MSc Programme in Urban Management and Development

Rotterdam, the Netherlands

August 2024

Thesis title: Open data and economic growth in French regions: a Qualitative Comparative Analysis

Name: Juliette Navarre

Supervisor: Dr. Sofia Pagliarin

Specialisation: DIGEC

Report number: 1852

UMD2023-24

IHS Making cities work Erasmus

Institute for Housing and Urban Development Studies of Erasmus University Rotterdam

Summary

The emergence of open data in the past few years pointed out its multiple potential benefits for society, especially in the economic field where it can impact, among others, innovation and efficiency. Nevertheless, the complexity of open data lies in the fact that publishing data will not automatically yield benefits. To that extent, this research looks into the factors that could contribute to enhance open data economic value, not only through the publication process but also in the way it is used. It addresses open data from a holistic approach considering at the same time the publication process, the policy context and the user dimension. The conceptual framework illustrates this interplay between technological, social and institutional factors to enhance open data economic value. In fact, while a lot of focuses is made on the benefits of open data, considerations for the complex process of releasing and using it, is often missing. A focus is made on the regional scale to emphasize the importance of contextual considerations.

The research focuses on a sample of 17 French regions, including overseas ones, to ensure diversity and variations within the case selection. Through QCA methodology, the study aims at identifying under what conditions can open data contribute to regional economic growth, looking at the same time into potential important regional discrepancies. In order to explain the economic growth of a region, QCA looks into the maturity of its open data portal, the regional governance effort to promote open data, and the share of skilled users among the regional population. While a qualitative assessment is conducted to collect data about regional open data portals and digital governance policies, quantitative datasets are collected regarding the share of skilled population and the regional GDP. The use of fuzzy sets allows for a more precise report of regional differences, especially regarding open data portals and governance mechanisms.

The results emphasized the importance of the social factor as a high share of skilled population appears to be the only necessary and sufficient conditions to produce a high regional economic growth. Thus, the resulting regional discrepancies can mostly be explained by the weak digital inclusivity of certain regions, suggesting that the major policy recommendations could be to focus on education and awareness regarding digital use and tools.

Keywords

Open data, open data portal, regional governance, economic growth, QCA

Acknowledgements

I would like to express my sincere gratitude to my supervisor Dr. Pagliarin, for the support and guidance throughout my writing process. Her insightful feedback and encouragement have been highly appreciated and crucial for the completion of this master thesis.

I would also like to extend my thanks to my fellow classmates and friends, their constant valuable support really helped me along this journey.

Table of contents

List of Figures

List of Tables

Abbreviations

Abbreviation	Full form
IHS	Institute for Housing and Urban Development Studies
EU	European Union
OGD	Open Government Data
PSI	Public Sector Information
RIS	Regional Innovation System
QCA	Qualitative Comparative Analysis
GDP	Gross Domestic Product
fsQCA	Fuzzy-sets Qualitative Comparative Analysis
HRST	Human Resources in Science and Technology
IDF	Ile-de-France
ARA	Auvergne-Rhône-Alpes
PACA	Provence Alpes Côte d'Azur
PRI	Proportional Reduction in Inconsistency
INUS	Insufficient but Necessary element of a path, which is in itself Unnecessary but Sufficient for the outcome

1. Introduction

1.1. Introduction to open data

The administrative accountability of governmental institutions has always been a topic of interest in French history. Already long before the digital revolution, the Declaration of the Rights of Man and the Citizen of 1789 proclaimed this accountability principle as a fundamental pillar of French Constitution. Later, in 1978, the right of access to government-held data was reinforced by the law on Freedom of Access to Administrative Documents establishing the framework of this accessibility (Publications Office of the European Union, 2023). Since the emergence of the digital age in the 2000s, the provision of administrative information is mostly digitalized through the form of data publicly accessible, more commonly known as open government data (OGD), and in 2016, the legal provision of open data was consolidated by the Digital Republic Law. It facilitates the opening and the free flow of administrative data, targets consumers right in the new digital society and advocates for an inclusive access to digital services.
Not only for administrative information, private sector and non-official organizations also progressively chose to release their data, at first for transparency purposes. Nevertheless, despite the emergence of this open data trend globally, there is still no official definition of this term. Open Knowledge Foundation has defined open data as "data that people are free to use, re-use and redistribute — without any legal, technological or social restriction" (n.d.). Similarly, OECD argued that open data is a "non-discriminatory data access and sharing arrangements where data is machine-readable and can be accessed and shared free of charge and used by anyone for any purpose" (2023, p. 8). A variety of definitions co-exist, most of them highlighting the key principles of open data as being: availability, transparency and absence of restrictions.

As a trendsetter, public sector is often pointed out as a central creator and collector of open data. Often referred to as OGD, the data released specifically by public bodies, is an important resource in many different domains and a driver for generating substantial value in those domains (Jetzek, Avital, & Bjorn-Andersen, 2014). Unlike many people tend to think, nowadays open data is not only about publishing data on a website for transparency purposes, but also about data reuse for economic and social improvement. Indeed, open data generates social value through transparency and participation mechanisms (Jetzek et al., 2014). Openness lowers the barriers between public institutions and citizens; by improving information visibility, it engages more external stakeholders. Additionally, as highlighted by the World Bank Group, open data can have quite a leverage on economic growth by providing "the raw material for innovative new data-driven businesses" (2015, p. 9) and by fostering job creation, among others.

The focus of this research is on the economic value of open data, mostly through OGD but not only, as nowadays, open data platforms also gather datasets coming from various private sources.

1.2. Problem statement & Relevance of the research topic

The massive emergence of open data in the past few years have led several authors to point out the many economic benefits it could generate. Among them, Zhang & al. especially emphasizes the opportunities brought by OGD for commercial re-use and entrepreneurship (2022). The information is not detained anymore only by the public sector, but also accessible for the "collective intelligence of the public" (Janssen, Charalabidis, & Zuiderwijk, 2012, p. 260) including the private sector. Resulting from those findings, the questions of how to properly

use open data and enhance its economic value arose among the academic as well as the social field. To that extent, this research aims to provide some elements of a response.

While a lot of literature focuses on the benefits of open data, considerations for the complex process of publishing it is often missing. Indeed, the complexity of open data lies in the fact that publishing data will not automatically yield benefits. As stated by Janssen & al. "open data on its own has little intrinsic value; the value is created by its use. Supporting use should not be viewed as secondary to publicizing data" (2012, p. 264). Thus, open data economic value appears to be part of a wider ecosystem. Interdependencies between "data, open data providers, open data users, material infrastructures and institutions" (Zuiderwijk, Janssen, & Davis, 2014, p. 22) need to be addressed in the publication process. To the same extent, it appears that the geographical environment could influence open data impact as the parameters mentioned above (institutions, users, infrastructures) are highly contextual. For example, as stated by Raymond & Kouper regarding US states' data initiatives "the policy landscape is (…) vastly different at the state level" (2023, p. 63).

Thus, it seems that an integrated approach is necessary to capture the full picture of the stimulation of economic value through open data. The many frameworks that already exist often miss a parameter. For example, the Open Data Maturity Report, published by the European Commission every year, addresses the governance mechanisms and the publication infrastructures but it is mostly an assessment from a data provider perspective and the user dimension is missing. To the same extent, OECD OURdata Index focuses on different stakeholders' engagement but lacks depth in the considerations of the technical publication features.

This research seeks to address open data economic value from a holistic approach considering at the same time the publication process, the policy context and the user dimension. The value of open data is real, but it remains uncertain which specific ecosystem is needed to foster its benefits. By expanding upon existing literature, this study aims at examining certain factors and their interrelation to contribute to open data economic value. Also, the case of France stands out as French open data framework is already known to be effective at the country level (Publications Office of the European Union, 2023), the study will focus on the regional scale, as well to emphasize the importance of contextual considerations.

1.3. Research objectives

Based on the above introduction of the research, the main objective of this thesis is to explain the conditions, or configurations of conditions, required for open data to stimulate regional economic growth in France. To that purpose, the thesis aims to attach a great importance to the different regional contexts regarding open data to explain differences in regional economic weight.

Therefore, the research objectives are:

- Identify the regional ecosystem that is necessary to stimulate open data economic potential.
- Examine regional economic performance discrepancies.

1.4. Research question

Based on the above, the main research question on this thesis is the following:

Under what conditions can open data contribute to regional economic growth in France?

1.5. Thesis structure

The thesis is structured into five chapters. Following this introduction, Chapter 2 presents a literature review focusing on the definition of open data especially through the ecosystem necessary for its use and the ensuing value for society. The conceptual framework of the research is introduced and detailed. The third chapter introduces QCA as the research methodology chosen to answer the research question. The concepts presented in the literature review are operationalized into measurable components. In Chapter 4, the results of the research are presented along with an analysis and a discussion. The last chapter concludes the thesis with some final remarks about the use of open data in the stimulation of economic growth.

2. Literature review and hypotheses

2.1. Open data scope

Despite the growing global popularity of open data, many authors pointed out the lack of clarity in its definition and in the ensuing key terms used in the literature and in public debates (Bonina, 2013). Nevertheless, it is often highlighted that open data needs to be accessible, without any restriction for use and re-use, free of charge and non-discriminatory (Bonina, 2013; Braunschweig, Eberius, Thiele, & Lehner, 2012; Hossain, Dwivedi, & Rana, 2016; Janssen et al., 2012; Kitchin, 2014; Welle Donker & van Loenen, 2017). As an example, Braunschweig & al. described the general understanding of open data as "data that is freely available and can be used as well as republished by everyone without restrictions from copyright or patents" (2012, p. 1). In that sense, private and confidential data are excluded (Janssen et al., 2012).

Open data exists in a lot of different forms and can come from a lot of different sources, which can bring some confusions in the academic field. As highlighted by Bonina, open data can range from small to medium and big data, and can be produced by government as well as private firms (2013). As most of open data is released by public bodies, some ambiguities can emerge between the wider open data term and open government data. Miletic & al. rightly emphasized that "public organizations produce and commission large amounts of data and information, which, when made available to citizens for further reuse is referred to as Open Government Data (OGD)" (2023, p. 1). So, a distinction is to be made between OGD, as a subset of open data, and open data itself, which can be sometimes a subset of big data *(Figure 1)*. Similarly, 'open private data' exist as the intersection between private sector information and open data. This research will consider all kind of open data, even if a focus will be made on OGD as "the government is the largest collector of data regarding aspects of economic and social life" (Bonina, 2013, p. 5). In fact, already in 2014, Jetzek & al. highlighted that "over 280 government data catalogs have been published and over a million datasets have been released by governments around the world" (2014, p. 101).

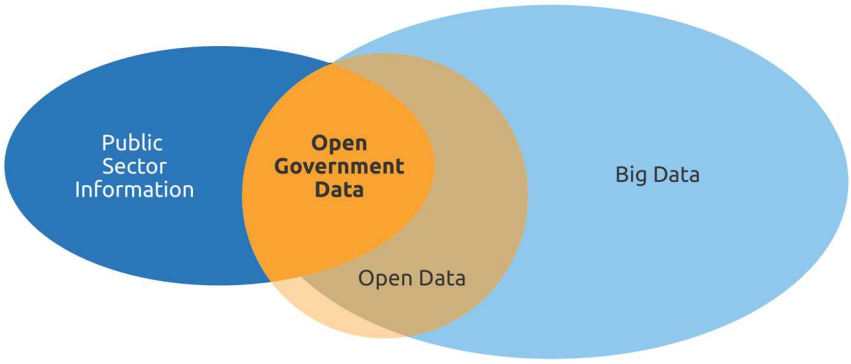


Figure 1: Scope of open data, from European Data Portal (2015)

The range of open data covers a lot of different types and domains: "primary (census data) or secondary (economic trend), real-time (such as traffic or weather data) or offline (government spending), location-based (toxic waste dumps) or generic (regional healthcare costs), reports, maps, satellite photographs, pictures and paintings, the genome, medical data, scientific formula, public sector budgeting, food-safety information, and so forth" (Hossain et al., 2016, p. 15). Likewise, Jetzek & al. emphasized how society has become dependent over data and open data, especially in application such as weather forecast or traffic (2014). This variety of application can explain the popularity encountered by open data in the past years as the data thus released benefits and creates value for a lot of different actors.

Nevertheless, traditionally data and, in a broader sense, knowledge have been closed or with restricted access, kept by the public institutions, which resulted in a certain historical frustration from civil society (Kitchin, 2014). Open data movement changed this mainstream boundary between public organizations and the general public by transitioning from a closed to an open system (Janssen et al., 2012). By definition, closed systems are easier to manage because they are unlikely to get affected by external factors and in that sense, they are predictable. Conversely, open systems enable free flow of information and are unpredictable, they allow additional views to be heard. Especially for open government data, "the opening of systems provides the opportunity for creating feedback loops in which the government can learn from the public" (2012, p. 259). This system theory argued by Janssen & al. introduced a complexity-informed framework for open data by highlighting the new interplay between governance instruments and user perspectives. To be fully adopted open data systems need more than provision and access to data, especially they require institutional measures, infrastructures and public engagement (2012).

The argument that public institutions (as well as private firms) efforts, on their own, to release data is not enough is shared by a lot of different authors. Among others, Jetzek & al. stated that "a common assumption when opening government data is that simply supplying more data freely and in more formats will lead to more use and value creation" (2014, p. 101). Similarly, Bonina emphasized that taking advantage of open data is not straightforward and represent a key challenge for governance structures (2013).

This first section introduces the central thread of this thesis: open data have value only if it is not only published but also used. How different authors have explored this argument and how the academic field have looked into enhancing open data value especially for economic purposes is going to be detailed along the Chapter.

2.2. Open data use, a contextual matter

For a better understanding of the academic argument that open data have value only if it is used, it is also important to understand the difference between data publication and data re-use. As defined by The guide to Open Data Maturity Model, data publication "addresses the organisational activities and processes that support the creation and management of datasets that are made accessible under an open licence", while data re-use "is concerned with the processes that support the effective reuse of third-party datasets" (Dodds & Newman, 2015, p. 12). In other words, considering only data publication means focusing on data supplier regardless of the user needs (Janssen et al., 2012).
From this observation, a lot of literature has sought to investigate open data ecosystem, meaning the factors necessary to enhance the use of public data, including as we just saw the supplier and the user.

In 2011, Vickery introduced the notion of 'PSI value chain' detailing the steps required for commercial re-use of public sector information (PSI) from the data creation to the delivery of a 'high-value' product, meaning a product ready to be re-used. The steps include the collection of the data, a proper organization of this raw data for storage and retrieval, processing the data and any editorial activities (for example the mapping of geo-spatial data), lastly the marketing and distribution of the edited data (2011). Thus, this value creation framework mostly emphasized the need for proper optimized ICT infrastructures providing the data, in order to foster open PSI commercial re-use. This observation remains on open data agenda as Raymond & Kouper pointed out in 2023 that "open data portals do not receive sufficient attention beyond simple technical implementation" (2023, p. 59). Similarly, Zuiderwijk & al. highlight the

importance of those ‘digital infrastructures'; but they nuanced Vickery’s point by adding that they are nevertheless part of a wider open data ecosystem, “characterized by multiple interdependent socio-technical levels (…) and components” (2014, p. 23), including among others policy, financing, legal framework and interactions between data suppliers and users. Thus Zuiderwijk & al. argued that some essential contextual elements are to be considered to enable easy publication and use of open data. Welle Donker & van Loenen also explored the notion of open data ecosystem and developed a holistic approach to the data release challenge, as they observed that the existing frameworks “only deal with single components of the open data ecosystem” (2017, p. 287). Alike Zuiderwijk & al., they emphasized that a well-functioning open data ecosystem must cover a technical, a legal and an organisational dimension, and hence decided to assess open data through the lens of the data supplier, the data governance and the user characteristics (2017).

Consequently, it appears that context in which open data is released seems to highly facilitate an efficient re-use: who provides the data, how is it provided and who is using it, matters. To understand how differences within those contextual elements could explain the gap between a successful use of open data and a poor one, a focus on a localised scale, especially regions, seems particularly relevant. In fact, those factors are significantly fluctuating depending on the regional environment. Also, the research question focuses on the stimulation of economic growth through open data use, and it appears that a lot of authors have pointed out the many advantages of regional scale focus regarding economic activities.

The combination of diverse regional specificities and experiences makes regions unique environments for the development of economic activities and thus is a proper scale to study differences in economic performance (Gössling & Rutten, 2007). Also, as stated by Krakowiak-Bal & Ziemiańczyk, “the heterogeneity across regions in their capacity to create knowledge and innovation, and also in their abilities to exploit ideas and technologies available across the European territory, motivates in-depth analyses of the territorial dimension of the knowledge economy” (2017, p. 8). To the same extent, Asheim & Gertler emphasized the importance of subnational level by pointing out that the circulation of new knowledge is highly localised (2005). The more attractive and dynamic locations (from an innovative perspective) will benefit from a rich local labour market and attract high-skills workers with good career opportunities, and thus engage in a virtuous cercle of growth at the expense of others. This reinforcing positive feedback loop between high-skilled workers and localized innovative potential can boost regional economic growth also through the use open data.

