Conceptualizing a national data infrastructure for Switzerland

Beat Estermann*, Marianne Fraefel, Alessia C. Neuroni and Jürgen Vogel
Bern University of Applied Sciences, Switzerland

Abstract. A national data infrastructure (NDI) provides data, data-related services and guidelines for the re-use of data to individuals and organizations. It facilitates efficient sharing of data, supports new business models, and is thus a key enabler for the digital economy, open research, societal collaboration and political processes. While several European countries have taken steps to set up data infrastructures cutting across institutional silos, approaches vary, and there is no common understanding of what a NDI exactly comprises. In Switzerland, activities are still at a conceptual stage. In order to foster a shared vision of what a NDI is about, stakeholder interviews were carried out with representatives of public administration, research, civil society, and the private sector. There is broad consensus among key stakeholders that a NDI is to be conceived as a nationwide distributed technical infrastructure allowing the sharing of data, based on predefined rules. Our findings also suggest that the notion of a NDI should be approached from four perspectives: a big data, a base register, an open data, and a mydata perspective. For its implementation, effective coordination across several dimensions (ethical, legal, political, economical, organizational, semantical, and technical) is crucial, which calls for a truly multidisciplinary approach.

Keywords: National data infrastructure, big data, open data, linked data, infrastructure resources

1. Introduction

Data play the role of an infrastructure resource in that they generate value when used as inputs into a wide range of productive processes the outputs of which are often public and non-market goods [54]. Just as electricity, streets and water are core infrastructures that serve citizens, companies and governments alike, so too a national data infrastructure (NDI) responds to a community-wide need in the big data era, similar to education, healthcare, and public services [66].

A NDI provides data, data-related services, and guidelines for data-reuse to public and private organizations and to citizens. As such, it allows the efficient sharing and effective use of data across organizational boundaries. It is beneficial for data providers and consumers alike, supports new business models, and is thus a key enabler of the digital society, enhancing societal collaboration and supporting political processes.

Fostering data-driven innovation requires the attention of policymakers. There is a need for a coherent data governance that balances openness and public benefit on the one hand, and privacy and data security on the other [54]. Good data governance not only needs to address the question of who gets access to what data for what purpose and under what conditions, but should also assign responsibilities and retribution mechanisms for data maintenance and enhancement in order to ensure the sustainability of the common data infrastructure. Furthermore, a NDI needs to be conceived so as to ensure interoperability

---

*Corresponding author: Beat Estermann, Bern University of Applied Sciences, Switzerland. E-mail: beat.estermann@bfh.ch.

and alignment with similar efforts at the international level. The European Commission perceives data policy to be one of the key issues to enhance the long-term competitiveness of Europe on a global scale. Data and data-related infrastructures for business, administration and society are thus an essential part of the European Digital Single Market Strategy [22]. Only recently, the Swiss government has adopted a "Digital Switzerland" Strategy that among other goals envisages the establishment of a NDI, the promotion of an equitable information ecosystem, the empowerment of Swiss citizens so that they have control over their own data, and the formulation of a coherent and future-oriented data policy for Switzerland [25].

As the governance issues related to the establishment and the maintenance of a NDI are still far from being fully understood by the key stakeholders in Switzerland, we carried out an exploratory study on the topic. The main goal of the study was to foster a shared vision among key stakeholders regarding the data infrastructure that is to be established and to draw their attention to the governance issues surrounding it. To this end we conducted sixteen qualitative interviews with stakeholders from various sectors. The questions guiding our research were: What are the key building blocks of a national data infrastructure? And how should inter-organizational collaboration be organized in view of the establishment and the maintenance of a national data infrastructure? – The study thus contributes to the conceptualization of a NDI and sheds light on the governance issues related to its establishment, providing valuable inputs in view of the formulation of a future-oriented data policy. At this stage, the study addresses relevant issues from a strategic perspective and does therefore not provide a detailed discussion of technical challenges.

The paper is structured as follows: In Section 2 we provide a summary of relevant research related to data infrastructures. Section 3 contains an overview of existing initiatives and activities towards an NDI specifically in Switzerland. Section 4 describes the methodological approach for data collection. In Section 5 we present notable findings from the interviews concerning the current understanding of a NDI and the outlook on its establishment, including stakeholder roles, coordination needs, and required knowledge. Our research findings suggest that the notion of a NDI should be approached from four distinct perspectives: a big data perspective, a base register perspective, an open data perspective, and a mydata perspective, which are covered in Section 6. We summarize our findings in Section 7 by addressing the research questions from above and by suggesting next steps.

2. Relevant research in the field

Over the last couple of years, several authors have addressed governance issues related to the establishment and the maintenance of a NDI or similar data infrastructures [41,45,51,52]. It appears from these publications that a multi-disciplinary approach is needed in order to gather a holistic view of the challenges at hand, drawing on different strands of research in disciplines such as political and social sciences, law, economics, and computer science. This section summarizes present considerations in the different strands of research relevant to the discussion of governance issues surrounding a NDI.

In order to clarify the notion of a NDI we set out by drawing on literature related to (data) infrastructure development. Infrastructures for data sharing and re-usage are built for different areas of application. It is therefore important to view them in relation to context. From a socioeconomic viewpoint, recent developments in big data analytics are of particular importance. One of the characteristics of big data is the growing variety of data that is being integrated and tapped into for all sorts of analyses. In that respect, technological trends and challenges should be taken into consideration, such as the ones related to open and linked data. When it comes to analyzing personal data, privacy issues are a major concern, a topic that should be approached both from a technical and a legal perspective. We conclude the section

with a short overview of contributions regarding the notion of data governance at the societal level, which may serve as an umbrella to integrate the different research perspectives.

2.1. Existing notions of data infrastructure

Neuroni et al. explore the notion of a national data infrastructure and related concepts, observing that the notion of a national data infrastructure bears connections to and overlaps with different concepts dedicated to infrastructure development in a digital environment, such as e-government infrastructures, national information infrastructures, and open data infrastructures. They also point to the fact that there is no common understanding in the research literature as to what a data infrastructure comprises [45,51].

There are at least three strands of research and practice that can be drawn upon when it comes to conceptualizing a NDI: e-government research, research on open government data (OGD), and the initiatives in various countries regarding the establishment of a "national information infrastructure" (NII) or a "national data infrastructure" (NDI).

In the context of e-government, the development of shared infrastructures enabling interagency collaboration is a core issue when it comes to improving public service provision [3,26]. This is illustrated by the European Interoperability Framework (EIF), which provides a conceptual model for public services and considers data-related services as a basic component for service provision [20]. The focus is on base registers that are legally controlled and maintained by public administrations and provide authentic sources of information on items such as persons, companies, buildings, or roads. According to the European Commission this information ought to be made available for re-use, and appropriate mechanisms for managing access and control should be put in place (ibid.). This is consistent with the view that authentic data sources are a key enabler for enhanced service provision as they allow improving user experience and administrative efficiency [9,12].

In the context of OGD, the focus is on infrastructures that allow third parties to re-use government data. As Neuroni et al. note, OGD benchmarks assess data availability and accessibility, user support and functionalities for stakeholders. Availability of data mostly relates to coverage of sectoral data (education, health, etc.), while the provision of basic government data (key registers) is hardly a topic. Also, the focus of OGD strategies generally is on public sector data that can be made available in an open access regime [51].

In contrast, NIIs or NDIs can comprise data from various types of data providers, such as government, businesses, or non-profit organizations. Also, they may be openly licensed, be made available to a restricted group of stakeholders, or be closed. Several European countries have adopted the concept of a NDI/NII. The different concepts have in common the sharing of core government data sets within and outside public agencies and the stimulation of their use across boundaries. The main motivation lies in the increase of public efficiency [70]. The concept of a NDI/NII can further be seen as a management framework for the strategically most important data held by government; it is composed of a set of guiding principles, a data set list, a governance structure, and baseline quality criteria [8]. Some initiatives or policies conceptually strengthen connections between government data held in base registers and OGD. This is the case for instance in the UK, in Denmark, and in the Netherlands [51].

No matter from which point of view a data infrastructure is approached, it can be understood as a community-wide need, respectively a good that should be provided as a public service [66]. Data itself has become an essential driver of innovation, and it can be considered as one of our society's central infrastructure resources [54]. As Frischmann observes, from an economic point of view, infrastructure resources are fundamental resources that do not get consumed when being used, and generate value when

used as inputs into productive processes. As their outputs are often public goods that generate positive impacts for society, it may be socially desirable to manage them in an openly accessible manner. Data meet the following characteristics of infrastructure resources: first, they may be consumed in a non-rivalrous fashion for some appreciable range of demand; second, the social demand for data is driven primarily by downstream productive activities that require them as an input; and third, they may be used as an input into a wide range of (private, public or social) goods and services [27,29].

