, 1153–1172.

Mazzocchi, Stefano. 2004. “Book Review - Open Innovation: The New Imperative For Creating and Profiting From Technology.” *Innovation: Management, Policy & Practice*.

Mina, Andrea, Di Minin Alberto, Martelli Irene, Testa Giuseppina, and Santoleri Pietro. 2021. “Public funding of innovation: Exploring applications and allocations of the European SME Instrument.” *Research Policy*.

OECD and European Commission. 2008. *Handbook on Constructing Composite Indicators: Methodology and User Guide*. OECD Manual. Paris: OECD Publishing, August.

Schumpeter, Joseph A. 1934. *The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest, and the Business Cycle*. Originally published in German as *Theorie der wirtschaftlichen Entwicklung* in 1911. Cambridge, MA: Harvard University Press.

Van Ark, Bart, Mary O’Mahony, and Marcel P. Timmer. 2008. “The Productivity Gap between Europe and the United States: Trends and Causes.” *Journal of Economic Perspectives* 22 (1): 25–44.