2.3. Conceptual framework

Based on the assertion, developed earlier in the literature review, that "the power of open data relies on its use" (Bonina, 2013, p. 19) and on the statement that different contextual elements matters in re-use processes, the research is based on the following conceptual framework:

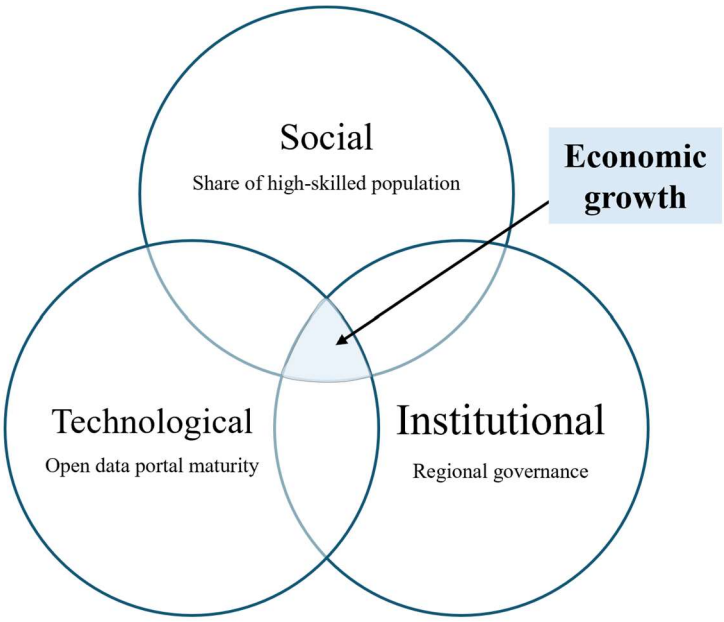


Figure 2: Open data ecosystem for efficient re-use

Data governance and the influence of institutions

First of all, data governance is at the forefront of data publication and use. In fact, public bodies responsibility regarding open data goes beyond providing the data in an efficient way, it is also to foster its access and usage through a supporting policy framework. As explained by OECD in its OURdata Index, securing open data benefits in the long run requires government's support through specific and sustainable policies (2023). It also emphasizes the necessity of engaging with stakeholders (mostly users) to fully realise the benefits of open data. In fact, the massive flow of information brought by digitalization impacted governance and "reinvented the way governments, public organisations and businesses interact with each other and with the public" (Vickery, 2011, p. 6). Besides, institutional differences can create significant gap between regions and consequently can partly explain discrepancies in regional economic performance. As stated by Boschma, "institutional rigidity leaves no room for experiments with new institutions that are required for the successful implementation of new ideas and innovations" (2005, p. 68). In other words, he explains that the effectiveness of institutional structures lies in a proper balance between stability, openness and flexibility specific to each region. As an example, Raymond & Kouper pointed out regarding US states' data initiatives that "the policy landscape is (…) vastly different at the state level" (2023, p. 63), consequently it can result in important discrepancies in the open data regional framework.

Data supply through performant technological infrastructures

Secondly, data supply, in other words the technological infrastructures providing the data, is clearly a significant component of data use. As stated by European Commission, public data as a common good is like an infrastructure, it is only an enabler for the economy, and to maximize its potential, it needs to be provided thoughtfully (2020b), most of the time the provider being

public bodies through OGD. Indeed, Zhang & al. highlighted that the source of the data including the data infrastructures is often a barrier to commercial re-use of open government data (2022). Likewise, Janssen & al. pointed out that "the existing structures are taken as a starting point, and the user needs for finding, processing, and using open data is neglected" in the process of data release (2012, p. 261). Consequently, ensuring strong and well-designed data infrastructures is a priority in open data provision to foster the re-use of the data. Those technical infrastructures are the representation of the degree of openness of the data by demonstrating the availability, the accessibility and the use of open data (Jetzek et al., 2014). To that extent, Zuiderwijk & al. highlighted that open data portals have been the focus of European and national bodies to explore the potential of open data (2014). At the regional level, Krakowiak-Bal & Ziemiańczyk indicated that "infrastructural support is important to build innovative advantage in a region" (2017, p. 12).

The importance of users' characteristics

Lastly, Jetzek & al. argue that achieving economic value from opening up data will require not only motivation (through governance) but also skills from public bodies as well as private organizations and citizens, and especially state that the "lack of appropriate governance mechanisms and a lack of insight into user's perspective can explain the gap between the promises of open data and what has actually been realized" (2014, p. 101). Thus, the social dimension is to be considered, Janssen & al. talked about the myth that every constituent can make use of open data, highlighting that not everyone has the resources or the expertise to make use of open data (2012). To the same extent, Zhang & al. also explained that besides the lack of skills, even "the mere usage of OGD (Open Government Data) is insufficient to ensure innovation" (2022, p. 393). As well as for infrastructures, the social dimension can influence regional economic performance, "if a region is rich in technology, in demanding customers and has an educated workforce, it will be easier and more rewarding for firms to develop new products and services" (Gössling & Rutten, 2007, p. 254). Also, relations, norms, values and interaction within a community can reinforce regional innovative capability (Doloreux & Parto, 2004).

Interdependencies

The literature shows that institutional, technological and social drivers are essential, not only on their own but above all as interdependent components. As highlighted by Zuiderwijk & al., the interdependencies between "data, open data providers, open data users, material infrastructures and institutions" (2014, p. 22) must be considered in the publication of open data. Similarly, Bonina reminded that "making businesses and creating competitive advantages from open resources is not straightforward and represent a key challenge when it comes to turn open data into a profitable business" (2013, p. 11).

The open data economic potential is wide, and to enhance its full value, it is necessary to regard open data through an integrated approach and to consider those potential benefits as the result of a combination of different interconnected and interdependent factors (Welle Donker & van Loenen, 2017). In other words, the economic growth that could result from data opening needs to be approached through a multi-dimensions framework considering the full picture of open data release complexity *(Figure 2)*.

2.4. Open data value, which benefits for society?

Making data available for the general public through open data placed the latter as a public good. In other words, according to the taxonomy of goods it is considered as a non-excludable and a non-rivalrous good. This argument is supported by Raymond & Kouper who stated that "data likely should be considered digital public goods, non rivalrous and non excludable resources that can be created and distributed without becoming depleted for free or a at low cost" (2023, p. 59). Especially, digitalization makes data non-rivalrous because it becomes available by multiple persons at a time, and openness makes it non-excludable as a shared resource, a 'digital commons' (Jetzek et al., 2014). Jetzek & al. also argued, to the same extent, that openness is the exact characteristic of open data that makes it valuable "as the antidote that can counteract the tendency of technology enactment to reproduce existing rules, routines, norms and power relations" (2014, p. 104). Similarly, Zhang & al. define data as a 'new gold mine' for the economy and explained that especially "open sharing (…) stimulate data value to create circumstances and provide opportunities for public entrepreneurship to develop commercial applications" (2022, p. 392). He even emphasized the role of European Union in encouraging the public, and especially businesses, to use public data to create value through the adoption of policies.

Thus, the value generated by the use of open data is estimated to have an impact on different domains of society, especially social (including political) and economic impacts. Even if the social side is not the focus of this thesis, it is important to grasp the implications at stake. Regarding social benefits, they mostly concern transparency and participation; for the economic ones they include innovation and efficiency. Among others, Kitchin highlighted that "open data create transparency and accountability; participation, choice and social innovation; efficiency, productivity and enhanced governance; economic innovation and wealth creation" (2014, p. 66). Likewise, Braunschweig & al. stated that opening data would ideally "lead to more transparency, participation and innovation throughout society" (2012, p. 1).

Regarding open data social value, Jetzek & al. defined it as "the generated improvements in the lives of individuals or society as a whole" (2014, p. 104). The many authors that have worked on the subject tend to agree on those improvements as transparency and participation, as already mentioned, but also trust and accountability. For example, Janssen & al. wrote that "sharing data openly and freely is often viewed as altruistic and advancing transparency and knowledge" (2012, p. 260). In fact, the quest for transparency fostered many open government initiatives, as it is often sought to strengthen democracy and fight corruption, at the same time contributing to the accountability of public bodies (Bonina, 2013). Governmental entities also seek for promoting participatory democracy by opening their data, encouraging informed participation and active citizenship in the decision-making environment (Kitchin, 2014). Not only the academic field but also public organisms recognized the impact of open data in governance mechanisms. As an example, the World Bank pointed out that "open data plays a critical role in improving governance by exposing and preventing mismanagement and corruption and promoting innovation in public administration" (2015, p. 15).

Alike social value, a lot has been researched on the potential economic value open data could generate. A lot of literatures of the past few years have sought to demonstrate the positive impact of open data on economy (Welle Donker & van Loenen, 2017). Vickery's report on PSI re-use estimated the value of open data re-use market to 28 billion euros (2011), thus corroborating the 27 billion euros found few years earlier by the MESPIR report (2006). More recently in 2020, European Union published a report on value creation of open data in Europe, estimating the market size of open data, meaning "the market size of products, services, and

content improved or enabled by open data", between 200 billion euros and 334 billion euros by 2025 (Publications Office of the European Union, 2020b, p. 18).
Those impressive figures are mostly explained by the massive potential for commercial re-use generated by releasing data. Vickery emphasized that PSI open access could increase the crossing between public and private sector information, and its use could enhance the development of new products, as a source of new knowledge and a raw material for innovation (2011). Similarly, Kitchin argued that "publicly generated data can be used to add value to existing business data, create new applications and services and thus new markets, and improve business knowledge and decision-making" (2014, p. 56). Thus, the many authors that have studied the re-use of open data, and mostly open PSI, tend to agree on its major economic benefits for society as being especially the stimulation of innovation, efficiency gain, job creation and cost savings, all of them resulting in economic growth. Miletic & al. highlighted the facilitation of innovative digital products, the creation of new jobs and the increase in value creation (2023). Caldag & Gökalp focused on OGD and emphasized its value for entrepreneurship and new job opportunities, also improving operational capacity and cost reduction (2022). Similarly, Zhang & al. investigated the success of entrepreneurial initiatives based on OGD and demonstrated that, under certain conditions, OGD can effectively stimulate the success of entrepreneurs even without a massive capital investment (2022).
Also, the idea that regions can be assimilated to engines of innovation especially relates to the concept of Regional Innovation System (RIS). RIS emerged in the 1990s to emphasize territorial specificities, and so the importance of regional scale and local resources in the innovative performance of regions (Krakowiak-Bal & Ziemiańczyk, 2017). Indeed, the strong systemic bond between firms and regions' knowledge infrastructure can reinforce localized learning within a region and support emerging economic activities (Asheim & Gertler, 2005). In other words, regional development "occurs in places where localized capabilities such as institutional endowment, built structures, knowledge and skills exist" (Doloreux & Parto, 2004, p. 2). Thus, RIS literature tends to emphasize the role of social and institutional conditions to explain regional economic competitiveness (Doloreux & Parto, 2004; Krakowiak-Bal & Ziemiańczyk, 2017), similarly to the open data ecosystem detailed in the previous section.

Open data not only profits entrepreneurs but can also be beneficial for a wide range of stakeholders with different appropriation opportunities: "established companies that may find new revenues from providing analysed data, individual developers and start-ups producing software applications for mobile and internet-based devices, customers benefiting from better services, journalists, educators and researchers that may use open data on their work" (Bonina, 2013, p. 5).

Thus, the value generated from open data use and the ensuing benefits covers a wide range of opportunities for both the social and the economic field. To that extent, Jetzek & al. developed an interesting framework showing possible value generation strategies to stimulate social or economic benefits from OGD *(Figure 3)*.

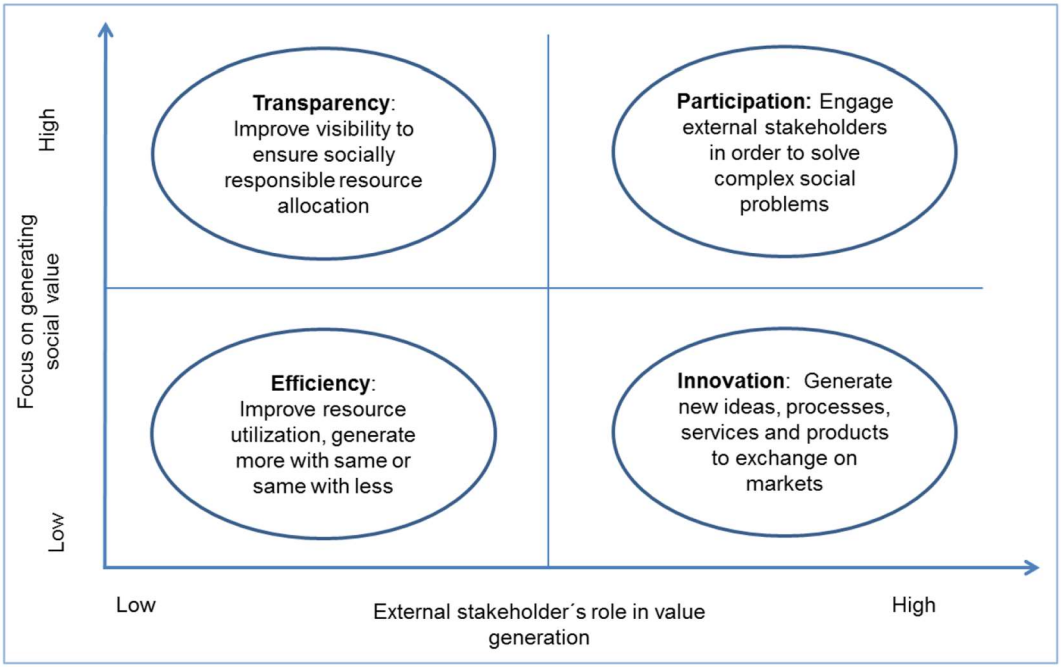


Figure 3: The OGD value generation framework, from Jetzek & al. (2014)

This framework shows not only an overview of the range of open data value, as detailed in this section, but also confirms that different parameters must be considered in the value generation mechanisms. Especially, Jetzek & al. highlight that the framework can be used "to explain the complex relationship between openness, data and value" (2014, p. 116), as well as the role of stakeholders.

2.5. Research hypothesis

This study takes part in the open data trend of the past few years but focuses especially on economic growth as the result of a combination of different factors previously mentioned. It is emphasized that those factors are regarded through an integrated approach, as a whole regional ecosystem influencing the outcome. Also, in this framework and in the literature in general, a lot of emphasize has been put on the user as a key element of the re-use chain. So, based on the literature review, this thesis will investigate the following hypothesis:

- H1: The combining effect of a mature open data portal, an important institutional regional governance effort to promote open data, and a high share of high-skilled users contribute to the economic growth of a region.

- H2: The presence of high-skilled users is a necessary condition for a high regional economic growth.

All hypotheses address the research question in finding out the conditions required for open data to contribute to regional economic growth in France.

3. Methodology & Research design

3.1. Description of the research method

The research focuses on open data and aims at explaining the conditions contributing to regional economic growth. Qualitative Comparative Analysis (QCA), as an explorative method, has been chosen to conduct the thesis and answer the research question.

This method was invented by Charles C. Ragin, who defines it as an "analytic technique designed specifically for the study of cases as configurations of aspects, conceived as combinations of set memberships" (Ragin, 1999, p. 1225). QCA as a configurational approach studies the different combinations of causal conditions linked to a selected outcome. Thus, the method "allows for an understanding of the differential driving mechanisms that lead to the same outcome and further reveals the complementary and substitutional relations" between chosen factors (Zhang et al., 2022, p. 394). The main argument of QCA use is that it is rooted between a quantitative and a qualitative research method, and thus is presented as "a middle ground between the two strategies of depth and breadth" (Ragin, 2000, p. 22). This diversity-oriented approach allows each case to be examined as a unique entity with its specific characteristics and dynamic, thus proposing a balanced approach between case-specific details and broader patterns (Ragin, 2000). Also, QCA bridges the gap between generality and complexity. Gerrits & Verweij define QCA as a "complexity-informed research method that mediates between the in-depth understanding of complex cases and knowledge of context-bound generality" (2016, p. 10). Similarly, Befani argues that QCA overcomes the trade-off between a fine-grained understanding of cases and the ability to generalize by "allowing the researcher to identify moderately complex causal relations while offering the opportunity to generalize and synthesize information from a small, medium, or large number of cases" (2013, p. 269). To that extent, QCA presents multiple advantages to cover the objectives of this research.