2.2. Potentials and policy implications of big data and the rise of data-driven services

As a result of rapidly increasing computing capacity, storage capacity and communication bandwidth, we are presently experiencing a situation where data collection, analysis, and storage are characterized by ever increasing volumes, variety, and velocity as well as by new methods of analysis, which are accompanied by a rise of data-driven services. These developments are often subsumed under the term "big data", while other terms, like "ubiquitous computing" or "ambient intelligence" point to the ever-present nature of computers and data analytics in our lives [6]. These developments are highly relevant in the context of a NDI – not only with regard to potential uses of the data and thus the usefulness of a NDI, but also with regard to the precautionary measures that need to be taken in the area of data protection and with regard to the ethical implications of big data. Thus, government and the public sector are concerned by these developments in various ways:

The public sector is an important data producer and thus an important source of information for big data analytics. Public sector sources, such as classifications and statistics need to meet high quality requirements, are often produced on a regular basis, and are therefore an essential resource for providing stability to big data models [73]. In turn, big data technologies may change the way official statistics are produced [71]. It is widely acknowledged that governmental data-provision may contribute to socioeconomic development and growth [53]. In 2014, the European Commission has released its vision of a data-driven economy, addressing the relevance of big data and calling for the adoption of open data policies and the development of public data infrastructures [21].

At the same time, the public sector is also a data user and as such may benefit from the implementation of big data concepts. Opportunities comprise: increasing efficiency and saving costs, increasing transparency, innovating service provision, supporting and improving decision-making, detecting and preventing mistakes and frauds, reducing security threats, and conducting policy impact assessments [7,16,23]. Typical domains include traffic planning and control, public security and intelligence services, social welfare and health care, assessment of environmental risks, disaster management, and strategic management. Depending on the application domain and the type of data generated, big data analytics in the public sector is closely related to the concept of smart cities [10,49].

There are several accounts of practical cases of (big) data analytics in the public sector [16,43,69]. As scholars point out however, many government data projects have so far rather been focusing on structured and stored data, instead of using e.g. real-time, or in-motion data [14]. Current gaps in exploiting the potential of big data in the public sector are also acknowledged by the European Commission, who in 2015 has launched an action on big data for public administrations. Its main objectives are to *"identify concrete big data and open knowledge opportunities and requirements in public administrations and in specific policy contexts"* and to *"promote cooperation (. . .) in order to accelerate the data-driven transformation"* [21].

Big data analytics not only offers opportunities for the public sector, but also raises challenges [10, 35,61]. The first set of challenges relates to big data governance: establishing a data-driven decision-making culture (create public value through big data), developing agreements for integrating data sources

across organizations to generate big data, and specifying roles and responsibilities for managing big data. The second set relates to big data implementation: recruiting specialized personnel (data scientists) and implementing adequate technological solutions. The third challenge relates to risk management: ensuring privacy in view of data aggregation (individual profiling) and ensuring security (combination of disparate data sources).

Irrespective of the field of application (private vs. public), big data raises a range of policy and ethical issues. Both, the OECD and the European Commission stress the need to provide adequate framework conditions for big data, including policy frameworks "*for the generation, collection, transport and use of data*" [53]. Regulatory issues in particular relate to the areas of user empowerment and privacy protection [21,53], whereas the longer-term implications of the ubiquitous use of data analytics raise important ethical issues, e.g. related to shifts in the balance of political and economic power, changes in our values and motives, the violation of personal borders by new surveillance and search technology, or changing attitudes towards the world that surrounds us [6,61].

2.3. Potentials and challenges of open and linked data

Another development, partly related to the one of big data is the breaking-up of data silos, i.e. the growing importance of the secondary use of data and the integration of data from various sources. This has also been recognized by the open data movement. Thus, while the call to open up public sector information can be seen as a logical extension of the freedom of information regulations that have been adopted by many countries since the 1990s, the open data movement is also driven by a technical and economical vision: transaction costs related to the secondary use of data are to be reduced and a *semantic web* is to be created by linking many 'open' datasets from various sources. Thus, 'linked open data' will serve as an infrastructure resource for third parties to build value-added services on top of it, such as new combinations of data, visualizations, or other data-driven services [2,5].

While open access to publicly funded data is generally associated with many benefits, such as greater returns from public investment, increased value creation through the downstream use of outputs, better access for policy makers to data needed to address complex problems, increased transparency, or improved citizen participation, there is still a considerable gap between these promises and reality. Current challenges comprise both data publication and access impediments (related to creating, opening, finding and obtaining the data) and data use impediments (preventing their effective use). Furthermore, the transformative nature of open data is more elusive than might be expected, and the implementation of a culture of open government difficult to achieve [39,79].

When combining data from different sources, a difficult obstacle lies in the need to first understand the particular encoding (syntax) and meaning (semantics) of the individual data fields, which are often not well-documented and require substantial domain knowledge. Even worse, for the same type of data, different sources may use different syntax and/or semantics. As a result, larger projects with several data sources require labor-intensive and error-prone data integration. As case studies in different industries reveal, most integration projects suffer from syntax-related issues and encounter semantics-related problems [24]. In order to overcome these challenges, experts and OGD researchers advocate publishing data as so-called linked data [5,30].

A vital prerequisite for linked data is the availability of standardized metadata in the form of ontologies [34]. While a substantial number of ontologies have been defined by standardization bodies and academic communities, many domains still lack appropriate ontologies. Also, serializing the data as linked data is difficult to automate, and induces considerable overheads in terms of storage and processing

power required to handle the data. On top of that a number of fundamental technical challenges remain: How to trace provenance on such platforms where pieces of data originating from various sources can be arbitrarily combined? – Even though RDF triples often contain a fourth piece of information to identify their source (or context), tracking such source information throughout the query processing pipeline is still an open research challenge [78]. Another fundamental problem that attracts the attention of researchers is the lack of data integration in data lakes, where multiple data types can coexist but cannot be interrelated or joined automatically [13,15]. And finally, user-friendly tools and software components are required to provide the typical functionalities of the data lifecycle to model, create, store, secure, publish, search, retrieve, integrate, process, and visualize data originating from a NDI; in many areas, mature and integrated solutions are still lacking [1].

2.4. Privacy protection and informational self-determination

The main legal issue in the context of big data is the protection of personal data in the face of the ever-increasing capacity to aggregate even trivial and non-sensitive information about a person to an extent that very precise knowledge is gained about that person [61]. A related challenge is the possible reindividualization of previously anonymized data due to the recombination of various datasets [55,75]. With a steadily growing number of successful de-anonymization cases [65,68], the common practice to anonymize data sets before analysis is increasingly being called into question by public authorities and within research [29,37]. One possible reaction is to qualify more data as personal data in the first place, thereby impeding re-usage and potential business cases. Another way of handling the problem is to develop new methods to ensure privacy beyond anonymization. IT research currently pursues new approaches, such as the combination of decentralization and encryption: The aim is to develop solutions that enable data analysis without prior decryption of data, thus reducing the risk of (unintentional) data exposure by data re-users [33,46].

Recent developments in data analytics have also led to calls to make data protection a basic requirement regarding the design of technical solutions. Thus, the "*privacy by design and by default*" principle has been included in the European General Data Protection Regulation [24], which enters into force in 2018. The new regulation also foresees a new right to data portability according to which "*the data subject shall have the right to receive the personal data concerning him or her*" (Art. 20). The "democratization" of personal data is also subject of current initiatives under the label of "Midata" or "MyData" that aim at establishing platforms allowing individuals to manage and share their personal data [31,64,67]. They represent an attempt to re-establish the power-balance between data subjects (the individuals whose data are processed) and data controllers (in charge of the collection, storage, processing, use and dissemination of data) in a situation brought about by technological developments where there is virtually no limit to the amount of information that can be recorded, where there is hardly any limit to the scope of analysis that can be done, and where the information may be stored virtually forever [62]. In this context, the notion of informational self-determination plays an important role. The term was first used in the context of a German constitutional ruling relating to personal information collected during the 1983 census (ibid.). The court ruled that individuals had a basic right to determine in principle the disclosure and use of their personal data. The notion of informational self-determination is not equivalent to the right to privacy in the United States tradition, which is sometimes defined as "the right to be let alone", following Warren and Brandeis' seminal article [72]. Informational self-determination rather reflects A. Westin's description of "informational privacy" as the "claim of individuals, groups or institutions to determine for themselves when, how, and to what extent information about them is communicated through

others" [36,62,76]. Some scholars argue that the right to informational self-determination confers individuals an "intangible property right" over their personal data, and that individuals could therefore legally 'sell' their personal information on the market. Others argue however that in such an interpretation the "value of data protection as an instrument aimed at fostering the autonomic capabilities of individuals would be lost" and that personal information therefore should not be something individuals "may dispose of or trade on the emerging market of personal information", because the good functioning of a free, democratic society is at stake [62, p. 52]. The issue thus has ethical implications that cannot simply be resolved by installing a market place for personal data.