Regarding the research framework, our sample, meaning the cases studied are the French regions. We have 3 causal conditions: the maturity of the regional open data portal (technological factor), the regional governance effort to promote open data (institutional factor), and the share of high-skilled population (social factor), as stated in the research objectives and explored in the literature review. Cases are unique configuration of these conditions producing a certain outcome, in our case economic growth measured by the regional Gross Domestic Product (GDP). Through QCA, this research aims at highlighting the different pathways to this outcome, since regions may seem similar in surface but contextual discrepancies between them could explain the difference in regional economic growth. The importance of context in QCA is often emphasized, Gerrits & Verweij stated that "context is explanatory for how cases as systemic wholes emerge and develop in certain (convergent or divergent) directions over time, even if they appear similar" (2016, p. 9). In other words, QCA can help highlighting that similar conditions can bring a different outcome (multifinality) or, on the contrary, that different conditions can bring similar outcome (equifinality). Similarly, Befani pointed out that "the effect changes when the context changes because different mechanisms are activated within it" (2013, p. 274). Besides, QCA can point out conjunctural causations, meaning that it can emphasized the effect of a condition in the context of the other conditions (Gerrits & Verweij, 2018c). Thus, by capturing nuances and variations in data, the diversity-oriented approach of QCA deals with this causal complexity and helps understanding how the different conditions interact and combine to produce specific outcomes (Gerrits & Verweij, 2018a; Ragin, 2000).

Different QCA types co-exist, for the purpose of this research we will be using fuzzy-sets QCA (fsQCA), meaning that in the calibration process fuzzy-sets membership values will be

attributed to the data collected. Fuzzy-sets QCA allows for a more fine-grained description of conditions (Gerrits & Verweij, 2016) because multiple values can be attributed instead of just fully-in or fully-out as in crisp-sets QCA. In our case, it especially allows to study more rigorously how the different open data portals and regional governance are managed.

The sample selection has been made to include as many French regions as possible in order to insure a certain diversity and variations within the cases. In that sense, out of 18 French regions, this research looks at 17 of them (including some overseas ones). The missing one, Mayotte, has been excluded due to the lack of data regarding its employment sector. The rest of the regions has been kept to include a broad variety of regional characteristics. For example, some of the overseas regions don't have a regional open data portal, some regions have a very low share skilled users compared to others. Reversely, some regions can be very similar. As explained by Gerrits & Verweij, "the comparison of cases where each case is considered as a fully unique data point would not bring any clear insights into the broader patterns across the cases" (2018b, p. 54).

The data collection methodology is presented in *Table 1*.

Condition	Type of data collection	Explanation
Open data portal maturity	Qualitative assessment of the researcher	Assessment of each regional open data portal based on different indicators (cf. section 3.2)
Regional governance effort	Qualitative assessment of the researcher	Assessment of each regional involvement based on different indicators (cf. section 3.2)
Share of high-skilled users	Quantitative data available	Data collected from an online public dataset of Eurostat. Data from 2022
Regional economic growth through GDP (outcome)	Quantitative data available	Data collected from an online public dataset of Insee.fr. Data from 2022

Table 1: Data collection methods used for the research

The research is going to be conducted using R and RStudio softwares to run the model, using also different QCA packages. The results provided by the software regarding the pathways to the outcome and possibly the necessity or sufficiency of certain conditions will be detailed and analysed in Chapter 4.

By covering most of French regions, the thesis considers a lot of different contexts and pathways to the outcome in the sample, thus reinforcing the ability to generalize the results. Also, as stated by Befani, like statistical methods QCA is suitable for generalization of the results either for a small, medium or large number of cases partly because it deals with the causal complexity mentioned earlier (2013).

3.2. Operationalization

Based on the conceptual framework proposed in Chapter 2, the research is going to explore 3 conditions that could contribute to regional economic growth. In this section, we are going to detail the choices made for each condition and the outcome included in the QCA analysis in order to attribute set membership values.

From here, the research will assume that, because one condition focuses on open data portals, the data mentioned is already publicly available, openly licenced, and free of charge.

3.2.1. Economic growth - *Outcome*

The research focuses on explaining how open data could contribute to economic growth through different conditions.

The regional economic growth is going to be operationalized and measured through the regional GDP. It is the most common used indicator to measure the economic performance of an area (country, region, etc.). In the case of regional GDP, it measures the regional economy's output for a given period and represents the value of all goods and services produced over a specific time period within a region. In other words, it is a tracker for the health of a region's economy (Kramer, 2024).

The research will use the Insee dataset: 'Produit Intérieur Brut en 2022'. The dataset provides a comparison of French regional GDP in 2022.

3.2.2. Open data portal maturity – *Condition 1*

Open data portals represent the main technological infrastructure available between the producer and the user of open data. They are defined as "websites set up by public sector bodies that serve as catalogues to support the discoverability of public data resources" by the Open Data Maturity Report (2023, p. 90). Welle Donker & van Loenen explained that "the chance that data are discovered may increase if the data are published in a well-known and accessible portal. Government information portals have been around for several decades, however, these are often poorly stocked, obsolete, and particularly user-unfriendly" (2017, p. 289). Thus, they are key elements in open data use to stimulate economic growth, as already mentioned in Chapter 2. As stated by Kitchin, open data movement should not only seek to make data available for wider reuse but also to provide "easy-to-use research tools that negate the need for specialist analytic skills" (2014, p. 48). To the same extent, Zuiderwijk & al. added that various tools and applications can bridge this gap and support data reuse. Especially, they emphasized that the better the functionalities of open data portals, the easier it is to stimulate the provision and use of open data and the realization of its advantages (2014).

Consequently, this condition aims at exploring the degree of open data portals maturity, meaning the way data is released by every region to ensure an efficient re-use. This first condition is measured based on 3 indicators: the provision of the data, the metadata available, the features provided on the portal *(Table 2)*.

	Basic (1 pt)	Standard (2 pts)	Advanced (3 pts)	Expert (4pts)
Provision	Limited number of datasets available. Few providers, exclusively from the public sector.	More than 200 datasets. Multiple providers.	More than 2000 datasets. Multiple providers. Provision of real-time datasets OR presence of private providers (or other non-official sources).	More than 2000 datasets including real-time ones. AND a lot of different providers from both public and private sectors (or other non-official sources).
Metadata	Tittle of the dataset, name of the provider.	Description of the dataset, release and/or last update, geographical coverage.	Advanced information on the dataset, analytics about its popularity.	User oriented information: recommendation for similar datasets, ratings, direct contact of the providers.
Features	Research section on the portal, at least one downloadable format per dataset.	Multiple research filters, multiple downloadable formats.	Possibility to visualize the data, API section, historic of the previous versions of the dataset.	User involvement: example of reuse already made, discussion feed on the dataset, multilingual option.

Table 2: Open data portal maturity measurement

The indicators have been chosen to reflect in the best possible way the maturity of the open data portals, meaning their degree of development. As reminded by the Guide to the Open Data Maturity Model, "a 'maturity model' generally provides a framework that allows an organisation to assess how well its processes conform to industry best practices" (2015, p. 8).

Regarding the provision of the data, it is the starting point of every portal. It is unlikely that a poor provider will release data in a very efficient manner. Thus, a lack of contribution or supply of data on the portal, and especially of OGD, can testify for a low awareness in the region and consequently for a low maturity of the open data portal. Besides, open data portals stand as unifying platforms for publishing open data in one place, the administrators and major providers are public bodies (Publications Office of the European Union, 2023). Consequently, the presence of non-official sources datasets shows an important degree of development, alike the provision of real-time datasets as it is data that changes and needs to be updated regularly. As highlighted by the Open Data Maturity Report, "such data is essential for several reuse cases, such as applications for optimising navigation in congested traffic or economic modelling" (2023, p. 102).

Regarding metadata, it is defined as data about the data. The metadata complement the datasets and enable them to be found and searched (Publications Office of the European Union, 2023). Janssen & al. explained that "data cannot be easily found if essential meta-data, like the publisher, authors, timeliness, and so on, is not available" (2012, p. 265). Many organizations have emphasized the importance of metadata in open data supply, among others the Open Data

Barometer and OPQUAST Good practices have drawn a particular attention to the implementation and accessibility of data through metadata. Similarly, the Open Data Portal Assessment Using User-Oriented Metrics report presents the publication of metadata as a major dimension to assess portals, and especially emphasizes the need for user oriented elements such as reviews to measure the usability of portals (Publications Office of the European Union, 2020a). To the same extent, Open Data Maturity Report pointed out that "analytics tools are used to gain insights into users' behaviour and the most and least consulted data categories" (2023, p. 90). Thus, at the minimum data needs to be "published with enough documentation and supporting metadata that re-users can understand the structure, scope and provenance of a dataset" (Dodds & Newman, 2015, p. 23). Analytics and user-oriented components are nevertheless necessary to stimulate a more efficient reuse of open data by targeting user behaviour on the portal and facilitating the use of the datasets.

Features of open data portals represent all the functionalities available to search and use a dataset. The more advanced those features are, the better the user experience. The Open Data Portal Assessment Using User-Oriented Metrics report pointed out "the importance for portals to extend their features with advanced search capabilities and, visualisation and analytics tools, as most portals only allow users to just download the available data" (2020a, p. 20). The goal of features is also to lower the barriers that may hinder the use of open data, for example visualization tools can help non-experienced users (Janssen et al., 2012), as well as API that helps to process easily the data (Publications Office of the European Union, 2023). Hence, the way data is implemented is essential to stimulate the way it can be reuse (OECD Public Governance Policy Papers, 2023; Opquast, 2011). Alike metadata, the more features are user-oriented the better the user experience on the portal. To that extent, Dodds & Newman proposed different process to engage with the users like building a re-user support process, for example through a discussion forum (2015).

Thus, the maturity of open data portals relies on the different indicators all assessing for the degree of development of the platform, from a basic platform to a user-oriented one. As in Colpaert & al. framework, the more collaborative the portal is, the easier its use is going to be hence offering more opportunities for open data reuse.

3.2.3. Regional governance effort to promote open data – *Condition 2*

This condition aims at measuring the extent to which each region is involved to foster the awareness about open data and guide its use. The implementation of a specific open data governance is relevant to facilitate open data use, especially through policies, processes and tools to structure the interactions between public institutions (as providers) and, private sector and citizens (Welle Donker & van Loenen, 2017). A good governance could not only promote open data use but also ensure its sustainability over time (Publications Office of the European Union, 2023). France, on the country-level, has a national digital strategy and some standards regarding the publication of open data, nevertheless most regions also have their own digital strategy to support and complement the national one.

This condition will be measure through 4 indicators *(Table 3)*.

Definition		Measurement			
Policy	Presence of specific policies oriented towards enhancing open data use.	No (0 pt)	Presence of a digital strategy mentioning open data. (1 pt)	Presence of a digital strategy with a marked focus on open data. (2 pts)	-
Visibility	Presence of the regional open data portal on social media.	No (0 pt)	Yes (1 pt)	-	-
Engagement	Share of datasets provided by the region on the regional open data portal	0% (0 pt)	From 1 to 34% (1 pt)	From 35 to 70% (2 pts)	From 71 to 100% (3 pts)
Guidance	Training or guidance program to support users.	No (0 pt)	Presence of digital support for the population (1 pt)	Presence of digital support with a focus on open data, involvement of the collectivity. (2 pts)	-

Table 3: Regional governance effort to promote open data measurement

In many open data framework assessment, policies and strategies have been at forefront because they represent good indicators of the governmental mechanisms in place to support the publication and use of open data. Open Data Maturity Report of EU highlighted that "in addition to national strategies and policies on open data, 20 Member States (74 %) stated that local and regional public entities also had an open data policy or strategy" (2023, p. 30). Among others, OECD and the Open Data Barometer looked at the presence of government policies and actions to release open data. Also, Dodds & Newman pointed out that "a mature open data organisation will have an open data strategy that clearly describes its ongoing commitments and policies relating to its open data practice" (2015, p. 30). Thus, not only the presence of a specific regional open data governance defines the legal context and standards to facilitate data use, but it provides also a stable and sustainable environment for the users (Welle Donker & van Loenen, 2017).

Ensuring a good visibility on social media of the regional open data portal enables each region to promote the activities of both publishers and the re-users community (Dodds & Newman, 2015). It is a social distribution channel for open data to connect directly with users and impact the portal use (Publications Office of the European Union, 2020a). It is a complement to a good open data governance by supporting data reuse through data promotion initiatives (OECD Public Governance Policy Papers, 2023).

The implication of regional bodies in open data publication is also important as they "serve as channels to create awareness about open data and encourage portal users to reuse open data"

(Publications Office of the European Union, 2023, p. 90). As a major provider, their involvement plays a part in fully showing the benefits of open data (OECD Public Governance Policy Papers, 2023).

The lack of public awareness about open data and the lack of skills to leverage it, are often highlighted as obstacles in open data use (Janssen et al., 2012; Publications Office of the European Union, 2023). Therefore, there is a need to support users through data awareness programmes or trainings, among others (OECD Public Governance Policy Papers, 2023). As emphasized by Dodds & Newman, "for data to retain its value in the long term, people who use it need to know about how it's collected, how it's processed and who owns it" (2015, p. 22). Regional support and guidance are thus an important component for making open data successful.

Through those 4 indicators, the condition 'regional governance effort to promote open data' demonstrates the need for institutional involvement in open data publication and use.

3.2.4. Share of high-skilled users – *Condition 3*

The last condition stands for the social dimension of the conceptual framework presented in Chapter 2. As already mentioned, the lack of skills among users is a barrier to open data use and its benefits. As stated by Welle Donker & van Loenen "having data supply and governance ticked off does not automatically mean that data will be re-used" (2017, p. 290).

The regional share of high-skilled workers is going to be operationalized through the Eurostat dataset 'Human resources in science and technology' (HRST). This dataset captures the share of the total regional population that has either successfully completed education at the third level and/or is employed in a science and technology occupation where a similar qualification is required (Eurostat, 2024). In other words, it presents the high- skilled part of the population.

3.3. Regions ranking & Calibration

3.3.1. Regions' assessment

This section details how the region ranking was conducted for the conditions 'Portal maturity' and 'Regional governance' based on the operationalization tables presented earlier. Two regions are used as examples: Ile-de-France and Nouvelle Aquitaine. The full assessment is presented in appendix.

- **Ile-de-France**

Open data portal (10 points)

Provision: The portal of IDF presents 900 datasets from 97 different providers. Those datasets are from public institutions, mostly municipal, departmental and regional bodies. No real-time dataset was found. Based on *Table 2*, the provision indicator was given the grade of 2.

Metadata: The datasets of the portal present each a specific information section containing the description, type and theme of the data, the number of downloads and popularity of the dataset, the last update, the geographical coverage, at least a generic contact for the provider and sometimes a direct one, and suggestions of similar datasets. Based on *Table 2*, the metadata indicator was given the grade of 4.

Features: The features found on the portal were also quite developed with a visualization and API section, a section to present the re-use made from the dataset, multiple formats to export

the data, a comment section for users. Based on *Table 2*, the feature indicator was given the grade of 4. The portal appears to be very user-oriented.

Regional governance effort to promote open data (5 points)

Policy: IDF has quite high regional ambitions in open data as it aims at becoming a 'smart region', leader in data and connectivity. Thus, a specific digital strategy has been set up to ensure proper services and meet citizens' needs in digital activities. Reinforcing and fostering open data is part of this strategy. Based on *Table 3*, the policy indicator was given the grade of 2.

Visibility: The regional open data portal is not present on social media by itself, only via the regional account. Based on *Table 3*, the visibility indicator was given the grade of 0.

Engagement: Out of 932 datasets, 395 are directly provided by the region itself, being 42% of the total datasets provided on the open data platform. Based on *Table 3*, the engagement indicator was given the grade of 2.

Guidance: The region is engaged in providing support and guidance to the general public. A budget has been allocated to foster digital accessibility and inclusion for beginners or for marginalized populations. Nevertheless, those programs are oriented towards very basic use of digital tools, and sensitizing people to the benefits of open data is not part of the regional action. Based on *Table 3*, the guidance indicator was given the grade of 1.

- **Nouvelle Aquitaine**

Open data portal (7 points)

Provision: The portal presents nearly 9000 datasets from a lot of different providers, including from private organizations. No real-time dataset was found on the portal. Based on *Table 2*, the provision indicator was given the grade of 3.

Metadata: The datasets of the portal present usually a lot of basic information such as a description and type of the data, the last update, the geographical coverage, and a generic contact for the provider. No analytic was found about the datasets (popularity, number of downloads, etc.). Based on *Table 2*, the metadata indicator was given the grade of 2.

Features: The features found on the portal were, as for metadata, quite basic, some research filters and different downloading formats for example. There is no visualization or API tool, no re-use section and no sign of any user-oriented components. Based on *Table 2*, the feature indicator was given the grade of 2.

Regional governance effort to promote open data (2 points)

Policy: No digital strategy is publicly provided by the region even if the regional website mentions some digital development axis. The few actions presented are mostly oriented towards supporting enterprises and innovation, more than data accessibility. Based on *Table 3*, the policy indicator was given the grade of 0.

Visibility: The regional open data portal is not present on social media by itself, only via the regional account. Based on *Table 3*, the visibility indicator was given the grade of 0.

Engagement: Out of 8525 datasets, 5 are directly provided by the region itself, being 0.06% of the total datasets provided on the open data platform. Based on *Table 3*, the engagement indicator was given the grade of 1.

Guidance: The region provides training courses and facilities to support digital inclusivity. Based on *Table 3*, the guidance indicator was given the grade of 1.

The assessment presented for IDF and Nouvelle Aquitaine was conducted for all 17 regions. The data for HRST and GDP was collected from public datasets. The data collection table is presented in conclusion of this section (*Table 8*).

3.3.2. Calibration of the raw data

As already mentioned in Section 3.3. calibration is to be conducted carefully to capture properly the data. Cases as configuration of conditions producing an outcome need to be studied as whole, meaning without decontextualizing, by preserving their complexity and the resulting diversity (Gerrits & Verweij, 2018b).

Following the guidelines provided by Gerrits & Verweij, we will first define the 'anchor points' (fully-in and fully-out of the set-membership), then we will determine what lies in-between for each case to be assigned a proper quantitative value. In fact, the whole purpose of calibration is to assess cases based on "the extent to which they feature a condition" (Gerrits & Verweij, 2018b, p. 56). The cross-over point will be defined at most based on external knowledge about the cases. In fact, as reminded by Schneider & Wagemann, calibration results from the combination of theoretical knowledge and empirical evidence, and it is the responsibility of the researcher to determine set-membership values of each case in a transparent way (2012).

For the calibration, we will be using only fuzzy-sets which allow for a more fine-grained calibration (Ide & Mello, 2022) and "may do more justice to the complexity of the empirical materials" (Gerrits & Verweij, 2018b, p. 67).

Resulting from the calibration process, each case will be assigned a set-membership score for every condition and for the outcome. These scores represent the "numerical expression for the belonging of a case to a set" (Gerrits & Verweij, 2018b, p. 57). Also, it is to keep in mind that set-membership values are "intrinsic to the research in which they are used" and are "closely linked to the research context" (Schneider & Wagemann, 2012, p. 34).

3.3.2.1. Calibration of Economic growth (GDP) – *Outcome*

The economic weight of French regions can vary significantly from a region to another (cf. *Table 8*). It is often highlighted that French GDP relies mostly on two main regions: Ile-de-France (IDF) and Auvergne-Rhône-Alpes (ARA) ('Poids Économique Des Régions', 2024). They are the most dynamic in France regarding the provision of goods and services especially in the Information & Communication sector ('La France et Ses Territoires', 2021). In 2022, they represent together 41% of the national GDP (29.7% for IDF and 11.6% for ARA).

On the contrary, Corse and Centre-Val de Loire are often considered the weakest ones in metropolitan France. Considering France globally, overseas regions are far behind gathering only 2% of the national GDP in 2022 for four regions together ('La France et Ses Territoires', 2021), the most important one being La Réunion with 0.8% of French GDP.

As part of this research, we consider "GDP" as being the set of French regions with a high GDP. Consequently, full membership is defined as at least 10% of the national GDP and full non-membership is defined as less than 1%.

Following IDF and ARA, five regions are also quite important, all of them weighting around 7% of the total French GDP ('Produit Intérieur Brut En 2022', 2024). For most of them, those

regions rely on an important industrial sector allowing them to perform economically at the country level. This gap between simultaneously the economically important region and the one lagging behind can also be observed empirically in the dataset.

Consequently, the cross-over point will be defined at 6% of the national GDP, qualitatively differentiating the regions with a high GDP set-membership from the regions with a low GDP set-membership.

Set-membership score	Percentage of the national GDP	Value in million €
1 – Fully in	10%	263 748
0.5 – Cross-over point	6%	158 249
0 – Fully out	1%	26 375

Table 4: Calibration of the outcome

3.3.2.2. Calibration of Portal maturity – *Condition 1*

Following the operationalization presented in section 3.2.2 and the qualitative assessment presented earlier in this Chapter, "Portal maturity" is the set of French regions with a mature open data portal. The full membership is defined by a level of maturity more than advanced. In other words, if the regional grade is strictly superior to 9 points, meaning that at least one indicator belongs to the Expert category.

On the contrary, the full non-membership is defined by a maturity level basic or less, meaning for a regional grade inferior to 3 points.

It has been shown that most regions obtained relatively high scores, most of them are above the average possible grade of 6 points. Hence, the cross-over point is defined by a maturity level superior to 7.5, ensuring that at least two indicators belong to the advanced category. The cross-over point separates the regions with a high portal maturity set-membership from the regions with a low portal maturity set-membership.

Set-membership score	Portal maturity grade
1 – Fully in	9
0.5 – Cross-over point	7.5
0 – Fully out	3

Table 5: Calibration for the condition 1

3.3.2.3. Calibration of Regional governance – *Condition 2*

Following the operationalization presented in section 3.2.3 and the qualitative assessment presented earlier in this Chapter, "Regional governance" is the set of French regions with a high regional governance involvement to promote open data. The full membership is set for the grades of regional governance effort superior or equal to 5 (materialized by a set-membership score of 4.7), ensuring that at least two indicators have the maximum possible grade.

On the contrary, the full exclusion is defined by no regional governance effort to promote open data, so a regional grade of 0.

On average, there are more regions under the average possible grade than above, meaning that based on the criteria defined French regions are, for most of them, not so good to promote open data. The average regional grade is 3.4 in the established ranking, and it appears to also represent a qualitative difference in the data as a noticeable threshold between the most performant regions and the others. This cross-over point thus differentiates the cases with a high regional governance involvement from the one with a low regional governance involvement.

Set-membership score	Regional governance effort grade
1 – Fully in	4.7
0.5 – Cross-over point	3.4
0 – Fully out	0

Table 6: Calibration for the condition 2

3.3.2.4. Calibration of HRST – *Condition 3*

The share of human resources in science and technology presents some major gaps between regions (cf. *Table 8*). Especially, IDF is by far the most skilled regions as it is the only one with more than half of its population either with tertiary education and/or employed in science and technology. In fact, almost half of all digital jobs in France are located in IDF, among them 6 out of 10 workers are with high-education compared to 3 out of 10 on average in other regions (Abboudi, Ferré, Pichard, & Lefranc, 2019). These observations reflect the weight of IDF region compared to the rest of France and thus allow us to define the fully-in set-membership to at least more than 50%, with "HRST" being the set of French regions with a high share of skilled population.

Reversely, some regions, in particular overseas ones, have a poor digitalization and have been struggling for years for digital inclusivity (Vall, 2020). Logically, those regions have a share of HRST inferior to 30%, which represent the full exclusion of this set-membership.

Nevertheless, in the rest of French regions digital professions require increasingly highly qualified people. In 2015, 52.7% of people working in digital professions were highly educated (Bailieul & Gamblin, 2019). In this digital race, a group of regions has been performing better. *Figure 4* shows that among others Occitanie and ARA for example especially increased a lot the share of their IT skilled workers in the past years. This group also happened to perform quite well in our 2022 datasets (cf. *Table 8*). Thus, the cross-over point is defined at 39.1% representing the threshold between trendsetter regions with a high share of skilled population (excluding IDF) and followers with a lower share of skilled population.

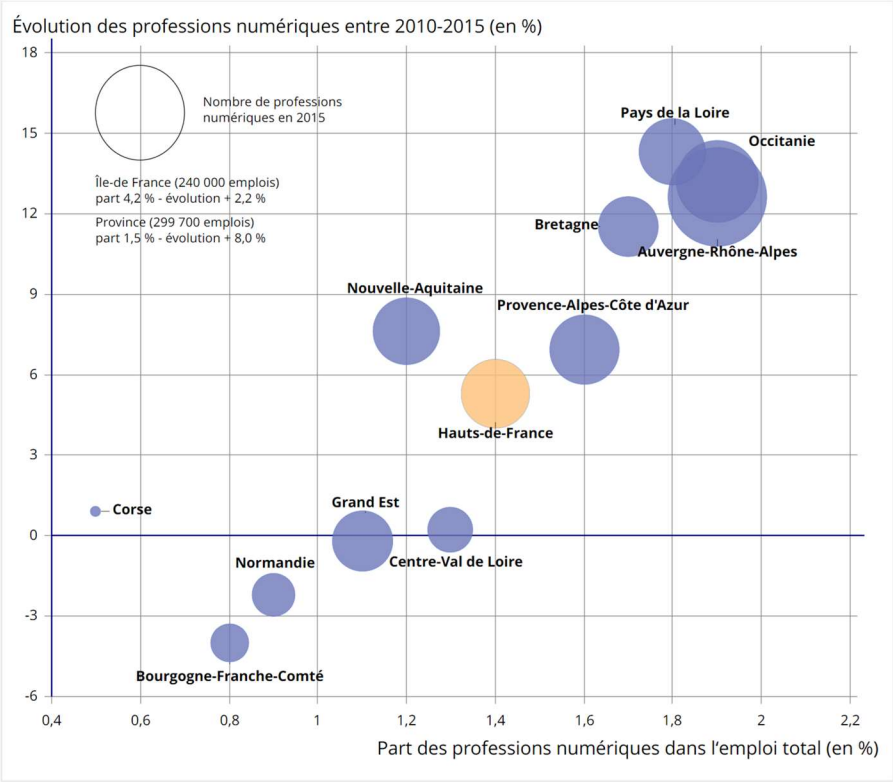


Figure 4: Number and share of digital professions in 2015 by region and 2010-2015 trends, Insee.fr

Set-membership score	HRST share
1 – Fully in	50%
0.5 – Cross-over point	39.1%
0 – Fully out	30%

Table 7: Calibration for the condition 3

3.3.2.5. Data matrix

The data matrix resulting from the regional assessment and the calibration process is presented in *Table 8*.

Id	Cases	Portal maturity		Regional governance		HRST (in % of total population)		GDP (in million €)	
		Raw	Calib.	Raw	Calib.	Raw	Calib.	Raw	Calib.
1	Ile de France	10	0.99	5	0.97	53	0.98	782 639	1
2	Centre – Val de Loire	10	0.99	4	0.80	35.8	0.26	79 992	0.15
3	Bourgogne – Franche-Comté	9	0.95	4	0.80	36	0.27	87 549	0.17

4	Normandie	6	0.27	3	0.41	35.3	0.23	107 192	0.24
5	Hauts de France	6	0.27	4	0.80	34	0.16	185 472	0.68
6	Grand Est	9	0.95	4	0.80	33.5	0.14	178 925	0.64
7	Pays de la Loire	11	1	3	0.41	37.3	0.36	137 051	0.38
8	Bretagne	9	0.95	6	1	39	0.49	115 184	0.28
9	Nouvelle Aquitaine	7	0.42	2	0.23	39.2	0.51	199 575	0.76
10	Occitanie	10	0.99	4	0.80	41.1	0.63	193 064	0.73
11	Auvergne-Rhône-Alpes	7	0.42	3	0.41	43.4	0.76	304 681	0.98
12	Provence-Alpes-Côte d'Azur	8	0.73	3	0.41	40.7	0.61	206 744	0.79
13	Corse	8	0.73	5	0.97	33.2	0.13	10 304	0.04
14	La Réunion	10	0.99	2	0.23	25.1	0.01	21 668	0.05
15	Guyane	0	0.01	3	0.41	21.5	0	4 562	0.03
16	Martinique	0	0.01	0	0.05	29.9	0.05	9 654	0.03
17	Guadeloupe	8	0.73	4	0.80	24.7	0.01	9 877	0.04

Table 8: Data matrix after calibration

3.4. Expected challenges and limitations

Even if very powerful, QCA methodology can also present some challenges. Calibration, as a major step of the process, is to be conducted carefully because it can impact strongly the QCA results. Calibration "creates standards against which data measurements become interpretable" (Gerrits & Verweij, 2018b, p. 52), and multiple 'rules' exist to guide the researcher in this process. Especially, calibration needs to be explained as mechanistic calibration is not a solution to properly reflect the data with the set-membership values (Rubinson, Gerrits, Rutten, & Greckhamer, 2019).

Among the most common interpretation challenges is limited diversity, when the number of combinations presented is inferior to the number of combinations logically possible. Limited diversity thus restrains the analysis because less results are available for the researcher. In fact, "when there is not enough empirical evidence to test certain theoretical expectations, then the algorithm may produce false-positive findings, leading to unreliable results" (Ide & Mello, 2022, p. 11). Gerrits & Verweij interpreted this empirical absence of configurations as "indicative for the nestedness of cases, that cases have unique properties but also exhibit similarities which makes certain configurations less likely" (2016, p. 16).

Logical contradiction can also impact the interpretation, one configuration may produce different outcome in cases. Despite, QCA's complexity-informed nature, some cases' boundaries may be too narrow to produce an outcome with just the conditions in place (Gerrits & Verweij, 2016). Resolving contradictory configurations needs to be addressed "based on the nature and strength of the evidence and the clarity of the guidance offered by relevant theoretical perspectives" (Ragin, 1999, p. 1236).

Regarding the limitations, the major one regarding QCA analysis is that it does not show results over time. As highlighted by Rihoux & Grimm, QCA fits only temporally static topic (2006). So, the temporal dimension is a limiting factor for the research question. In the context of this thesis, the datasets used are focusing on the year 2022.

4. Results, analysis and discussion

In this chapter, we are going to detail the QCA analysis conducted on R software to identify patterns and causal recipes across cases. As already mentioned, the specificity of QCA methodology allows us to account for the complexity of cases and to look at the conditions on their own and in the context of the other conditions (Gerrits & Verweij, 2018c). Especially, QCA set-theoretic analysis tests not only individual conditions but also combinations of conditions for their necessity and/or sufficiency (Mello, 2021).

The analysis is conducted along 3 main steps. First, we focus on the analysis of the truth table obtained from the data matrix presented in the previous chapter, then this truth table is minimized to obtain the solution formula of the research, eventually this formula is interpreted to answer the research question. The analysis is conducted for the occurrence of the outcome. The non-occurrence is presented in appendix.

4.1. Results and analysis

4.1.1. Necessary conditions

The analysis of potential necessary conditions, as a first step in the analytical process of QCA even before engaging with the truth table, is often considered as a good practice. Knowing about necessity could influence the researcher's decisions for the rest of the procedure, besides the observation of necessity is not straightforward from the analysis of sufficient conditions (Mello, 2021). Also, as highlighted by Ragin, "necessary conditions have very powerful policy implications" (1999, p. 1228).

Necessity is defined as a superset relation, whenever the outcome Y is present the necessary condition X is also present. In other words, a condition is considered necessary if its presence is required for the outcome to occur. Thus, the analysis of necessity is conducted to test "whether a specified number of conditions–in their absence or presence–are individually necessary for the outcome" (Mello, 2022, p. 21).

The consistency theoretical threshold for necessity is conventionally set at 0.9. The consistency evaluates the reliability of a condition to lead to the outcome. It allows "one to assess to which degree the cases that share a given condition X agree in displaying the outcome Y" (Gerrits & Verweij, 2018c, p. 111).

	Cons.Nec	Cov.Nec	RoN
MATURITY	0.797	0.489	0.490
GOVERNANCE	0.784	0.532	0.582
HRST	0.725	0.905	0.956
~MATURITY	0.383	0.479	0.796
~GOVERNANCE	0.528	0.551	0.774
~HRST	0.607	0.372	0.439

Table 9: Necessity table for the occurrence of the outcome

As presented in *Table 9*, as part of this research none condition passes the consistency threshold, meaning that no condition on its own is enough to produce the outcome. As an example, the condition MATURITY has the higher consistency with 0.797. By plotting the necessity relation

between this condition and the outcome GDP (*Figure 5*), we notice that in fact even if most cases conform to the superset relation of X ≥ Y, too many are way above the 'threshold line'. The others measures of fit, coverage and relevance, are not meaningful for the research given the low consistency scores (Mello, 2021).