2.5. Data governance at the societal level

From a cross-organizational perspective, governance can be defined as *"the means by which to infuse order in a relation where potential conflict threatens to undo or upset opportunities to realize mutual gains"* [72]. Thus, governance refers to the rules, roles and behavioral patterns which configure the way collective action occurs. The collective action problems identified in political science and institutional economics can roughly be attributed to four main types: free riding/opportunism, blockage, congestion, and depletion [19]. From an economic point of view, collective action problems tend to lead to underproduction or to excessive consumption of a good or a resource, putting at stake overall utility. Collective action problems vary in function of their context (e.g. available technologies) and of the characteristics of the goods or services concerned [27,56]. All these factors have an influence on the effectiveness of governance structures in a given setting and need to be taken into account when deciding about public or private provision and on the access regimes of the various elements of a NDI. Important aspects to be taken into account in this context are the immaterial character of data, exempting it from the problem of depletion, the gradual shift away from hierarchical towards network forms of organization, and the rise of the (digital) commons in the Internet era [4,59].

While company-wide data governance frameworks have been a topic in information science for decades [74], the big data era confronts us with the same task, but worldwide and in a setting where power structures are less regulated. Data governance as overarching perspective therefore needs to focus on the key elements that are relevant for a cross-boundary common view of the reality, and the underlying architecture models need to support a vision that satisfies all relevant stakeholders [60]. Government agencies taking a leading role are expected to hold up a form of governance that has been termed "collaborative governance" or "network governance" [17,18] and has been broadly defined as *"the processes and structures of public decision making and management that engage people constructively across the boundaries of public agencies, levels of government, and/or the public, private and civic spheres in order to carry out a public purpose that could not otherwise be accomplished"* [18]. There is no doubt that seizing the benefits from data driven innovation requires the willingness to collaborate. In order to create public and economic value, incentive systems for cooperation and collective action need to be developed, covering the entire data life-cycle [3,11,32,47,50]. Coherent policies are needed to encourage investments, promote sharing and reuse, and reduce barriers to cross-border flows that could interrupt global data value chains [41,42]. The elements to be considered include data access and reuse, portability and interoperability, linkage and integration, quality and curation, ownership and control, motivation and responsibility, as well as data value and pricing [54]. Thereby it is important to strike the right balance between the social benefits of enhanced reuse and sharing of data, and individuals' and organizations' concerns about such openness [38,54].

Experiences show that non-technological design rationalities are as important as technological issues [3,44]. The roadmap for a NDI in Denmark for example is built on fundamental choices regarding

governance models (how to keep all stakeholders inside and outside government committed and on board), financing models (mitigating the fact that benefits of open data rarely accrue with the data holders making the effort) and active outreach to potential data (re-)users (recognizing them as important stakeholders), and lets those three aspects mutually reinforce each other [41].

3. The present situation in Switzerland

In Switzerland, a series of steps have already been taken in view of the establishment of a NDI [45]:

– Since 2003 a cooperative effort has been undertaken between the Confederation, the cantons and further public and private stakeholders to create a National Spatial Data Infrastructure.[1]
– In 2013, the Confederation launched a pilot portal for open government data, and the year after, the Federal Council adopted an Open Government Data Strategy. The open government data strategy at the federal level has been complemented by similar strategies at cantonal and municipal levels – however still by far not by all the cantons or cities.
– In 2015, the Rectors' Conference of Swiss Universities launched a pilot Open Research Data Platform,[2] while the State Secretariat for Economic Affairs (SECO) launched the LINDAS linked data service[3] permitting structured data from various sources to be integrated and networked using a cross-organisational approach, and commissioned an observatory for Digital Single Market infrastructures.
– In 2016, the Confederation launched the official opendata.swiss portal[4] for government data and data of state-affiliated businesses throughout Switzerland [75]. It also published its "Digital Switzerland" Strategy [25] which, among other things, calls for a *"coherent and future oriented data policy in Switzerland"*, including *"regulatory coordination on a national, and where necessary international basis, in relation to access to data and digital content as well as data processing and security"*, for the *"construction of a national data infrastructure"* to help the *"open data principle permeate the entire administration, the research sector and parts of the private sector"*, and for the implementation of the *"fundamental right of an individual to exercise self-determination with regard to information"* by giving Swiss citizens control over their own data.

Besides that, there is a range of data infrastructures where practices of data sharing among different administrations occur, but an overarching view or strategic vision is still lacking.

4. Research questions and methodological approach

In spring 2016, the Bern University of Applied Sciences, in cooperation with the opendata.ch association, carried out a study to explore how key stakeholders in Switzerland conceive of a NDI and the governance issues surrounding it. The two research questions at the center of the study were:

1. What are the key building blocks of a national data infrastructure?
2. How should inter-organizational collaboration and participation be organized in view of the establishment and the maintenance of a national data infrastructure?

---

[1] http://www.e-geo.ch.
[2] http://openresearchdata.ch.
[3] http://lindas-data.ch.
[4] https://opendata.swiss.

Table 1
Overview of interview partners

No.	Organization(s)	Function(s) of the Interviewee(s)	Sector(s)
1	National Council; Private sector company providing ICT and Internet services	Member of Parliament; Member of the Administrative Board	Politics; Private sector
2	National Council	Member of Parliament	Politics
3	Federal Department of Economic Affairs, Education, and Research	Secretary General	Public administration
4	State Secretariat for Economic Affairs	Head of E-Government for SME	Public administration
5	Federal Office of Communications	Director	Public administration
6	Swiss Federal Archives	Director	Public administration
7	Federal Chancellery	Federal Chancellor	Public administration
8	Swiss Federal Railways	Enterprise Architect	State-affiliated enterprises
9	Swiss Post	Head of the CIO office, responsible for ICT processes and service management; Head of Corporate Strategy, Data Management and Governance; Information Architect responsible for big data	State-affiliated enterprises
10	Swisscom	Head of Big Data Business Development Senior Manager for Public Communication	State-affiliated enterprises
11	Insurance company	Head natural hazards; Co-director of a company-sponsored university lab for natural hazards	Private sector; Universities
12	Insurance company	Head IT Supply Chain Management	Private sector
13	Banking company	Chief Information Officer	Private sector
14	Opendata.ch association	President	Civil society
15	Swissuniversities	Co-directors of the program SUC-P2 "Scientific information: access, processing and safeguarding"	Universities
16	University of Applied Sciences; eCH association	Head of the E-Government Institute; Board member	Universities; Standardization organization

In order to address these questions, 16 semi-structured interviews were carried out with key stakeholders from various backgrounds, representing the perspectives of public administration, state-affiliated enterprises, private enterprises, research, civil society, and politics (see Table 1 for an overview of interview partners). The interview guideline comprised questions on four over-arching themes:

– Data-based value creation: Where do interviewees situate themselves in the value chain with regard to the secondary use of data? How do they experience the interactions between agents along the value chain? What works well? What does not? What measures should be taken to improve the cooperation between different players? How to improve the value of the data in view of their secondary use?

– Building blocks and stakeholders of a NDI: To what extent do interviewees consider data as an infrastructure resource? What are the building blocks of a NDI? What is its purpose? What use would the interviewees' organizations derive from a NDI? What conditions need to be satisfied for their organizations to benefit from a NDI? Which building blocks of a NDI should be put in place first? Which data should be included? What other aspects need to be tackled with the highest priority? What contribution could the interviewees' organizations make? Which contributions would they expect from other entities?

– Governance issues in the context of a NDI: What governance aspects need to be tackled? How to improve cross-organizational cooperation between the different players? How to promote data-

based innovation? How best to achieve informational self-determination?
– Next steps in view of the establishment of a NDI: Which stakeholders should be involved in the process? To what extent would interviewees' organizations be willing to take an active part in the establishment and maintenance of a NDI? In what areas do they expect research to provide the necessary know-how for the establishment of a NDI?