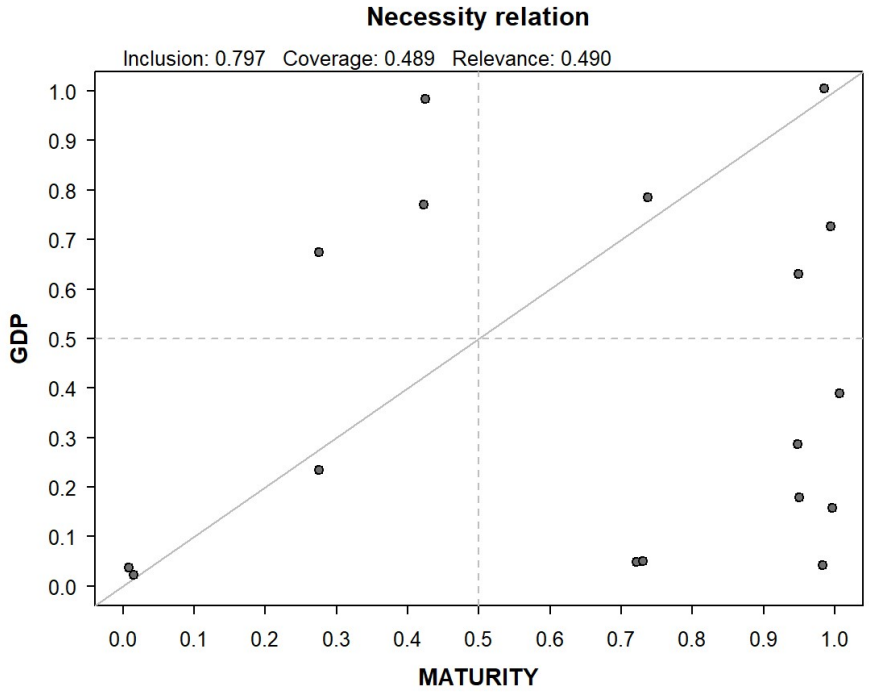


Figure 5: Necessity relation for the condition MATURITY

This first analytical step shows that neither the presence nor the absence of a mature open data portal, an important regional governance effort or a high share of high-skilled population is necessary on its own to contribute to an important economic growth.

4.1.2. Truth table

Following the analysis of necessity, the truth table is a major tool in the comparative process of QCA and results in a solution formula identifying the patterns leading to the outcome. The table displays all the logically possible configurations for the conditions to produce the outcome. In our case, 8 rows are listed as 2^k combinations are possible, with k = 3 representing the number of conditions. In the process, each case is assigned to a row, and each row is assigned an outcome either 1 or 0, expressing whether the combination of conditions produces the outcome or not based on the empirical data (Gerrits & Verweij, 2018c). The truth table provides different information about the empirical distribution of cases, how they connect to the outcome and thus analytically explores the data (Mello, 2021).

The truth table obtained is presented in *Table 10*.

Maturity (MAT)	Governance (GOV)	HRST	OUT	n	incl	PRI	Cases
0	0	1	1	2	0.990	0.978	Nouvelle-Aquitaine, Auvergne-Rhône-Alpes
1	0	1	1	1	0.972	0.925	Provence-Alpes-Côte d'Azur
1	1	1	1	2	0.889	0.790	Ile de France, Occitanie
1	0	0	0	2	0.680	0.333	Pays de la Loire, La Réunion
0	1	0	0	1	0.642	0.376	Hauts-de-France
1	1	0	0	5	0.491	0.148	Centre - Val de Loire, Bourgogne-Franche-Comté, Bretagne, Corse, Guadeloupe
0	0	0	0	3	0.455	0.213	Normandie, Guyane, Martinique
0	1	1	?	0	-	-	-

Table 10: Truth table for the occurrence of the outcome

Before presenting the results, it is important to highlight that, as discussed in section 3.4, some challenges regarding the reliability and quality of the truth table must be taken into consideration in the analysis. As part of the research, the first challenge faced is related to contradictory configurations, meaning that similar configurations can present different outcome. As pointed out by different authors, contradictions don't have to be seen as problems but more as opportunities for the researcher to get engage deeper with the cases through the resolution process (Gerrits & Verweij, 2018c; Rihoux & Ragin, 2009a). As stated by Ragin, QCA is an iterative process that requires a dialogue between ideas and evidence, in other words between the theory and the data. It is even more true as this research have a relatively small number of cases and it is thus conceivable to gain familiarity with each case.

Different strategies exist to deal with contradictory configurations, in our case, we had especially to reexamine in a more qualitative way some cases involved in the contradictions in order to identify what had been missed and could differentiate those cases (Rihoux & Ragin, 2009a). As an example, regions Normandy and Haut-de-France presented very similar pattern in the three conditions but a different outcome. Thus, the assessment made for the open data portal maturity and the governance effort was re-conducted in order to understand what could possibly differentiate the 2 regions and lead to such different outcomes. By going back to the cases, some distinctions could be made and the ranking was reviewed, hence resolving the contradiction. This strategy is very case-oriented, but some other ones also require going back to the theory, to review a calibration for example.

Nevertheless, sometimes strategies are not enough for the resolution of contradictory configurations and some cases must be removed. Taking the example of the configuration 110 leading to outcome 0, the six cases below (*Table 11)* were initially assigned to this row. We can see in *Table 11*, that all the cases present very similar patterns, but Grand Est is the only case producing the outcome. The fact that Grand Est set-memberships were so close to the ones of Centre – Val de Loire, and Bourgogne-Franche-Comté for example, excluded the possibility to adjust calibration thresholds as a resolution strategy. Going back to the cases and re-explore the data collected did not prove conclusive.

Cases	Maturity (MAT)	Governance (GOV)	HRST	GDP
Centre — Val de Loire	0.99	0.80	0.26	0.15
Bourgogne-Franche-Comté	0.95	0.80	0.27	0.17
Grand Est	0.95	0.80	0.14	0.64
Bretagne	0.95	1.00	0.49	0.28
Corse	0.73	0.97	0.13	0.04
Guadeloupe	0.73	0.80	0.01	0.04

Table 11: Cases initially assigned to the configuration 110

Out of the 17 French regions studied (and 18 in total), Grand Est is the only one producing the outcome with the configuration MAT*GOV*~HRST. By checking the membership of this case to all the possible combinations of conditions, it appears that it has a 0.80 membership to this configuration. Based on those observations, Grand Est stands as an outlier in the open data French regional scope. Nevertheless, one can wonder why such a notable difference with the GDP values of the 5 other regions presented in *Table 11*, despite the similarities exposed by the operationalization of the different QCA conditions. Going back to French economic patterns on a wider scale, it appears that a strong link exists between the economic situation of a region and its number of inhabitants. In France, a high GDP is most of the time correlated with a highly populated region, even if it is obviously not the only factor for economic success. The seven richest regions are also the most populated (Insee). In 2022, Grand Est had 1,6 times more inhabitants than Bretagne, 2,2 times more than Centre-Val de Loire and nearly 15 times more than Guadeloupe. So, the decision was made to remove Grand Est from the truth table to conduct the minimization process without contradiction. We can assume that it is the only region producing the outcome with the configuration 110 not because of its portal maturity or its governance in open data, but because its economic relative success is partly rooted in its regional characteristics.

The second challenge concerns limited diversity, i.e. some configurations are not observed in the truth table due to the lack of empirical evidence. Those empty rows are logical remainders. In our case, the truth table presents one logical remainder for the combination of conditions ~MAT*GOV*HRST (011). It appears that no case of the research sample shows this pattern, leading or not to the outcome. Similarly to contradictory configurations, different strategies exist to deal with limited diversity, including adding more cases to the research and reducing the number of conditions (Gerrits & Verweij, 2018c). In our case, they are not conceivable given that all possible French regions are already included in the research, and three conditions is a minimum in QCA methodology. Logical remainders can be factors for inconsistencies between the truth table and the Boolean minimization, so deciding whether they should be included for the minimization procedure or not is to be considered by the researcher, mostly based on theory (Cooper & Glaesser, 2016). This point will be further developed in the next section.

Going back to the truth table, we can see that 3 combinations of conditions are sufficient to produce the outcome:

~MAT*~GOV*HRST → GDP

MAT*~GOV*HRST → GDP

MAT*GOV*HRST → GDP

Those configurations all passed the consistency threshold defined at 0.75 indicating "which rows of the truth table should be treated as positive instances of the outcome and included in the analysis" (Mello, 2021, p. 12). The three configurations present a consistency of at least 0.889, which appears to be quite a high consistency, 1 being the perfect score. In fact, "the higher the consistency score, the more evidence is in line with the statement" (Gerrits & Verweij, 2018c, p. 101). As pointed by Ragin & Rihoux, "social science data are rarely perfect, so it is important to assess the degree to which the empirical evidence is consistent with the set theoretic relation in question" (2009b, p. 22). The rest of the configurations have too low consistency to be considered a subset of the outcome and be included in the analysis.

The PRI (Proportional Reduction in Inconsistency) score is provided as an additional consistency indication. A low PRI score signals that a configuration could be sufficient for both the outcome and the non-outcome. By considering the PRI score, the researcher prevents for asymmetric causality, meaning including a truth table row in both the outcome and the non-outcome minimization (Gerrits & Verweij, 2018c). In our case, the PRI are also quite high, especially for the first two combinations.

So, the truth table indicates that out of 8 possible configurations, 3 appears to be sufficient to produce the outcome. It ensues the following solution term:

~MAT*~GOV*HRST + MAT*~GOV*HRST + MAT*GOV*HRST → GDP

Sufficiency is defined as a subset relation, whenever the sufficient condition X is present the outcome Y is also present. In other words, a condition is sufficient if its presence guarantees the occurrence of the outcome.

This expression can be simplified by the minimization process to facilitate the understanding of the results.

4.1.3. Minimized solution formula

The truth table draws a first overview of the generic patterns across cases producing the outcome. Nevertheless, to closely analyse necessity and sufficiency of conditions or combinations of conditions it is better to work with the minimized solution formula. The minimization consists in a bottom-up pairwise comparison to simplify the truth table solution formula by combining rows that differ in only one causal conditions, but yet produce the same outcome (Ragin, 1999).

While the consistency threshold, presented in the previous section, specifies the rows producing the outcome that should be included in the minimization, the second step is to decide whether the logical remainders should be also included in the process. Three different approaches exist: the conservative solution excluding all the logical remainders, the parsimonious solution including them all, and the intermediate solution that sits in-between the two previous ones. As part of this research, we chose to run the model with both the conservative and the parsimonious solution. The results are presented in *Table 12*.

Path	incls	PRI	covS	covU	Cases
Conservative solution					
MAT*HRST	0.898	0.820	0.706	0.266	Provence-Alpes-Côte d'Azur, Ile de France, Occitanie

~GOV*HRST	0.968	0.926	0.483	0.044	Nouvelle-Aquitaine, Auvergne-Rhône-Alpes, Provence-Alpes-Côte d'Azur
M1 MAT*HRST + ~GOV*HRST	0.900	0.830	0.750		
Parsimonious solution					
HRST	0.903	0.838	0.776	-	Nouvelle-Aquitaine, Auvergne-Rhône-Alpes, Provence-Alpes-Côte d'Azur, Ile de France, Occitanie

Table 12: Minimization results

4.1.3.1. Conservative solution

The conservative solution works exclusively with the data at hand, thus refraining from making any assumptions about the logical remainders (Gerrits & Verweij, 2018c). This solution is often seen as the more complex one, as it usually yields longer solution terms because it includes less rows for the pairwise comparison (Mello, 2022). The minimization process for this solution is presented in *Figure 6*.

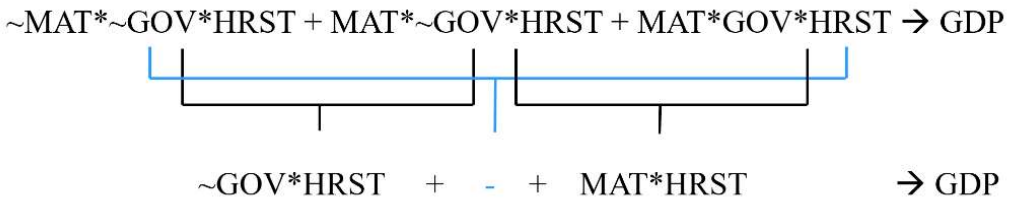


Figure 6: Minimization process for the conservation solution

The results presented in *Table 12* allow us to deduce that the configurations MAT*HRST and ~GOV*HRST are sufficient to produce the outcome. The solution formula can also be written in the factored form **HRST(MAT*~GOV)**, as HRST is shared by both pathways. Also, it appears that MAT, HRST and ~GOV are INUS-conditions.

INUS-condition refers to a condition that is 'Insufficient but Necessary part of a configuration, which is in itself Unnecessary but Sufficient to the outcome'. As an example, the condition MAT is a necessary element of the configuration MAT*HRST, but it is not sufficient as the condition HRST is also needed. At the same time, MAT is part of the sufficient configuration MAT*HRST, but this configuration is not necessary to produce the outcome as the configuration ~GOV*HRST also produces it.

This minimized formula highlights two major features of QCA: conjunctural causation and equifinality. In fact, none condition on its own is sufficient but rather two different combinations of conditions both producing the same outcome. Regarding multifinality, it must be studied by also looking at the non-occurrence of the outcome, it will not be covered in the thesis.

Both configurations present a high consistency and a high PRI (cf. *Table 12*) which, as already mentioned, comfort in their reliability to lead to the outcome. Regarding the raw coverage, it shows "how much of the outcome is explained by each of the paths in a solution formula separately" (Gerrits & Verweij, 2018c, p. 112). It appears to be quite high for the first

configuration and lower for the second one, but it is to be noted that unlike consistency there is no coverage threshold. Thus, the main point to be highlighted is that the outcome is explained at 70.6% by the configuration MAT*HRST. The unique coverage expresses the proportion of the outcome that is uniquely covered by the configuration, it considers the existing overlapping between the different paths. For example, the case PACA can be explained by both combinations of conditions, so the unique coverage of MAT*HRST is 26.6% because it only explains the part of the outcome represented by the cases IDF and Occitanie, as no other path covers those cases.

Regarding the 0.90 consistency of the whole model, it confirms that it is reliable for interpretation. Similarly, the solution coverage explains, how much of the outcome can be explained by the overall solution formula. It is quite high with 0.75 meaning that 75% of the outcome is covered by the entire solution term (Rihoux & Ragin, 2009a). Those results reinforce the reliability of the solution formula.

4.1.3.2. Parsimonious solution

The parsimonious solution considers all the logical remainders of the truth table in the minimization process. As explained by Mello, "since the parsimonious solution is allowed to examine all logical remainders, it can work with the broadest pool of configurations and thus its minimization procedure tends to yield the least complex solution terms" (Mello, 2021, p. 14). The minimization process for this solution is presented in *Figure 7*.

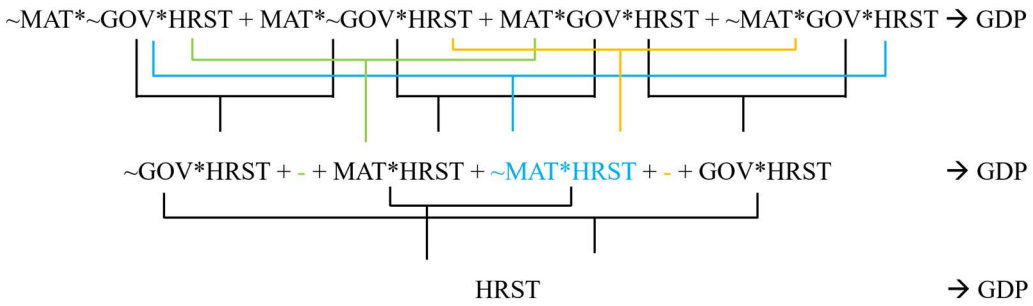


Figure 7: Minimization process for the parsimonious solution

The results presented in *Table 12* show that the condition **HRST** is sufficient to produce the outcome. The solution appears to be consistent with the set theoretic relation with a consistency score of 0.903. Similarly, the PRI score is also high, preventing from asymmetric causality as already mentioned. The coverage highlights that the outcome is explained at 77.6% by the configuration HRST. In total, five cases are covered by the parsimonious solution, those regions all present a high share of skilled population thus ensuring the presence of the outcome high GDP.

4.1.3.3. Counterfactual analysis

The counterfactual analysis aims at hypothesizing configurations that are not observed in the empirical data and determine whether they are more likely to lead or not to the outcome. In our case, we must decide whether the configuration ~MAT*GOV*HRST could produce the outcome, even if there is no empirical evidence in the research sample. This reasoning is to be made based on theoretical and substantive knowledge, and by going back to the cases (Mello, 2021). By evaluating the plausibility of the logical remainder, the counterfactual analysis will

determine which solution (conservative or parsimonious) could explain the successful cases in the most meaningful way.

In favour of the conservative solution, we must remind that the logical remainder (011) is not in line with the theoretical expectation of the research that the three conditions are needed to produce the outcome. So, based on the research framework the plausibility of this configuration producing the outcome is not so strong to include the row in the minimization.

Nevertheless, by looking at the truth table it appears that the sufficient rows tend to suggest that this configuration could also show the outcome. In fact, the condition HRST is shared by those three pathways. The conservative solution HRST(MAT*~GOV) also highlights the central role of HRST in producing the outcome. Additionally, from a theoretical perspective, if having a high share of skilled population on its own shows the outcome (first row of the truth table), then why having not only a high share of skilled population and an important governance effort wouldn't produce it also?

Based on this reasoning, we will assume that the configuration ~MAT*GOV*HRST produces the outcome, and we will consider the solution formula of the parsimonious solution as the minimized solution formula of the research: **HRST → GDP**
Also, by assuming that the logical remainder (011) produces the outcome, we place HRST not only as a sufficient condition but also as a necessary one, as whenever the outcome is present, this condition is also present.

4.2. Interpretation & Discussion

Based on the analysis presented in section 4.1, this part focuses on interpreting and discussing the results. Especially, the findings will be confronted to the research hypothesis. This step of the QCA methodology is not focusing anymore on processing the quantitative data, instead it is a return to the qualitative cases to understand the meaning of the results obtained (Verweij, 2015).

The interpretation relies mostly on the minimized solution formula as the most relevant result for the researcher "to ask more focused 'causal' questions about ingredients and mechanisms producing (or not) an outcome of interest, with an eye on both within-case narratives and cross-case patterns" (Rihoux & Ragin, 2009a, p. 34). The minimized solution formula highlights that a high share of skilled population is sufficient and necessary to contribute to a high regional GDP.

This pathway indicates that a region having an important share of high-skilled people among its population appear to produce a high GDP, regardless of the maturity of its open data platform and of its governance effort to promote open data on the regional scale.

Going back to the conceptual framework detailed in Chapter 2, it is interesting to compare it to the Venn diagram of the minimized solution formula (*Figure 8*).

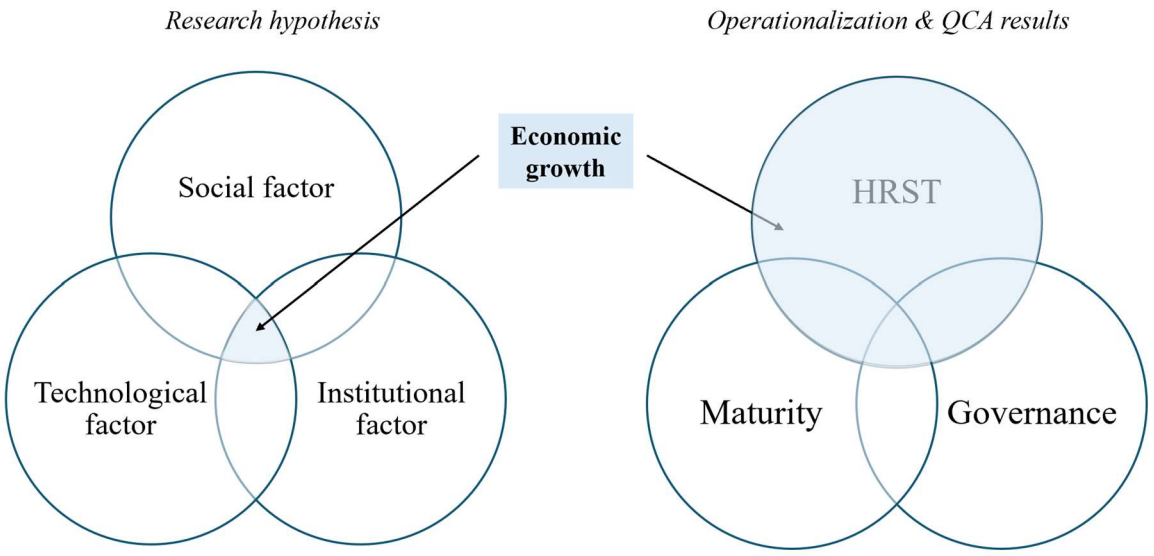


Figure 8: Comparison of the research hypothesis and the QCA results

The main hypothesis of this research was that the combining effect of a mature open data portal, an important institutional regional effort to promote open data, and a high share of high-skilled users contributes to the economic growth of a region. This assumption is verified as this pathway in fact produces the outcome, but it is not the only causal recipe, and more importantly its presence is not necessary for the outcome to occur.
A focus was also made on the role of the social factor by assuming that the presence of a high-skilled population was a necessary condition for a high GDP. The condition HRST is in fact necessary and sufficient to contribute to a high GDP, and consequently a central component to consider in the regional policy framework.

Among the successful cases showing a high regional GDP, Ile-de-France and Occitanie are the ones presenting the best open data potential. They exemplify by presenting not only a performant and user-oriented open data portal, but also an important institutional regional involvement embodied especially by strong digital policies, and lastly a regional population among the most educated and/or IT-skilled of the French scope. In the case of PACA region, it illustrates by showing a mature open data portal and a high-skilled population. The regional governance lacks in terms of proper digital policies supporting open data use, but the research shows that it is not an obstacle for successful regional economic growth. Lastly, Nouvelle-Aquitaine and ARA both presents a high share of skilled population and produce a high GDP despite the lack of maturity of their open data portal and the lack of proper governance mechanisms to promote open data on the regional scale.

By plotting the results (*Figure 9*), we can see that despite the sufficiency of the condition HRST to produce the outcome, the regions presenting the highest GDP are not fundamentally the ones having also a high portal maturity and high governance. Indeed, as we just mentioned, IDF is in fact a very complete region regarding open data but ARA, as the second economic French region, does not prove to have neither a performant open data platform nor a good regional governance.

To the same extent that the results point out that some disparities exist among metropolitan regions as some of them produce the outcome and some don't. They also especially emphasize that no overseas regions prove to be successful in presenting a high GDP. Among the four overseas cases, two of them don't show any condition (Martinique and Guyane), La Réunion

only presents good technological infrastructures with a mature open data portal, and Guadeloupe shows a mature open data portal and an important governance support but fails to produce the outcome as its population seems to lack of IT-skilled persons. This observation provides interesting insights to deal with regional economic discrepancies.

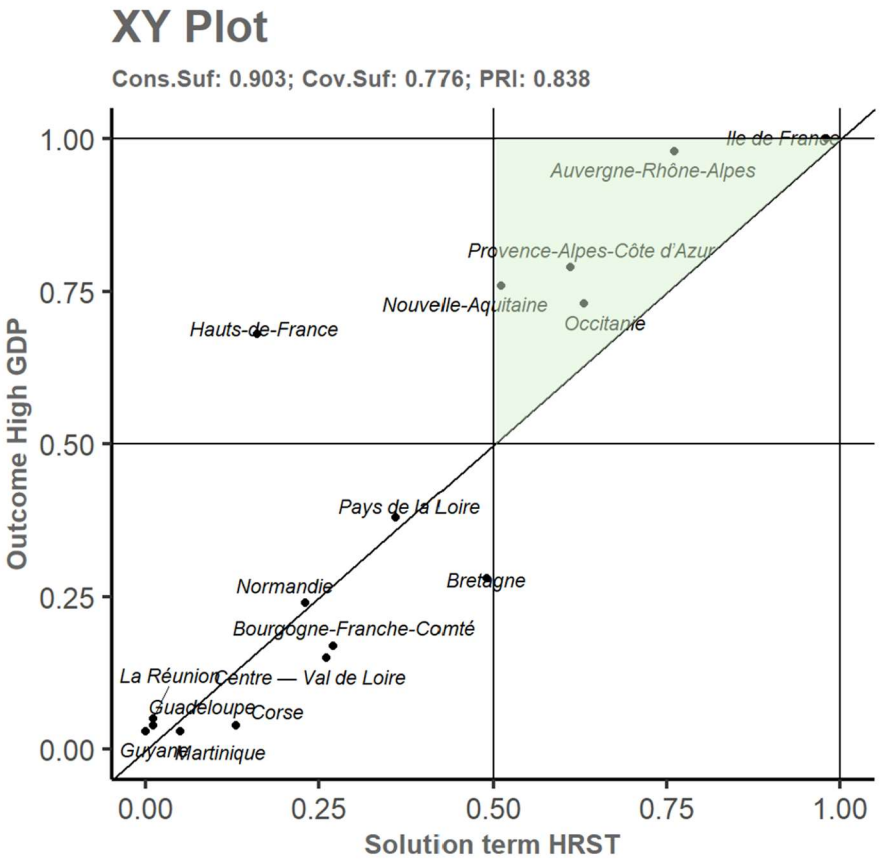


Figure 9: XY plot for the minimized solution formula

To conclude the results obtained and materialized in *Figure 8* highlight that there is not a single path producing the outcome but multiple mechanisms. Having a high share of skilled population is sufficient and necessary but the different regional contexts then articulate in different ways to produce a high regional GDP. The ensuing policy implications and recommendations will be further detailed in the next Chapter.

5. Conclusions

This study aims at addressing open data and its economic value on a regional scale, through a holistic approach considering at the same time the publication process, the policy context and the user dimension. Based on the argument that open data has value only if it is used, the hypothesis was made that to enhance its benefits the combining effect of a mature open data portal, an important institutional regional effort to promote open data, and a high share of skilled users contribute to the economic growth of a region. Also, a focus was made on the social dimension by assuming that within this regional framework the presence of high-skilled users was a necessary condition for high regional economic performances.

The research emphasized that contextual considerations are important, thus the findings are mostly valid within the scope of France. Generalization, to other European countries for example, needs to be conducted carefully by considering the plausible differences at the national and regional levels. This study gives an overview of French regional open data framework. It contributes to the scientific field by adding an assessment of open data economic potential on a localized scale. Above all, it reinforces the statement that open data benefits come from its use by highlighting the central role of users.

Through fsQCA, the analysis developed along the thesis provides some major findings:

Finding 1: Having a high share of skilled population is a necessary and sufficient condition to produce a high regional GDP.

Finding 2: A mature open data portal and an important regional governance effort to promote open data are neither necessary nor sufficient condition for a region to show a high GDP.

Finding 3: Some important regional discrepancies exist within France, especially regarding overseas regions.

As an answer to the research question, those findings point out that what contributes to regional economic growth in France is a highly skilled population, comprising highly educated people or/and people qualified in the science and technology field. Open data in itself does not contribute directly to economic growth, but rather through its users and the skills and knowledge they have to make a proper use of it, independently of the quality of the open data platform or of the support of regional governance.

A potential explanation of these findings could be that France is considered as a forerunner in open data since the country has been ranked every year since 2015 in the European ‘trendsetters’ by the European Commission Open Data Maturity Report, highlighting at the same time the quality of its national open data infrastructures and the supportive policy framework on the country level. Consequently, we can assume that the open data framework is sufficiently developed and structured on the national scale to cover regional needs. In other words, the national open data platform can replace the regional one, and the national governance is strong enough to support regional initiatives. As an example, assuming that a user or a company wants to make use of open data, the national open data framework could replace the regional one, contrarily to knowledge that remains highly localized.

Various practical implications and policy recommendations result from this analysis. First of all, in a society where digital is becoming omnipresent especially for economic purposes, this research emphasized that people are the central element, since as users, they are the key link to process the raw data into a valuable product. Thus, a focus on education, formation, training courses, and raising awareness is required from public institutions. Besides, this focus needs to be even stronger and consistent for the regions lagging behind in terms of digital inclusivity, especially overseas ones, through financial support for example.

By highlighting the importance of the social dimension in the process of value creation, suggestions for further research could be to explore users' needs more in-depth. Technological infrastructures and governance mechanisms proved to be rather shallow in the open data regional ecosystem, so what could support users' initiatives? Similar studies could also be conducted with other European countries to potentially highlight the impact of the national ecosystem compared to the regional one.

Bibliography

Abboudi, M.-C., Ferré, T., Pichard, L., & Lefranc, J. (2019). Près de la moitié des emplois du numérique localisés en Île-de-France. Retrieved from Insee website: https://www.insee.fr/fr/statistiques/4259562#graphique-figure1

Asheim, B. T., & Gertler, M. S. (2005). The geography of innovation: Regional innovation systems. In *The Oxford handbook of innovation*. Oxford ; New York: Oxford University Press.

Befani, B. (2013). Between complexity and generalization: Addressing evaluation challenges with QCA. *Evaluation*, *19*(3), 269–283. https://doi.org/10.1177/1474022213493839

Bonina, D. C. M. (2013). *The value of open data: Definitions, challenges and opportunities*.

Boschma, R. (2005). Proximity and Innovation: A Critical Assessment. *Regional Studies*, *39*(1), 61–74. https://doi.org/10.1080/0034340052000320887

Braunschweig, K., Eberius, J., Thiele, M., & Lehner, W. (2012). *Limits of Current Open Data Platforms*.

Çaldağ, M. T., & Gökalp, E. (2022). The maturity of open government data maturity: A multivocal literature review. *Aslib Journal of Information Management*, *74*(6), 1007–1030. https://doi.org/10.1108/AJIM-11-2021-0354

Cooper, B., & Glaesser, J. (2016). Analysing necessity and sufficiency with Qualitative Comparative Analysis: How do results vary as case weights change? *Quality & Quantity*, *50*(1), 327–346. https://doi.org/10.1007/s11135-014-0151-3

Dekkers, M., Polman, F., te Velde, R., & de Vries, M. (2006). *Measuring European Public Sector Information Resources*.

Dodds, L., & Newman, A. (2015). *A guide to Open Data Maturity Model: Assessing your open data publishing and use*. Open Data Institute.

Doloreux, D., & Parto, S. (2004). *Regional Innovation Systems: A Critical Review*.

Eurostat. (2024). Human resources in science and technology. Retrieved from https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Human_resources_in_science_and_technology

Gerrits, L., & Verweij, S. (2016). Qualitative Comparative Analysis as a Method for Evaluating Complex Cases. *Zeitschrift Für Evaluation*.

Gerrits, L., & Verweij, S. (2018a). Chapter 2: The case. In *The evaluation of complex infrastructure projects: A guide to qualitative comparative analysis* (pp. 15–51). Retrieved from https://www-elgaronline-com.eur.idm.oclc.org/monochap/9781783478415/chapter02.xhtml

Gerrits, L., & Verweij, S. (2018b). Chapter 3: Calibration. In *The evaluation of complex infrastructure projects: A guide to qualitative comparative analysis* (pp. 52–80). Retrieved from https://www-elgaronline-com.eur.idm.oclc.org/monochap/9781783478415/chapter03.xhtml

Gerrits, L., & Verweij, S. (2018c). Chapter 4: Comparison. In *Transport Economics, Management, and Policy*. *The evaluation of complex infrastructure projects: A guide to qualitative comparative analysis* (pp. 81–118). Cheltenham, UK: Edward Elgar Publishing.

Gössling, T., & Rutten, R. (2007). Innovation in Regions. *European Planning Studies*, *15*(2), 253–270. https://doi.org/10.1080/09654310601078788

Hossain, M. A., Dwivedi, Y. K., & Rana, N. P. (2016). State-of-the-art in open data research: Insights from existing literature and a research agenda. *Journal of Organizational Computing and Electronic Commerce*, *26*(1–2), 14–40. https://doi.org/10.1080/10919392.2015.1124007

Ide, T., & Mello, P. A. (2022). QCA in International Relations: A Review of Strengths, Pitfalls, and Empirical Applications. *International Studies Review*, *24*(1), viac008. https://doi.org/10.1093/isr/viac008

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management*, *29*(4), 258–268. https://doi.org/10.1080/10580530.2012.716740

Jetzek, T., Avital, M., & Bjorn-Andersen, N. (2014). Data-Driven Innovation through Open Government Data. *Journal of Theoretical and Applied Electronic Commerce Research*, *9*(2), 100–120. https://doi.org/10.4067/S0718-18762014000200008

Kitchin, R. (2014). Open and Linked Data. In *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*. 1 Oliver's Yard, 55 City Road, London EC1Y 1SP United Kingdom: SAGE Publications Ltd. https://doi.org/10.4135/9781473909472

Krakowiak-Bal, A., & Ziemiańczyk, U. (2017). Factors Influencing the Level of Regional Innovation—Qualitative Comparative Analysis. *Regional Barometer. Analyses & Prognoses*, *15*(4), 7–14. https://doi.org/10.56583/br.397

Kramer, L. (2024). What Is GDP and Why Is It So Important to Economists and Investors? Retrieved from Investopedia website: https://www.investopedia.com/ask/answers/what-is-gdp-why-its-important-to-economists-investors/

La France et ses territoires. (2021). Retrieved from Insee website: https://www.insee.fr/fr/statistiques/5039891?sommaire=5040030#graphique-figure3

Mello, P. A. (2012). *A Critical Review of Applications in QCA and Fuzzy-Set Analysis and a 'Toolbox' of Proven Solutions to Frequently Encountered Problems*.

Mello, P. A. (2021). Set-Theoretic analysis. In *Qualitative Comparative Analysis: An Introduction to Research Design and Application*.