The interviews were exploratory in nature and aimed at understanding different stakeholder perspectives. In terms of Emerson's framework of collaborative governance [18], the focus of the research was on "discovery" (revealing of individual and shared interests, concerns, and values) and "definition" (building shared meaning by articulating common purpose and objectives; agreeing on concepts and terminology; clarifying and adjusting tasks and expectations among stakeholders). The interview partners were selected from a list of potential interlocutors based on a stakeholder analysis according to the typology proposed by Mitchell et al. [48], favoring stakeholders that are characterized by high levels of power, legitimacy and urgency with regard to the establishment of a NDI, as this would at the same time allow us to identify the core stakeholders that are needed to further advance the establishment of a NDI in Switzerland. Furthermore, we made sure that the selection equally comprised data providers, data users, intermediaries, and other players (e.g. politicians).

5. Results of the interviews

5.1. Assessment of the status quo

In order to set the grounds for discussing issues related to the establishment of a NDI in Switzerland, the interviewees were asked to position their organization/themselves along the data value-chain and to assess the current situation regarding secondary data use. As was to be expected due to the selection process, many interviewees indicated that their organizations fulfill multiple roles as data providers, data users and as intermediaries. For some of them, one of these roles clearly prevails, while for others the roles are context-dependent. And finally, the sample included three interviewees who are neither data providers nor data users, nor typical intermediaries, but are mainly involved in setting the framework conditions (e.g. through lobbying or as members of parliament).

When asked to assess the current interplay between stakeholders of the data value chain, interviewees put positive emphasis on the following aspects: Governmental data provision in selected areas (geospatial data, statistical data, and environmental data) is appreciated, in particular if the data is made available through central portals in a coordinated manner (international, federal). Several stakeholders, especially from the private sector, have established processes in place for re-using data directly from third parties (government, others) that generally work well. Interviewees also pointed to the favorable conditions in Switzerland regarding the basic communication infrastructure and legal certainty as prerequisites for exploiting data value chains. The ongoing political dialogue on data-related issues is considered as important and the role of the open data community as catalyst for raising political awareness is appreciated in that respect.

Stakeholders also pointed out some negative aspects of the current situation: Governmental data provision shows deficiencies with regard to data accessibility due to heterogeneous data formats and pricing models as well as with regard to data interoperability across organizational, federal and national boundaries, resulting in unnecessary effort and expenditures for data users. For government data providers, it is not always clear who is to fund the activities necessary to change this situation. Interviewees also

Differentiating characteristics of data				
Provenance	State	State-affiliated enterprises	Private enterprises	Private individuals
Production context	Public		Enterprise	Personal
Accessibility	Open		Shared	Closed
Data protection requirements	Particularly sensitive		Non-sensitive	
Purpose of use	Functioning of the state		Other purposes	
Scope	Possibly all data		Demand-oriented data	

Fig. 1. Morphological box with the criteria used to differentiate the data potentially within the scope of a NDI.

pointed to a lack of knowledge about existing data value chains (who uses which data for what purpose?). Related to this, there are still observable reservations towards opening up data on the one hand and uncertainties concerning data access on the other.

Accordingly, interviewees identified room for improvement. There is a broad understanding that the full potential of the secondary usage of data has not been tapped yet and that associated opportunities and risks have not been fully understood, neither in the public nor in the private sector. The adoption of an ecosystem perspective is considered essential to enhance value generation through data re-usage (e.g. through data combination). Also, several interviewees perceive a need for a broader social debate and for raising public awareness, especially on data privacy issues and data ethics in general (e.g. informational self-determination). According to them, politics should develop a coherent position on the provisioning of open government data (including the clarification of financial aspects), as clear conditions and rules would be favorable both to data provision and usage.

5.2. Understanding of a national data infrastructure

In order to reach a better understanding of the way various stakeholders conceive of a "national data infrastructure", interviewees were asked to what extent data should be considered an "infrastructure resource" (cf. Section 2.1). While it was undisputed that data are a highly valuable resource that is prone to be re-used by third parties, identifying them as an "infrastructure resource", similar to the electricity grid, the water system, roads, railroads, etc. was not straightforward for all interviewees.

When it came to deciding which data should belong to a NDI, various interviewees applied different criteria (see the morphological box in Fig. 1 for an overview), which had an impact on their view of the necessary governance mechanisms and the question to what extent data provision is to be considered a public responsibility: While some reckoned that a NDI should only include data held by public sector organizations, others considered that it should be extended to include also data held by private enterprises and by individuals. For some it was not so much a question of who held the data, but by whom the data was produced or provided. While some expressed the view that a NDI should only include data that can be provided as open data, others stressed the importance of allowing data providers to share data among selected users, or perceived of the NDI as a system which may also include data that remains closed (e.g. at the discretion of individuals exercising informational self-determination). Interviewees who considered the NDI first of all as an infrastructure for the provision of open data were mainly talking about data that is not sensitive in terms of data protection requirements, while those approaching it also from a mydata perspective had also sensitive data in mind. Yet another criterion was applied by some of the interviewees who argued that a NDI should comprise the data that is necessary for the proper functioning of the state, and finally there were diverging views on whether all eligible data should be provided through a NDI by default or whether data provision should be demand driven.

Participants were also invited to point out important components of a NDI. Again, the notion of a NDI was not equally familiar to all interviewees. While some had a rather clear vision of what a NDI is, others would have preferred to be presented with a pre-defined concept. Some interviewees adopted an overarching perspective; others rather put a focus on specific areas of interest. In sum, the interviews provided broad input with regard to a common understanding of a future NDI that converges on the following definition:

> *A national data infrastructure is a nationwide distributed technical infrastructure (comprising portals, platforms, and services) that allows the access to and exchange of data based on predefined rules.*

Thereby it is important to stress that a national data infrastructure is not a monolithic block (but has a distributed architecture) and does not exist in isolation (but is inter-connected at the international level). These two aspects were particularly emphasized by the interlocutors from the research sector, where communities tend to organize across borders, but were also mentioned by several other interviewees.

Participants also provided concrete input on what a NDI should or could comprise and what issues its implementation would entail. As for the different types of data, there is a common understanding that government data are at the core. Correspondingly, provision is at least in part considered to be of public responsibility. However, the question which other types of data should be made available via a NDI and what responsibilities arise from this requires further clarification. Further potential elements of a NDI that need to be addressed in a coordinated manner are presented below (cf. Section 5.4).

Most interviewees would agree that the purpose of a NDI is to support data-driven value creation and to help realize the potential value of existing data. The interviewees were asked about the benefits they would gain from a future NDI. The following potential benefits were most salient:

- Efficiency gains (access to more data, standardized data, inter-organizational exchange of data and shared infrastructures);
- Better services for users (new or improved services thanks to new insights or better integration of information across organizational boundaries);
- An improved image of government agencies and state-affiliated enterprises;
- Reduced costs related to data acquisition.

For several interviewees, however, the concrete purpose and use of a NDI still needs clarification, particularly from the point of view of private enterprises. Some suggested to develop the business case for a NDI and to analyze its economic and societal benefits in more depth, e.g. based on use cases.

5.3. Key stakeholders and their roles

In order to explore the perceptions different stakeholders have regarding the key agents and their roles in setting up a NDI, the interviewees were asked what contribution their organization could make to establishing a national data infrastructure in Switzerland and what they would expect from other stakeholders in that respect. Again, answers ranged from clearly formulated expectations to rather reluctant statements. The results underline that the implementation of a NDI requires the participation of various stakeholders with well-defined roles, some of which are already rather clear, while others remain to be clarified.

There were no notable gaps regarding self-perception and external perception of the various stakeholders in this regard. Overall, the following key roles emerged: the interviewees expect from politics to provide the necessary framework conditions and to issue mandates to public administration. Public

Table 2
Fields of action requiring coordination

Dimensions	Fields of action requiring coordination
Ethical	Ethical norms regarding the use of personal data
Legal	Regulation of data protection and security Regulation of the usage of government data (e.g. fees, conditions limiting its usage) Regulation of data provision (attribution of tasks, regulation of liability issues) Regulation of the use of personal data
Political	Legal foundations/legal mandate (clarification of government agencies' responsibilities) Coordination (across sectors and federal levels)
Economical	Funding for data provision/basic investment Business models for data refinement Competitive situation of state-owned enterprises Data monetization (as an incentive for data provision)
Organizational	Access regimes (open vs. club model, private data) Data-lifecycle management Processes (data provision, data usage, etc.) Knowledge management, sharing of know-how, support to data providers and users
Semantical	Shared ontologies Metadata
Technical	Basic infrastructure (for the storage, transport, and processing of data) Infrastructure components (data portals, platforms and interfaces for service provision) Identity and access management Knowledge management functionalities

administration in turn shall provide and publish data, foster the debate, play a coordinating role, and contribute to the setup of the technical infrastructure. Civil society organizations are expected to foster the networking and the dialogue between stakeholders, while academia is expected to provide data and infrastructure components, whereas the exact role of state-affiliated and private enterprises still needs clarification.