Mello, P. A. (2022). *R Manual for QCA*. https://doi.org/10.7910/DVN/KYF7VJ

Miletić, A., Kuveždić Divjak, A., & Welle Donker, F. (2023). Assessment of the Croatian Open Data Portal Using User-Oriented Metrics. *ISPRS International Journal of Geo-Information*, *12*(5), 185. https://doi.org/10.3390/ijgi12050185

OECD Public Governance Policy Papers. (2023). *2023 OECD Open, Useful and Re-usable data (OURdata) Index*. OECD. Retrieved from OECD website: https://www.oecd-ilibrary.org/docserver/a37f51c3-en.pdf?expires=1715351482&id=id&accname=guest&checksum=F8B4593591BE5A986AABA5804A4DB864

Open Knowledge. (n.d.). What is Open Data? Retrieved from Open Data Handbook website: https://opendatahandbook.org/guide/en/what-is-open-data/

Opquast. (2011). *Opquast Bonnes pratiques Open Data*.

Poids économique des régions. (2024). Retrieved from La finance pour tous website: https://www.lafinancepourtous.com/decryptages/politiques-economiques/economie-francaise/le-poids-economique-des-regions/

Produit intérieur brut en 2022. (2024). Retrieved from Insee website: https://www.insee.fr/fr/statistiques/2012723#tableau-TCR_062_tab1_regions2016

Publications Office of the European Union. (2020a). *Sustainability of (open) data portal infrastructures: Open data portal assessment using user oriented metrics.* LU: Publications Office. Retrieved from Publications Office website: https://data.europa.eu/doi/10.2830/51660

Publications Office of the European Union. (2020b). *The economic impact of open data: Opportunities for value creation in Europe*. LU: European Commission. Retrieved from European Commission website: https://data.europa.eu/doi/10.2830/63132

Publications Office of the European Union. (2023). *Open data maturity report 2023.* LU: European Commission. Retrieved from European Commission website: https://data.europa.eu/doi/10.2830/384422

Ragin, C. C. (1999). *Using Qualitative Comparative Analysis to Study Causal Complexity*.

Ragin, C. C. (2000). Diversity-oriented research: Between complexity and generality. In *Fuzzy-set social science* (Nachdr.). Chicago: Univ. of Chicago Press.

Raymond, A. H., & Kouper, I. (2023). Open Governments, Open Data: Moving toward a Digital Commons Framework. In B. M. Frischmann, M. J. Madison, & M. R. Sanfilippo (Eds.), *Governing Smart Cities as Knowledge Commons* (1st ed., pp. 58–80). Cambridge University Press. https://doi.org/10.1017/9781108938532.005

Rihoux, B., & Grimm, H. M. (Eds.). (2006). *Innovative comparative methods for policy analysis: Beyond the quantitative-qualitative divide*. New York, NY: Springer.

Rihoux, B., & Ragin, C. (2009a). Crisp-Set Qualitative Comparative Analysis (csQCA). In *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*. 2455 Teller Road, Thousand Oaks California 91320 United States: SAGE Publications, Inc. https://doi.org/10.4135/9781452226569

Rihoux, B., & Ragin, C. (2009b). Qualitative Comparative Analysis using Fuzzy Sets (fsQCA). In *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*. 2455 Teller Road, Thousand Oaks California 91320 United States: SAGE Publications, Inc. https://doi.org/10.4135/9781452226569

Rubinson, C., Gerrits, L., Rutten, R., & Greckhamer, T. (2019). Presenting qualitative comparative analysis: Notation, tabular layout, and visualization. *Methodological Innovations*, *12*(2), 205979911986211. https://doi.org/10.1177/2059799119862110

Schneider, C. Q., & Wagemann, C. (2012). Set-theoric methods: The basics. In *Set-Theoretic Methods for the Social Sciences: A Guide to Qualitative Comparative Analysis* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139004244

Vall, R. (2020). *L'illectronisme ne disparaitra pas d'un coup de tablette magique!* Retrieved from https://www.senat.fr/rap/r19-711/r19-711-syn.pdf

Verweij, S. (2015). *Producing satisfactory outcomes in the implementation phase of PPP infrastructure projects: A fuzzy set qualitative comparative analysis of 27 road constructions in the Netherlands*.

Vickery, G. (2011). *Review of Recent Studies on PSI Re-Use and Related Market Development*. Retrieved from https://www.researchgate.net/publication/265573154_Review_of_Recent_Studies_on_PSI_Re-Use_and_Related_Market_Development

Welle Donker, F., & van Loenen, B. (2017). How to assess the success of the open data ecosystem? *International Journal of Digital Earth*, *10*(3), 284–306. https://doi.org/10.1080/17538947.2016.1224938

World Bank Group. (2015). *Open Data for Sustainable Development*. The World Bank. Retrieved from The World Bank website: https://thedocs.worldbank.org/en/doc/999161440616941994-0190022015/original/OpenDataforSustainableDevelopment.pdf

Zhang, H., Kang, F., & Li, H. (2022). Configurational path of successful entrepreneurship based on open government data: A QCA analysis. *Transforming Government: People, Process and Policy*, *16*(4), 391–404. https://doi.org/10.1108/TG-04-2022-0041

Zuiderwijk, A., Janssen, M., & Davis, C. (2014). Innovation with open data: Essential elements of open data ecosystems. *Information Polity*, *19*, 17–33. https://doi.org/10.3233/IP-140329

Appendix 1: IHS copyright form

In order to allow the IHS Research Committee to select and publish the best UMD theses, we kindly ask you to fill out and sign this copyright form and make it an annex to your final thesis.

Criteria for publishing:

1. A summary of 400 words should be included in the thesis.
2. The number of words should not exceed 15,000.
3. The thesis should be edited.

By signing this form you are indicating that you are the sole author(s) of the work and that you have the right to transfer copyright to IHS, except for items cited or quoted in your work that are clearly indicated.

I grant IHS, or its successors, all copyrights to the work listed above, so that IHS may publish the work in The IHS thesis series, on the IHS web site, in an electronic publication or in any other medium. IHS is granted the right to approve reprinting.

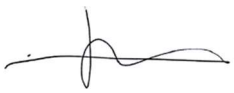
The author(s) retain the rights to create derivative works and to distribute the work cited above within the institution that employs the author.

Please note that IHS copyrighted material from The IHS Thesis series may be reproduced, up to ten copies for educational (excluding course packs purchased by students), non-commercial purposes, providing full acknowledgements and a copyright notice appears on all reproductions.

Thank you for your contribution to IHS.

Date : 01/08/2024

Your Name(s) : Juliette Navarre

Your Signature : 

IHS
Making cities work

Appendix 2: Regional assessment

	Open data portal maturity						
	Provision	Grade	**Metadata**	Grade	**Features**	Grade	**TOTAL**
Ile de France	933 datasets / 97 producers (nearly exclusivly public) Not really real-time datasets	2	Information page for every dataset with title, description, identification number, number of downloads, key words, type of licence, language of the dataset, last update, theme, producer (and sometimes how to contact), geographical coverage and if applicable temporal one, similar datasets Possibility to search a dataset per popularity = analytic ON THE DATASET	4	- Multiple filters - Information section - Tab with the data - Visualization section (not everytime) - Analytics section (not everytime and for the data not the use of it) - Export section with multiple formats to download the data - API section - Comment section - Possibility to see some re-use if there are some	4	**10**
Centre — Val de Loire	312 datsets / 28 producers (nearly exclusivly public) No real-time	2	Similar structure than IDF Except no similar datasets but the rest is super precise Possibility to search a dataset per popularity = analytics ON THE DATASET	4	Similar structure than IDF Re-use section but no re-use example found	4	**10**
Bourgogne-Franche-Comté	1 229 datasets / 50 providers (nearly exclusivly public) No real-time	2	- Description - Tab with a lot of info about dataset (including granularity, update, contact info, etc.) but even if a lot of boxes not all are filled - Analytics about downloads and views ON THE DATASET	4	- Multiple filters and format - Visualization en carte (mais de du tableau des données brutes) - API Example of re-use but no involvement of the users	3	**9**
Normandie	55 datasets / 2 producers It seems only public No real-time	1	- Description - Update, geographical coverage - Similar datasets - Few info about the producers - NO analytics and very few about the provider	2	- Multiple filters and format - Visualization + tab with data - API No re-use no user involvement	3	**6**
Hauts-de-France	293 datasets / 4 producers Only public No real-time	2	Very few details: - Description - Update - Provider with no further details But redirection towards another portal (geonetwork, more detailed but it is not the main one)	1	- Research filters, formats to download - Visualization (map and tab, graph), API - No re-use section, no user involvement	3	**6**
Grand Est	6673 datasets / 30 providers Only public Yes real-time datasets	3	- Description - Update - Geographical coverage - Contact for provider - analytics about the use (in the filters)	3	- Visualization - API section - Re-use section, even if no re-use found - NO user involvement	3	**9**
Pays de la Loire	1804 datasets / 14 producers Only public but emphasize on citizen in the re-use page Yes real time (and multiple)	4	Similar structure than IDF Except no similar datasets + no contact provider Possibility to search a dataset per popularity = analytics SUR LE DATASET	3	Similar structure than IDF Specific page dedicated to re-use = easier to find them	4	**11**
Bretagne	195 datsets / 24 producers 1 private No real time	3	Similar structure than IDF Except no similar datasets and no contact found Possibility to search a dataset per popularity = analytics ON THE DATASET	3	Similar structure than IDF Re-use example found but nut no comment section	3	**9**

Nouvelle-Aquitaine	8517 datasets / A lot of providers, some private Real-time datsets appear in the research but no data really on the portal only redirecing the user to another website with the data (ex towards Bordeau Métropole open data)	3	- Description - Update - Organization + contact - granularity - Keywords No analytics	2	- Research filters, formats But no visu, no API, no re-use, no user involvement	2	**7**
Occitanie	1246 datasets / 97 producers At leat 1 private (Weathernews) No real-time found	3	Similar structure than IDF Except no similar datasets and no contact found Possibility to search a dataset per popularity = analytics ON THE DATASET	3	Similar structure than IDF Re-use section but no re-use example found	4	**10**
Auvergne-Rhône-Alpes	6276 datasets / 25 providers Only public Real-time but it's only a redirection to another portal (and doesn't really work)	2	- Description - Update - Contact for providers - Geographical coverage - NO analytics about the use but popularity in the filters	3	Very few features because redirection to other portal (especially a geo portal for visualization) - No direct visu, no API, no user involvement	2	**7**
Provence-Alpes-Côte d'Azur	3583 datasets / A lot of producer and some are private Real-time datsets appear in the research but no data really on the portal only redirecting the user to another website with the data (ex towards Aix Métropole open data)	3	- Description - Update - Organization - granularity - Keywords - Similar datasets No analytics	2	- Filters - Multiple formats to download - Visu - API No user involvement	3	**8**
Corse	794 datasets / 64 producers Only public Some real-time datasets mais not updated	2	Similar structure than IDF Possibility to search a dataset per popularity = analytics ON THE DATASET	3	- Multiple filters - Information section - Tab with the data - Visualization section (not everytime) - Export section with multiple formats to download the data - API section No re-use, no user involvement	3	**8**
La Réunion	250 datasets / 78 producers 1 private found (SysDevRun) 1 real-time	4	Similar structure than IDF No similar datasets Possibility to search a dataset per popularity = analytics ON THE DATASET	3	- Multiple filters - Information section - Tab with the data - Visualization section (not everytime) - Export section with multiple formats to download the data - API section - Analytics section (sometimes) No re-use, no user involvement	3	**10**

Guyane	x	0	x	0	x	0	**0**
Martinique	x	0	x	0	x	0	**0**
Guadeloupe	123 datasets / 45 producers Only public 3 real-time but not working	2	Similar structure than IDF No similar datasets Possibility to search a dataset per popularity = analytics ON THE DATASET	3	- Multiple filters - Information section - Tab with the data - Visualization section (not everytime) - Export section with multiple formats to download the data - API section - Analytics section (sometimes) No re-use, no user involvement	3	**8**

	Regional governance								
	Policy	Grade	**Visibility**	Grade	**Engagement**	Grade	**Guidance**	Grade	**TOTAL**
Ile de France	The website only mentionned the national policies Ambition Smart Regions à laquelle la plateforme open data appartient	**2**	NO (juste pour région)	**0**	395 sur 932 = 42,4%	**2**	Digital inclusivity strategy: - "Le Plan régional d'inclusion numérique, juillet 2020, 367 000 euros, pour amener les débutants à l'autonomie et accompagner les exclus du numérique" - "30 associations ont été soutenues pour des actions relevant notamment de la lutte contre l'illectronisme et la fracture numérique dans les quartiers populaires pour un montant total de 251 013 euros" Financial regional support: "L'engagement de la Région Île-de-France se matérialise aussi avec un budget de deux millions d'euros maximum étalé sur quatre ans. Piloté par le PTNUM, ce budget permettra de donner les moyens aux équipes pour respecter les obligations légales et favoriser l'inclusion et l'accessibilité"	**1**	**5**
Centre — Val de Loire	-The website only mentionned the national policies - Regional digital strategy 2023-2028: "AXE 2, Objectif 8: Développer l'utilisation des données publiques et exploiter leur potentiel au bénéfice de l'innovation publique (ouverture des données, services numérique mobilités, mobilisation des données pour créer des services nouveaux ou les améliorer [type territoires intelligents et durables])" But no specific mention for open data	**1**	NO (juste pour région)	**0**	111 sur 312 = 35,6%	**2**	Digital strategy strategy: "AXE 4, Objectif 13: Favoriser l'inclusion numérique (réseau et soutien aux lieux de médiation numérique)" "AXE 4, Objectif 15: Agir pour l'égalité femmes-hommes dans le numérique (mentorat, guide de formations numériques, outils pédagogiques, appel à projets dédié à l'égalité sous toutes ses formes dans le numérique)" "4,5 Millions € sur les 3 prochaines années sur les usages numériques liés à la cybersécurité, à l'inclusion numérique, à la sobriété numérique, au développement des usages éthiques de la donnée sur les territoires et à la créativité numérique" --> But not specific to open data	**1**	**4**
Bourgogne-Franche-Comté	Digital regional strategy "soutenir le developpement et usage du numérique (pour citoyens), accélérer le déploiement d'infrastrucutures (accès internet notamment), accompagnement des collectivités dans la démarche de territoires intelligents et durables" Guide for properly openning data in 6 steps	**2**	NO (juste pour région)	**0**	39 sur 1229 = 3,2%	**1**	- Digital third-place - Network for digital inclusivity --> Focus on digital but not on open data	**1**	**4**
Normandie	Digital strategy: - "**Datalab** pour expérimenter des projets basés sur big data et open data" Très bref objectif de développer la relation citoyenne et les services numériques des collectivités entre autres par l'open data Objectif du bloc innovation: Construire une organisation performante de référence pour la Data : infra, service, régulation,open Data et construire le patrimoine de données normand (notion de patrimoine immatériel) --> Open data mentionned but not through a clear and focus objective	**1**	NO (juste pour région)	**0**	Hard to know	**1**	COMMUNOTIC network Digital third-place	**1**	**3**
Hauts-de-France	Innovation strategy 2014-2020 but no focus on open data, rather digital	**1**	NO (juste pour région)	**0**	Hard to know but nearly all	**2**	Digital inclusivity considered but for basic needs "Pass numérique: L'objectif de ce dispositif est de répondre aux enjeux majeurs de l'inclusion numérique et de la lutte contre l'illectronisme, et de permettre aux utilisateurs de devenir plus autonome dans leurs usages du numérique. Il permet de payer une séance d'initiation ou d'accompagnement au numérique, sur des sujets très variés"	**1**	**4**
Grand Est	Project OD4Growth "renforcer la disponibilité et le traitement de l'Open Data (donnée ouverte) pour accompagner la croissance locale et la transformation urbaine" MAIS "projet porté par l'UE avec aussi d'autres villes européennes" --> Clear focus on developping open data through european projects	**2**	YES	**1**	264 sur 6673 = 4%	**1**	Focus on supporting company and not people Nothing else mentioned	**0**	**4**