5.4. Fields of action requiring coordination

As mentioned in Section 3, Switzerland does not start from zero when it comes to establishing a NDI, and as the new "Digital" Switzerland Strategy as well as ongoing debates about data policy illustrate, time is ripe to advance its conceptualization and implementation. As the interviews indicated, the creation of a NDI should be seen as a step-by-step endeavor that combines a bottom-up and a top-down approach: While publishing first datasets, involving citizens and private companies, and developing use cases can be done by many government agencies in a bottom-up approach, several interviewees argued that the strategic framework should be provided in a top-down manner, ensuring that it is backed by political decisions providing a clear mission to the government agencies involved in the establishment of a NDI. To be effective, the two approaches need to be synchronized; several interviewees pointed to the fact that civil society ought to play the role of a mediator between the two.

Based on the stakeholder interviews, seven fields of action could be identified that need to be tackled in a coordinated manner. The most important issues raised in the interviews are presented in Table 2; they have been categorized according to an extended version of the European Interoperability Framework's interoperability levels (EIF) [20]:

– On the ethical dimension, it has been pointed out that given recent technological developments, a broad public debate is needed regarding the ethical norms that should govern the use of personal data and the adaptation of the legal framework in this area.

- On the legal dimension, interviewees have called for an adaptation of the legal framework in the areas of data protection and data security to take into account the recent developments in the area of big data analytics and to empower citizens to decide how their personal data is to be used. Adaptations are also needed in areas where current ordinances prohibit government agencies from releasing data as open data due to fees regulations or other usage restrictions. Furthermore, it has been pointed out that clarification is needed as to which type of actors are responsible for what aspects of data provision and to what extent they could be made liable for shortcomings in the data.
- On the political dimension, coordinative action is required not only in terms of providing the legal foundations and the legal mandate to public administration, but also to ensure coordination across sectors and federal levels.
- On the economic dimension, the main concern is about financing models for basic investments as well as the development of viable business models for data provision and data refinement. There is also a certain level of uncertainty on how the public and the private sectors will cooperate in this area. While large economic benefits have been predicted for example in the field of open government data, it remains unclear who exactly will be able to capture these benefits and to what extent the entities making the effort to provide the data will be able to recoup on their expenses. With regard to data sharing in the private sector it is still rather unclear to the interviewees what incentives will be put in place. Here, some reflections go in the direction of data monetization, which leaves state-owned enterprises wondering to what extent they should embark on the open government data journey, and to what extent they should play by the rules of the private sector.
- On the organizational dimension, it must be clarified which data should be governed by which access regimes (open access vs. club model vs. private data exchanged on a data market). Another aspect is the organization of data-lifecycle management which involves a series of players along the data value-chain. In this context, standard processes and best practices should be established. Several interviewees also pointed to the importance of knowledge management, the sharing of know-how and the provision of support to data providers and users.
- On the semantical dimension, interviewees mentioned the need for a standardized data model for data catalogues (metadata of data sets) as well as the development of shared ontologies in various fields to ensure semantic interoperability of the data across organizational boundaries.
- On the technical dimension, basic infrastructure is required for the storage, transport, and processing of data, complemented by a set of infrastructure components (data portals, platforms and interfaces for service provision). The data needs to be secured, and appropriate identity and access management needs to be put in place which works across organizational borders. The various dimensions need to be supported by knowledge management systems that facilitate collaboration in inter-organizational settings.

Many interviewees expressed their views on how coordination should take place. Thus, it has been observed that the dialogue and the establishment of a 'code of conduct' among the different stakeholder groups are at least as important as legal regulations. There is no agreement as to the right mix between incentives and legal obligations when it comes to getting government agencies to publish their data in form of open data. Several interviewees expressed the view that public agencies need to go ahead with regard to opening up their data, setting an example in order to encourage other players (e.g. private enterprises) to follow suit. Others stressed the importance of successful use cases, as only demonstrable usefulness of providing data via a national infrastructure (as open or shared data) could create the momentum necessary to trigger a self-carrying dynamism. And finally, in view of ensuring governance in a collaborative context, it was suggested to elaborate several possible variants regarding the overall objective before deciding on a common roadmap.

Table 3
Fields of interest related to NDIs

Focus	Fields of interest
Ecosystem	Multiple perspectives on the ecosystem (governance, legal, organizational and technical issues)
	International best practices and implications for Switzerland
	Approaches to defining responsibilities in a multi-stakeholder context
	Assessment of existing cooperation models (public/private)
	Assessment of existing models for data brokerage
	Ways of ensuring the public value of a NDI (e.g. alignment between providers and users of data)
Economic aspects	Analysis/provision of evidence on the benefits of a NDI
	Assessment and prioritization of potentials for value creation
	Overview and assessment of business models (along the range of types of data and access regimes: from personal to public, and from closed to open)
	Approaches to balancing investments and benefits in a multi-stakeholder context (public/private)
	Assessment of adequate pricing models for data re-use
	Assessment of models for remunerating re-use of personal data
	Assessment of suitable (cooperation) models for operating data infrastructures (central/decentral etc.)
Ethical aspects	Permissibility/legitimacy of linking disparate data (especially with regard to personal data)
	(Non-)legitimacy of analysis/validity of conclusions
	(Limits of) acceptance within society (big data, privacy)
	Definition of adequate basic conditions for the analysis of personal data and ways to enforce them
Privacy	Overview of the range of personal data traces and (technical) possibilities for generating insights
	Overview of recent developments in the field of privacy at an international level
	Overview and assessment of mechanisms aimed at securing control over personal data
	Overview and assessment of technological possibilities required to install trusted third parties as intermediaries regarding the provision of/access to personal data
Technology	Overview of the technological possibilities and challenges of a NDI
	Clarification of differences between network infrastructures and data infrastructures
	Assessment of options for organizing data management, e.g. the (dis-)advantages of centralization
	Assessment of different design options and their (dis-)advantages
	Overview and assessment of approaches to increase data quality in complex domains (sectorial data)
	Requirements towards the technical infrastructure regarding the management of big volumes of data
	Assessment of the potential of blockchains for managing/regulating data access (from a technical and legal perspective)

5.5. Required knowledge base

The interviews were also used as an opportunity to get better insights on the knowledge base needed by key stakeholders to engage in conceptualizing a NDI and in order to identify potential research gaps. Needs of information touch upon various fields of interest (cf. Table 3). Participants stated that further considerations on establishing a NDI should be based on an ecosystem perspective, and that the cost-benefit ratio of a NDI should be assessed in more detail, e.g. based on use cases. Also, there was a strong interest in issues related to data ethics and promising approaches to strengthen privacy protection.

6. Discussion

6.1. Four perspectives on a national data infrastructure

As the review of the existing literature and the interview results suggest, the notion of a NDI should be approached from four distinct, but complementary perspectives (see Fig. 2):

– The big data perspective stresses the role of novel analytical methods, spurred by an ever-increasing quantity of data collected and facilitated by ever larger data storage capacities and a trend towards
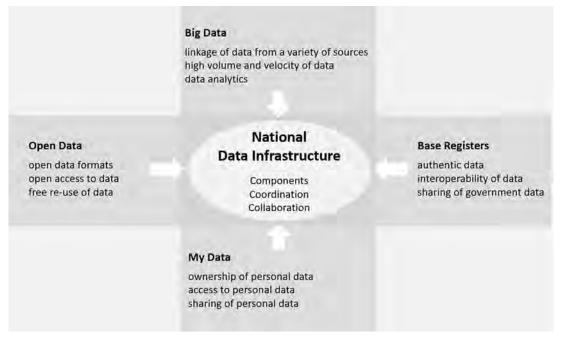
Fig. 2. Four perspectives on a NDI.

breaking up data silos. Data analysis requires access to all sorts of data (authentic, open, personal). At the same time, big data analytics challenges traditional approaches to data protection.

– The base register perspective focuses on data that is collected on the basis of a specific legal mandate and that is trusted, authentic, and under access control [20]. Infrastructures for basic government data have long been in place, are shared among authorized parties, but are in most cases not open to third parties.