Pays de la Loire	Stratégie innovation 2014-2020 trouvée mais aucun mention open data Vraiment rien même en cherchant données ouvertes et open data dans la barre de recherche	**0**	NO (juste pour région)	**0**	634 sur 1817 = 35%	**2**	Digital training for basics The rest is focused on helping company	**1**	**3**
Bretagne	Page on the regional website to present data and the different tools available (portal, etc.) Regional digital strategy with focus on open data	**2**	NO (juste pour région)	**0**	162 sur 194 = 83,5%	**3**	Support: "Le Visa Internet Bretagne s'adresse tous les bretonnes et bretons qui souhaitent s'initier gratuitement aux usages numériques, quels que soient leur âge ou leur situation. La Région soutient également financièrement les lieux labellisés Espaces publics numériques, qui mettent en place ces initiations"	**1**	**6**
Nouvelle-Aquitaine	No strategy found No mention of open data on the website	**0**	NO (juste pour région)	**0**	5 sur 8525 = 0,06%	**1**	Digital training for basics	**1**	**2**
Occitanie	Digital strategy: "Open Data Région Occitanie: projets de développement du portail http://data.laregion.fr ou d'ouverture de jeux de données impliquant des traitements en base et tendant vers l'industrialisation de flux sortants de données" "Usages des données ouvertes: projets de développement de services exploitant les données ouvertes ou structurant des modèles partagés de données (transports, alimentation...)" "Stratégie Data: contribution de l'Open Data à la stratégie Data en Occitanie incarnée par Occitanie Data"	**2**	NO (juste pour région)	**0**	106 sur 1244 = 8,5%	**1**	Digital training for basics	**1**	**4**
Auvergne-Rhône-Alpes	No mention of open data on the website Open data platform really separated from the website	**0**	YES	**1**	1067 sur 6291 = 17%	**1**	Campus Région with digital trainings for all ages and profiles	**1**	**3**
Provence-Alpes-Côte d'Azur	No mention of open data on the website Open data platform really separated from the website	**0**	YES	**1**	154 sur 3891 = 4%	**1**	Digital training centers	**1**	**3**
Corse	Digital Strategy: SMART ISULA Data focus with some action plans but it is more about data in general than open data	**1**	NO	**0**	382 sur 797 = 48%	**2**	Programm for digital inclusivity Within the strategy, focus on train not only population but also local bodies	**2**	**5**
La Réunion	No mention of open data Focus more on company than citizens	**0**	NO (juste pour région)	**0**	30 sur 250 = 12%	**1**	Digital inclusivity policy	**1**	**2**
Guyane	Porject of open data portal in the digital strategy	**2**	x	**0**	x	**0**	Focus on digital inclusivity within the strategy	**1**	**3**
Martinique	Regional website under construction so difficult to get info Some initiatives were found but not implemented by the regional bodies	**0**	x	**0**	x	**0**	Oriented more towards company than citizens	**0**	**0**
Guadeloupe	Specific open data axis within the digital strategy	**2**	NO (juste pour région)	**0**	4 sur 123 = 3,2%	**1**	Digital training centers	**1**	**4**

Appendix 3: R code

```
## 711520 Thesis
## Open data and Economic growth

## Remove everything
rm(list = ls())

## Set working directory
setwd("C:/Users/julie/Documents/3- ROTTERDAM/1- Cours/6- Thesis/3- R")
##dir()

##dir.exists("C:/Users/julie/Documents/3- ROTTERDAM/1- Cours/6- Thesis/3- R")


## Import data file
thesis1 <- read.csv("DataThesis.csv",fileEncoding = "UTF-8",sep=";", header = T)
#"header" specifies whether the first row contains variable names
View (thesis1)
```

```
      ### DESCRIPTIVE STATISTICS ####

#str(thesis1)
#head(thesis1)
#names(thesis1)
summary(thesis1)

  ## Portal maturity
#mean(thesis1$Portal.Maturity)
#summary(thesis1$Portal.Maturity)
quantile(thesis1$Portal.Maturity,c(0.25,0.5,0.75))
#sd(thesis1$Portal.Maturity)
#hist(thesis1$Portal.Maturity)

  ## Regional governance
#mean(thesis1$Regional.Governance)
#summary(thesis1$Regional.Governance)
quantile(thesis1$Regional.Governance,c(0.25,0.5,0.75))
#sd(thesis1$Regional.Governance)
#hist(thesis1$Regional.Governance)

  ## HRST
#mean(thesis1$hrst)
#summary(thesis1$hrst)
quantile(thesis1$hrst,c(0.25,0.5,0.75))
#sd(thesis1$hrst)
#hist(thesis1$hrst)

  ## GDP
#mean(thesis1$gdp)
#summary(thesis1$gdp)
quantile(thesis1$gdp,c(0.25,0.5,0.75))
#sd(thesis1$gdp)
#hist(thesis1$gdp)
```

```
      ### QCA ###

    ## CALIBRATION of variables into conditions ##

##install.packages("QCA") #To install package
##install.packages("SetMethods")
library(QCA) # To load QCA package
library(SetMethods)
```

```
  ## Calibrating Portal Maturity
thesis1$MATURITY <- round(calibrate(thesis1$Portal.Maturity, thresholds = "e=3, c=7.5, i=9", include.lowest = TRUE), 2)

# Replace 0.50 with 0.51
#thesis1$MATURITY[which(thesis1$MATURITY == 0.5)] <- 0.51

  ## Calibrating Regional Governance
thesis1$GOVERNANCE <- round(calibrate(thesis1$Regional.Governance, thresholds = "e=0, c=3.4, i=4.7", include.lowest = TRUE), 2)

# Replace 0.50 with 0.51
#thesis1$GOVERNANCE[which(thesis1$GOVERNANCE == 0.5)] <- 0.51


  ## Calibrating HRST
thesis1$HRST <- round(calibrate(thesis1$hrst, thresholds = "e=30, c=39.1, i=50", include.lowest = TRUE), 2)

# Replace 0.50 with 0.51
thesis1$HRST[which(thesis1$HRST == 0.5)] <- 0.51


## Calibrating outcome GDP
thesis1$GDP <- round(calibrate(thesis1$gdp, thresholds = "e=26375, c=158249, i=263748", include.lowest = TRUE), 2)
```

```r
#Display the set-membership values
thesis1_qca <- thesis1[, c(1:2, 7:10)]
View(thesis1_qca)


    ## Analysis of NECESSITY  ###

## Performed for each individual condition
## Goal: checking superset relations

conds <- subset(thesis1_qca, select=c(3:5)) # select conditions
names(conds)

#OUTCOME
nec_y <- QCAfit(conds, thesis1_qca$GDP, names(conds), necessity = TRUE) # SetMethods package
#nec_y <- pof(conds, thesis1_qca$GDP, relation = "nec") # QCA package
nec_y

#write.csv(nec_y, "Necessity.csv")

## Alternatively, it is possible to perform the analysis of necessity for each individual conditions separately:
#pof("maturity_calibrated <= GDP", data = thesis1_qca) # pof: parameters of fit
#pof("~maturity_calibrated <= GDP", data = thesis1_qca)

  ##Plot necessity
thesisVisu <- thesis1[c(7:10)]
View(thesisVisu)

# For HRST condition
XYplot(HRST, GDP, thesisVisu, relation = "necessity", mguides = TRUE, jitter = TRUE, enhance = FALSE, model = FALSE)


    ## TRUTHTABLE ##


## Assign rownames as a combination of case name with outcome set membership value
rownames(thesis1_qca) <- paste(thesis1_qca[[2]],thesis1_qca[[6]],sep="|")

## Generate for the occurrence of the outcome (Y)
tt_thesis1 <- truthTable(thesis1_qca, outcome = "GDP",
                      conditions = "MATURITY, GOVERNANCE, HRST",
                      incl.cut = 0.75, complete = TRUE, show.cases = TRUE,
                      sort.by = "incl, n+")
tt_thesis1
#write.csv(tt_thesis1$tt, "TruthTable.csv")

# Show the membership to all the possible configurations of condition.
# Only one is higher than 0.5 and is the one displayed in the truth table
tt_thesis1$minmat

#Deleting the contradictory case
thesis1_qca <- thesis1_qca[-6 , ]
View(thesis1_qca)


  # MINIMIZATION

## Conservative solution for the occurrence of the outcome
c_sol_y <- minimize(tt_thesis1, include = "1", details = TRUE,  use.tilde = TRUE)
c_sol_y

# Parsimonious solution for the occurrence of the outcome
c_soly <- minimize(tt_thesis1, include = "?", details = TRUE,  use.tilde = TRUE)
c_soly


    ## GRAPHS ##

# Create plots for the conservative solution
# Create terms in the dataset to make plots
thesis1_qca$gov <- 1 - thesis1_qca$GOVERNANCE # create ~GOV


thesis1_qca$path1_csol_y <- pmin(thesis1_qca$MATURITY, thesis1_qca$HRST) # MAT*HRST
thesis1_qca$path2_csol_y <- pmin(thesis1_qca$gov, thesis1_qca$HRST) # ~GOV*HRST
thesis1_qca$csol_y <- pmax(thesis1_qca$path1_csol_y, thesis1_qca$path2_csol_y) # overall solution

thesis1_qca$hrst_csol_y <- pmax(thesis1_qca$HRST) # HRST

# Plot for path1 in csol_y
path1_csol_y <- xy.plot("path1_csol_y", "GDP", data = thesis1_qca,
                        main = paste("XY Plot for Pathway 1 MAT*HRST"),
                        xlab = "High portal maturity and high share in skilled population",
                        ylab = "High GDP",
                        labs = rownames(thesis1_qca),
                        jitter = TRUE, fontsize = 3
)
path1_csol_y

# Plot for path2 in csol_y
path2_csol_y <- xy.plot("path2_csol_y", "GDP", data = thesis1_qca,
                        main = paste("XY Plot for Pathway 2 ~GOV*HRST"),
                        xlab = "Low governance and high share in skilled population",
                        ylab = "High GDP",
                        labs = rownames(thesis1_qca),
                        jitter = TRUE, fontsize = 3
)
path2_csol_y
```

```
# Plot for the overall solution
rownames(thesis1_qca) <- paste(thesis1_qca[[2]])

csol_y <- xy.plot("csol_y", "GDP", data = thesis1_qca,
                        main = paste("XY Plot for the overall minimized solution formula"),
                        xlab = "HRST(MAT*~GOV)",
                        ylab = "GDP",
                        labs = rownames(thesis1_qca),
                        jitter = TRUE, font = "sans", fontsize = 3
)
csol_y


# Plot for the parsimonious solution
# Plot for hrst in csol_y
hrst_csol_y <- xy.plot("hrst_csol_y", "GDP", data = thesis1_qca,
                        main = paste("XY Plot"),
                        xlab = "Solution term HRST",
                        ylab = "Outcome High GDP",
                        labs = rownames(thesis1_qca),
                        jitter = TRUE, fontsize = 3
)
hrst_csol_y

# ------------------------------------------------------------------------------------

    #NON-OUTCOME

#Necessity
nec_ny <- QCAfit(conds, thesis1_qca$GDP, names(conds), necessity = TRUE, neg.out = T)
nec_ny
#write.csv(nec_ny, "NecessityNY.csv")

##Plot necessity
thesisVisu <- thesis1[c(7:10)]
View(thesisVisu)

XYplot(~HRST, ~GDP, thesisVisu, relation = "necessity", mguides = TRUE, jitter = TRUE, enhance = FALSE, model = FALSE)


## Generate truthtable for the non-occurrence of the outcome (y or ~Y)
rownames(thesis1_qca) <- paste(thesis1_qca[[2]],thesis1_qca[[6]],sep="|")

tt_thesis1_ny <- truthTable(thesis1_qca, outcome = "~GDP",
                         conditions = "MATURITY, GOVERNANCE, HRST",
                         incl.cut = 0.75, complete = TRUE, show.cases = TRUE,
                         sort.by = "incl, n")
tt_thesis1_ny
#write.csv(tt_thesis1_ny$tt, "TruthTableNY.csv")


## Conservative solution for the non-occurrence of the outcome
c_sol_ny <- minimize(tt_thesis1_ny, include = "1", details = TRUE,  use.tilde = TRUE)
c_sol_ny
```

Appendix 4: Analysis for the non-occurrence of the outcome

1. Necessary conditions

Similarly to the necessity test for the occurrence of the outcome, analysing the conditions necessary for the absence of the outcome could provide some interesting insights for the research. The same consistency threshold of 0.9 applies to determine whether a condition is necessary for the non-occurrence of the outcome.

As presented in *Table 13*, the condition ~HRST is necessary for the absence of the outcome. In other words, whenever the regional GDP is low, the share of skilled users among the regional population will be low also. Regarding the coverage and the relevance, there is no clear threshold defined but the usual rule of thumb indicates that below 0.5 the necessary condition may be trivial (Mello, 2021), it is not the case in our analysis.

	Cons.Nec	Cov.Nec	RoN
MATURITY	0.708	0.622	0.565
GOVERNANCE	0.699	0.680	0.670
HRST	0.285	0.509	0.806
~MATURITY	0.418	0.746	0.889
~GOVERNANCE	0.518	0.775	0.872
~HRST	0.947	0.832	0.745

Table 13: Necessity table for the non-occurrence of the outcome

This first finding tends to reinforce the results provided in Chapter 4 that the regional GDP and the share of skilled population are closely related.

2. Truth table

The truth table for the non-occurrence of the outcome is presented in *Table 14*.

Maturity (MAT)	Governance (GOV)	HRST	OUT	n	incl	PRI	Cases
1	1	0	1	5	0.912	0.852	Centre - Val de Loire, Bourgogne-Franche-Comté, Bretagne, Corse, Guadeloupe
0	0	0	1	3	0.852	0.787	Normandie, Guyane, Martinique
1	0	0	1	2	0.840	0.667	Pays de la Loire, La Réunion
0	1	0	1	1	0.768	0.595	Hauts-de-France

1	0	1	0	1	0.655	0.075	Provence-Alpes-Côte d'Azur
1	1	1	0	2	0.581	0.210	Ile de France, Occitanie
0	0	1	0	2	0.534	0.022	Nouvelle-Aquitaine, Auvergne-Rhône-Alpes
0	1	1	?	0	-	-	

Table 14: Truth table for the non-occurrence of the outcome

Four combinations of conditions are sufficient to produce the non-outcome:

MAT*GOV*~HRST → ~GDP

~MAT*~GOV*~HRST → ~GDP

MAT*~GOV*~HRST → ~GDP

~MAT*GOV*~HRST → ~GDP

Those configurations all passed the consistency threshold of 0.75 and all present a high coverage and PRI which comfort in the reliability of the results. The following solution term can summarize the truth table:

MAT*GOV*~HRST + ~MAT*~GOV*~HRST + MAT*~GOV*~HRST + ~MAT*GOV*~HRST → ~GDP

There is one logical remainder: the configuration (011) does not present empirical evidence within the research sample.

3. Minimized solution formula

In order to conduct the minimization, it has to be decided whether the logical remainder should be included or not in the analysis.

The four sufficient rows of the truth table tend to suggest that the configuration (011) could not produce the non-outcome. In fact, the condition ~HRST is shared by the four sufficient pathways, so it questions the plausibility of a configuration including the condition HRST to produce the non-outcome ~GDP. Besides, as explained by Mello, in order to avoid contradictory counterfactual between the analysis of the outcome and the one of the non-outcome, which are conducted separately, the use of the logical remainder should be limited to the analysis where it is plausible (2021). In our case, the logical remainder was included in the analysis of the outcome because it was highly plausible that despite the lack of empirical evidence the configuration could produce the outcome. Consequently, this analysis is conducted with the conservative minimization formula, excluding the logical remainder. The results are presented in *Table 15*.

Path	incls	PRI	covS	covU	Cases
Conservative solution					
~HRST	0.865	0.824	0.945	-	Normandie, Guyane, Martinique, Hauts-de-France, Pays de la Loire, La Réunion, Centre - Val de Loire, Bourgogne-Franche-Comté, Bretagne, Corse, Guadeloupe

Table 15: Minimization results for the non-occurrence of the outcome

The resulting minimization solution formula is: **~HRST → ~GDP**

It highlights that ~HRST is a necessary and sufficient condition to produce the non-outcome ~GDP. The consistency and coverage results are high enough to comfort in the robustness of the result. In fact, the coverage highlights that the non-outcome is explained at 94.5% by the configuration ~HRST.

4. Interpretation & Discussion

The analysis of the non-outcome is often neglected but it represents a good practice to validate the theoretical argument and strengthen the confidence in the results by comparing both analysis of the outcome and the non-outcome (Mello, 2012). It also emphasizes the causal asymmetry possible in QCA.

As part of this research, the results obtained for the analysis of the non-occurrence of the outcome confirm the analysis conducted in Chapter 4. Both truth tables are symmetric as well as minimized solution formula:

- Outcome: HRST → GDP
- Non-outcome: ~HRST → ~GDP

A high share of skilled population contributes to a high GDP, and a low share produces a low GDP.

While causal asymmetry is common in QCA due to the complexity of social phenomena, symmetric results can occur. In our case, we can assume that the relationship between the share of high-skilled people among the regional population and the GDP is so strong that its effect overpasses the other conditions. To that extent, it could be interesting to run an analysis adding or modifying some conditions in order to capture additional effects. A suggestion could be for example to differentiate the highly educated people from the IT-skilled people (as the dataset HRST merges both) to nuance the results and gain deeper knowledge into the regional mechanisms. Another possibility could be also to target a more specific outcome and study for example the regional innovation capacity. In fact, it could be easier to capture the direct effect of open data use on innovation than on economic growth, which stands as a more indirect outcome, a consequence of the innovation capacity.

The non-outcome analysis adds some more understandings about the cases and allows to reflect further about the configuration of the research, that despite emphasizing the key role of users, may be too simplistic to fully capture the impact of open data on economic growth.

Institute for Housing and
Urban Development Studies of
Erasmus University Rotterdam