– The open (government) data perspective highlights the quality of 'openness' of the data on a technical and legal scale. In recent years, national, local, and sectoral infrastructures (e.g. spatial data) have been built to provide open government data. Evidence on the impact is still limited.

– The mydata perspective stresses the re-use of personal data in combination with an increased control of individuals over their own personal data [57]. This implies a paradigm shift in the area of data protection, requires a special focus on societal acceptability, and calls for new technological solutions.

The four perspectives enjoy different levels of prominence in the present debate about a NDI in Switzerland: Open data has been the object of a federal strategy adopted in 2014, and similar strategies have been adopted in some cantons and larger cities; the topic has been promoted by a community-based lobby-organization since 2010 and has received further impetus through the launch of a national open data portal in 2013. Over the past years, increasing efforts have been made by the federal administration to provide the data also in form of linked data. Big data has been the object of a study commissioned by the Federal Office of Communications in 2015 which evaluated the opportunities and risks brought about by the recent developments in this area and the resulting needs for action. The study recommended measures along the following lines: (1) establishment and maintenance of a national data infrastructure based on the open data principle; (2) measures to reduce the risk of abuses; and (3) implementation of the principle of informational self-determination [40]. Mydata has been promoted in Switzerland mainly

by the MIDATA cooperative, which was founded in 2015. The organization lobbies in favor of inscribing the principle of informational self-determination in the national legislation and is in the process of developing a platform that allows people to selectively share their health data. While the sharing of base registers has taken place within the federal administration and in cantonal administrations for years, the sharing has typically been limited to selected public agencies, and it is only in the context of recent efforts in the area of open government data and in the context of discussions about the establishment of a NDI, that the idea of a more generalized approach to sharing base register data has surfaced.

On an abstract level, each perspective shows a number of peculiarities with regard to types of data, existing infrastructure components as well as prevailing and/or emerging technologies. Also, in accordance with its target stakeholders and its focus, each perspective will bear specific policy implications.

Reaching agreement over the right mix of the four perspectives and the pace at which each of them shall be developed will be crucial when it comes to advancing the cause of a NDI. For this purpose it will be useful to reflect on the interactions and inter-relations between the different perspectives. Thus, adopting the base register perspective may re-focus the efforts in the field of open government data on areas that are particularly likely to entail efficiency gains both within public administration and among private enterprises. As the Danish example shows [41], this may indeed unlock the discussion about the actual (economic) use of open data. Furthermore, it may spark reflections about the pooling of data as 'shared data' among various public authorities, possibly even with the participation of private enterprises. At the same time, it can be expected to spur the discussion about the re-use of personal data, which is also at the center of the mydata perspective. Here, however, the preoccupation is not primarily with efficiency gains, but with ethical questions and with ensuring informational self-determination. In fact, the informational self-determination and the empowerment of users endorsed by the mydata perspective is two-pronged: On one hand, it is about civil liberties, about the right to know and to regain control over who is doing what with one's personal data. On the other hand, it is about economic rights, namely the right to put one's own personal data to profit – be it by releasing it for re-use for a social cause or by trying to monetize its use by profitable enterprises in the context of big data applications. And finally, this two-pronged nature of the mydata perspective shares similarities with the open data perspective's two-fold objective of ensuring transparency and good governance and of unleashing the slumbering (economic) potential related to the re-use of existing data within public administration and beyond thanks to new developments in data analytics.

6.2. Need for collaborative governance

As expected, the provision of the technical components of the NDI represent only a fraction of the aspects that require coordination. While for many aspects, public administration and research institutions are expected to act as the main drivers in view of the establishment and the management of a NDI, the interviews have also made clear that a NDI is the result of a collaborative effort across sectors and across federal levels. Furthermore, its establishment needs to be coordinated with similar initiatives in other countries. As a result, the government agencies taking a leading role are expected to engage in "collaborative governance". Concretely, government agencies are supposed to take the initiative by publishing data, involving citizens and companies, and by developing use cases. First movers are expected to lead by example, and several fields of action requiring coordination are characterized by a network mode of governance, such as cross-organizational data lifecycle management, the agreement on standard processes for data provision and data usage, the development of metadata standards and shared ontologies, or inter-organizational learning and knowledge management. In the same vein, some interviewees

stressed the importance of generally recognized codes of conduct as opposed to legal regulations, which is characteristic of the network mode of governance [58]. Several public sector representatives stressed nonetheless the importance of the hierarchical relationship their activities are inscribed in; hence the emphasis on clarifying government agencies' responsibilities by providing them with an explicit mandate from parliament. While some instances of the market mode of governance are seen as impediments, such as fees regulations hampering the release of government data as open data or the competitive situation of state-owned enterprises preventing them from fully embracing an open data policy, market mechanisms are also cited as incentives for data publication or data refinement and could therefore play a vital role in creating the basis for a flourishing data infrastructure. Furthermore, it should be kept in mind that the network mode of governance and cross-organizational collaboration also have their downsides: While the barriers to data sharing among governance agencies may be perceived as a major impediment when it comes to realizing the benefits of big data in the public sector, effective privacy regulations may in fact require informational separation of power with the effect that any transfer of personal data from one state entity to another must be handled as an exception covered by a legal provision [36]. Similarly, in the private sector, collaborative arrangements among competitors may serve as a seedbed for anticompetitive practices and therefore be detrimental from a macroeconomic point of view [58]. As these examples illustrate, changes to existing governance structures need to be approached with caution. Besides that, there are several other framework conditions of a NDI that require thoughtful deliberation and broad societal debate: the ethical norms and rules governing the use of personal data, the regulation of data security as well as the regulation of the liability of data publishers and other actors along the data value-chain.

6.3. Informational self-determination as a salient aspect of a NDI

As noted above, the mydata perspective, i.e. the issue of informational self-determination, has emerged as a salient aspect of a NDI, which in this form had not been mentioned in the existing literature on national data infrastructures. The issue is highly topical and the object of an unfolding political debate, where many stakeholders have not adopted a clear position yet. As a recent study on big data in Switzerland shows [40], there is broad support for a new paradigm in the area of personal data, which consists in giving individuals greater control over the data that pertains to them. Giving individuals a 'right to a copy' of the personal data held by companies and organizations would be a first step towards this new paradigm. The issue has been debated in the Swiss parliament and is currently subject of a study mandated by the federal government. A similar development can be observed within the EU, where the European General Data Protection Regulation [24] implements a new right to data portability according to which "the data subject shall have the right to receive the personal data concerning him or her". A further step promoted by some would consist in giving individuals some form of property right over the personal data concerning them, empowering them to exercise greater control over what use third parties are allowed to make of their data. In view of the implementation of a NDI the following questions should be asked:

– What role do personal data play in the context of a NDI?
– What approach should be taken towards data anonymization in the face of a growing risk of de-anonymization?
– To what extent could a NDI serve as a catalyzer for user empowerment through the provision of platforms for the management of personal data and the creation of trustworthy intermediaries between holders and potential re-users of personal data?

- What has a NDI in store for private individuals (e.g. in terms of a 'right to a copy' or in terms of monetization of one's personal data)?
- How to balance the different interests of (1) data-collecting organizations, (2) the public (open data, transparency; maintaining a free, democratic society), and (3) individuals (mydata, informational self-determination)?

6.4. Implications for further research activities

As has been pointed out by previous research and has again been underlined by the present study, the establishment of a NDI clearly must be tackled from an interdisciplinary perspective. The research in the diverse fields of interest from the point of view of key stakeholders is not equally developed. While some areas are fairly well studied (e.g. pricing models) other research fields are only emerging (e.g. systematic analysis of the impact of open data, new privacy mechanisms). Overall, the interviews suggest that there is a need to strengthen knowledge transfer from research to practice and vice versa. Furthermore, there are opportunities to learn from the experiences in other countries; in the case of Switzerland this is particularly the case with regard to some countries' focus on base registers that has not received much attention so far by Swiss authorities.

7. Conclusions and outlook

7.1. What are the key building blocks of a national data infrastructure?

The findings regarding our first research question can be summarized as follows:

A national data infrastructure can be conceived of as a nationwide distributed technical infrastructure (comprising portals, platforms, and services) that allows the access to and exchange of data on the basis of predefined rules. It is not a monolithic block, but has a distributed architecture, and it is not confined to one country alone, but is inter-connected at the international level. Also, the notion of a NDI should be approached from four distinct, but inter-related perspectives: a big data perspective, a base register perspective, an open data perspective, and a mydata perspective. Each of the four perspectives shows a number of peculiarities with regard to types of data, existing infrastructure components as well as prevailing and/or emerging technologies. Also, in accordance with its target stakeholders and its focus, each perspective will bear specific policy implications.

With regard to the data a NDI shall comprise, there is general agreement that non-sensitive government data and research data shall be part of it. With regard to other types of data, such as data from private enterprises or private individuals, further clarification is needed.

7.2. How should inter-organizational collaboration and participation be organized in view of the establishment and the maintenance of a national data infrastructure?

There is general agreement that a step-by-step implementation process combining bottom-up and top-down elements is most likely to lead to success. There is also general agreement as to the role of different stakeholder types: the parliament is expected to provide the necessary framework conditions and to issue mandates to public administration. Public administration in turn shall provide and publish data, foster the debate, play a coordinating role, and contribute to the setup of the technical infrastructure.

Civil society organizations are expected to foster the networking and the dialogue between stakeholders, while academia is expected to provide data and infrastructure components and to feed current research developments back into practice. At the same time, it also appeared from the interviews that the role of state-affiliated and private enterprises still needs to be clarified.

The study has also allowed to identify a set of concrete issues on which coordinating action is needed. It has been shown that such coordinating action is needed along seven dimensions: political, ethical, legal, economical, organizational, semantic, and technical. Given the prominence of the issues related to open data, personal data, and informational self-determination, the central role of data governance should not be underestimated when it comes to balancing the different interests of data-collecting organizations, the public (open data; maintaining a free, democratic society), and individuals (mydata). Reaching a good balance in this respect is one of the keys to the success of a NDI. Another one is the demonstration of its usefulness through the example of concrete use cases.

7.3. Next steps

Our research project is contributing to the discussion of a highly topical issue in Switzerland: the blueprint of a national data policy. The focus is on key stakeholders within the big data ecosystem in their diverse roles as data providers, data users and analyzers, legislators and regulators, as well as providers of public goods and guarantors of security. With view to the economic expectations and societal concerns associated with big data, developing a concept of a NDI serves first of all to outline the frame conditions within which development can be driven forward. This includes the clarification of roles of the different stakeholders, the definition of rules and decision rights, the agreement on government frameworks, the implementation of appropriate accountability mechanisms, as well as the identification of topics that need broader public debate.

In close collaboration with key stakeholders we will further develop a common vision of a NDI and identify the need of actions from a governance perspective. Furthermore, as one of the next steps, the concrete infrastructure components (data portals, platforms, key services) need to be defined and the overall technical architecture of the NDI designed. As our research has confirmed, a multidisciplinary approach is crucial. While there is plenty of existing research to be drawn upon in the various domains, the real challenge lies in painting a holistic picture, taking into account the interactions and the interdependencies between the various dimensions and perspectives. The research presented in this paper has allowed us to stake out the ground and to develop an analytical framework that can be used as a boundary object in further interactions with and among stakeholders.

References

[1] Auer, S., Bühmann, L., Dirschl, C., Erling, O., Hausenblas, M., Isele, R., Lehmann, J., Martin, M., Mendes, P.N., van Nuffelen, B., Stadler, C., Tramp, S., & Williams, H. (2012). Managing the life-cycle of linked data with the LOD2 stack. International Semantic Web Conference, Berlin, Heidelberg: Springer.

[2] Bauer, F., & Kaltenböck, M. (2001). Linked Open Data: The Essentials: A Quick Start Guide for Decision Makers. Semantic-Web Company.

[3] Bekkers, V. (2009). Flexible information infrastructures in Dutch E-Government collaboration arrangements: Experiences and policy implications. Government Information Quarterly, 26(1), 60-68.

[4] Benkler, Y., & Nissenbaum, H. (2006). Commons-based peer production and virtue. Journal of Political Philosophy, 14(4), 394-419.

[5] Berners-Lee, T. (2009). Putting Government Data online. Available from: https://www.w3.org/DesignIssues/GovData. html.

[6] Bohn, J., Coroamă, V., Langheinrich, M., Mattern, F., & Rohs, M. (2005). Social, economic, and ethical implications of ambient intelligence and ubiquitous computing. In: Weber, W., Rabaey, J.M., & Aarts, E. (eds) (2005). Ambient Intelligence, Berlin, Heidelberg: Springer, pp. 5-26.

[7] Bundesrechenzentrum (2015). Big Data in der öffentlichen Verwaltung. White Paper. Bundesrechenzentrum GmbH. Available from: https://www.brz.gv.at/downloads/studien/2015-03-30_Big_Data_in_der_oeffentlichen_Verwaltung_ v1.2_pub.pdf?55m9h7.

[8] Cabinet Office (2015). National Information Infrastructure Implementation Document. Second Iteration. UK Cabinet Office. Available from: https://www.gov.uk/government/publications/national-information-infrastructure.

[9] Capgemini, IDC, Sogeti, & Politecnico di Milano (2015). Future-proofing eGovernment for a Digital Single Market. European Commission. https://doi.org/10.2759/32843.

[10] Chen, Y.C., & Hsieh, T.C. (2014). Big Data for Digital Government: Opportunities, Challenges, and Strategies. International Journal of Public Administration in the Digital Age, 1, 1-14.

[11] Cordella, A., & Bonina, C.M. (2012). A public Value Perspective for ICT Enabled Public Sector Reforms: A Theoretical Reflection. Government Information Quarterly, 29(4), 512-520.

[12] de Vries, M. (2012). Funding of a System of Key Registers in a PSI-conomics and Contemporary Perspective. The Dutch Experience in a Danish Context. Danish Ministry for Housing, Urban and Rural Affairs.

[13] Demartini, G., Difallah, D.E., & Cudré-Mauroux, P. (2012). ZenCrowd: Leveraging Probabilistic Reasoning and Crowd-sourcing Techniques for Large-scale Entity Linking. In: Proceedings of the 21st World Wide Web Conference, ACM, pp. 469-478.

[14] Desouza, K.C., & Jacob, B. (2014). Big Data in the Public Sector: Lessons for Practitioners and Scholars. Administration & Society, 49(7), 1043-1064.

[15] Duggan, J., Elmore, A.J., Stonebraker, M., Balazinska, M., Howe, B., Kepner, J., Madden, S., Maier, D., Mattson, T., & Zdonik, S. (2015). The BigDAWG Polystore System. ACM Sigmond Record, 44(2), 11-16.

[16] Eckert, K.P., Henckel, L., & Hoepner, P. (2014). Big Data – Ungehobene Schätze oder digitaler Alptraum. Fraunhofer FOKUS. Available from: https://www.oeffentliche-it.de/documents/10181/14412/Big+Data+ungehobene+Sch%C3% A4tze+oder+digitaler+Albtraum.

[17] Eggers, W.D. (2008). The changing nature of government: network governance. In: O'Flynn, J., & Wanna, J. (eds.) (2008). Collaborative Governance: A new era of public policy in Australia? Canberra: ANU E Press, pp. 23-28.

[18] Emerson, K., Nabatchi, T., & Balogh, S. (2012). An integrative framework for collaborative governance. Journal of Public Administration Research and Theory, 22(1): 1-29.

[19] Estermann, B., Riedl, R., & Neuroni, A.C. (2009). Integrated and Transcendent E-government: Keys for Analyzing Organizational Structure and Governance. In: Proceedings of the 10th Annual International Conference on Digital Government Research, dg.o, pp. 162-171.

[20] European Commission (2010). European Interoperability Framework (EIF) for European public services. Annex 2. Bruxelles: European Commission.

[21] European Commission (2014). Towards a Thriving Data-Driven Economy. COM(2014) 442 final. Available from: https://ec.europa.eu/digital-single-market/en/news/communication-data-driven-economy.

[22] European Commission (2015). A Digital Single Market Strategy for Europe. Brussels: European Commission, 6.5.2015.

[23] European Commission (2016). Providing Big Data Opportunities for Public Administrations [website]. Available from: http://ec.europa.eu/isa/actions/01-trusted-information-exchange/1-22action_en.htm.

[24] European Union (2016). Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of such Data, and Repealing Directive 95/46/EC, L 119.

[25] Federal Office of Communications (2016). "Digital Switzerland" Strategy. Bern: Federal Office of Communications.

[26] Fraefel, M., Selzam, T., & Riedl, R. (2013). Organizational requirements for building up national E-government infras-

tructures in federal settings. In: 46th Hawaii International Conference on System Sciences (HICSS), IEEE, pp. 1642-1651.

[27] Frischmann, B.M. (2006). An Economic Theory of Infrastructure and Commons Management. Minnesota Law Review, 89, 917-1030.

[28] Frischmann, B.M. (2012). Infrastructure: The Social Value of Shared Resources. Oxford Scholarship Online.

[29] Graham, S.L., Press, W., Gates jr, J.S., Lander, E.S., Gorenberg, M., Mundie, C., & Schmidt, E. (2014). Big Data and Privacy: A Technological Perspective. President's Council of Advisors on Science and Technology.

[30] Gschwend, A., Neuroni, A.C., Gehrig, T., & Combetto, M. (2015). Publication and Reuse of Linked Data: The Fusepool Publish-Process-Perform Platform for Linked Data. In: Tambouris, E. et al. (eds.) (2015). Electronic Government and Electronic Participation. Amsterdam: IOS Press, pp. 116-123. https://doi.org/10.3233/978-1-61499-570-8-116.

[31] Hafen, E., Kossmann, D., & Brand, A. (2014). Health data cooperatives – citizen empowerment. Methods of Information Medicine, 53(2), 82-86.

[32] Harrison, T.M., Guerrero, S., Burke, G.B., Cook, M., Cresswell, A., & Helbig, N. (2012). Open Government and E-Government: Democratic Challenges from a Public Value Perspective. Information Polity, 17(2), 83-97.

[33] Hayes, B. (2012). Alice and Bob in Cipherspace. American Scientist, 100(5), 362-367.

[34] Hitzler, P., Krotzsch, M., & Rudolph, S. (2010). Foundations of semantic Web technologies. Boca Raton: CRC Press.

[35] Höchtl, J., Parycek, P., & Schöllhammer, R. (2016). Big data in the policy cycle: Policy decision making in the digital era. Journal of Organizational Computing and Electronic Commerce, 26(1-2), 147-169.

[36] Hornung, G., & Schnabel, C. (2009). Data protection in Germany I: The population census decision and the right to informational self-determination. Computer Law & Security Review, 25, 84-88.

[37] Horvath, S. (2013). Aktueller Begriff Big Data. Wissenschaftliche Dienste des Bundestages; 2013. Available from: https://www.bundestag.de/blob/194790/c44371b1c740987a7f6fa74c06f518c8/big_data-data.pdf.

[38] Janssen, M., & van den Hoven, J. (2015). Big and Open Linked Data (BOLD) in Government: A Challenge to Transparency and Privacy? Government Information Quarterly, 32, 363-368.

[39] 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.

[40] Jarchow, T., & Estermann, B. (2015). Big Data: Chancen, Risiken und Handlungsbedarf des Bundes. Ergebnisse einer Studie im Auftrag des Bundesamts für Kommunikation. Bern: OFCOM.

[41] Jetzek, T. (2016). Managing Complexity Across Multiple Dimensions of Liquid Open Data: The case of the Danish Basic Data Program. Government Information Quarterly, 33(1): 89-104.

[42] Jetzek, T., Avital, M., & Bjørn-Andersen, N. (2014). Generating Sustainable Value from Open Government Data in a Sharing Society. In: Bergvall-Kåreborn, B., & Nielsen, A.P. (eds.) (2014). Creating Value for All Through IT. Berlin, Heidelberg: Springer, pp. 62-82.

[43] Kim, G.H., Trimi, S., & Chung, J.H. (2014). Big-data Applications in the Government Sector. Communications of the ACM, 57(3), 78-85.

[44] Klievink, B., Bharosa, N., & Tan, Y.H. (2016). The Collaborative Realization of Public Values and Business Goals: Governance and Infrastructure of Public – Private Information Platforms. Government Information Quarterly, 33(1), 67-97.

[45] Klievink, B., Neuroni, A., Fraefel, M., & Zuiderwijk, A. (2017). Digital Strategies in Action: a Comparative Analysis of National Data Infrastructure Development. In: Proceedings of the 18th Annual International Conference on Digital Government Research, ACM, pp. 129-138.

[46] Lindell, Y., & Pinkas, B. (2009). Secure Multiparty Computation for Privacy-preserving Data Mining. Journal of Privacy and Confidentiality, 1(1), 59-98.

[47] Millard, J. (2015). Open Governance Systems: Doing More with More. Government Information Quarterly. Available from: http://www.sciencedirect.com/science/article/pii/S0740624X15300034.

[48] Mitchell, R.K., Agle, B.R., & Wood, D.J. (1997). Toward a Theory of Stakeholder Identification and Salience: Defining the Principle of Who and What Really Counts. Academy of Management Review, 22(4), 853-888.

[49] Neirotti, P., De Marco, A., Cagliano, A.C., & Mangano, G. (2014). Current Trends in Smart City Initiatives: Some Stylised facts. Cities, 38, 25-36.

[50] Neuroni, A.C., Fraefel, M., & Riedl, R. (2011). Inter-organizational Cooperation in Swiss e-Government. In: Janssen, M. et al. (eds.) (2011). Electronic Government. Proceedings of the 10th International Conference on Electronic Government, EGOV, pp. 259-272.

[51] Neuroni, A., Fraefel, M., Estermann, B., Jarchow, T., & Golliez, A. (2016). Exploring the notion of a national data infrastructure and the governance issues surrounding it. Joint proceedings of ongoing research of IFIP EGOV and ePart, 191-199.

[52] Nugroho, R.P., Zuiderwijk, A., Janssen, M., & de Jong, M. (2015). A comparison of national open data policies: lessons learned. Transforming Government: People, Process and Policy, 9(3), 286-308.

[53] OECD (2013). Exploring Data-Driven Innovation as a New Source of Growth. Mapping the policy issues raised by "Big

Data". OECD Digital Economy Papers, No. 222, Paris: OECD Publishing. http://dx.doi.org/10.1787/5k47zw3fcp43-en.

[54] OECD (2015). Data-Driven Innovation: Big Data for Growth and Well-Being. Paris: OECD Publishing.

[55] Ohm, P. (2009). Broken Promise of Privacy: The Surprising Failure of Anonymization. UCLA Law Review, 57, 1701-1777.

[56] Ostrom, E. (2003). How Types of Goods and Property Rights Jointly Affect Collective Action. Journal of Theoretical Politics, 15(3), 239-270.

[57] Poikola, A., Kuikkaniemi, K., & Honko, H. (2015). MyData – A Nordic Model for Human-Centered Personal Data Management and Processing. Open Knowledge Finland. Available from: http://urn.fi/URN:ISBN:978-952-243-455-5.

[58] Powell, W.W. (1990). Neither Market nor Hierarchy: Network Forms of Organization. Research in Organizational Behavior, 12, 295-336. [1] OECD. Data-Driven Innovation: Big Data for Growth and Well-Being. Paris: OECD Publishing; 2015.

[59] Rifkin, J. (2014). The zero marginal cost society: The internet of things, the collaborative commons, and the eclipse of capitalism, Macmillan.

[60] Ross, J.W., Weill, P., & Robertson, D.C. (2016). Enterprise Architecture as Strategy: Creating a Foundation for Business Execution. Boston: Harvard Business School Press.

[61] Rouvroy, A. (2008). Privacy, Data Protection, and the Unprecedented Challenges of Ambient Intelligence. Studies in Ethics, Law, and Technology, 2(1). https://doi.org/10.3233/978-1-61499-570-8-11610.2202/1941-6008.1001.

[62] Rouvroy, A., & Poullet, Y. (2009). The Right to Informational Self-Determination and the Value of Self-Development. Reassessing the Importance of Privacy for Democracy. In: Gutwirth, S., Poullet, Y., De Hert, P., de Terwangne, C., & Nouwt, S. (2009). Reinventing Data Protection? Dordrecht: Springer, 45-76. https://doi.org/10.1007/978-1-4020-9498-9.

[63] Schmidt, A., Otto, B., & Österle, H. (2010). Integrating information systems: case studies on current challenges. Electronic Markets, 20(2), 1-14.

[64] Shadbolt, N. (2013). Midata: towards a personal information revolution. Digital Enlightenment Yearbook, 202-224.

[65] Shapiro, Z. (2015). Big Data, Genetics, and Re-Identification. Bill of Health, Harvard Law. Sept. 24 2015. Available from: http://blogs.harvard.edu/billofhealth/2015/09/24/big-data-genetics-and-re-identification/.

[66] Shin, D.-H. (2007). A Critique of Korean National Information Strategy: Case of National Information Infrastructures. Government Information Quarterly, 24, 624-645.

[67] Su, X., Hyysalo, J., Rautiainen, M., Riekki, J., Sauvola, J., Maarala, A.I., & Honko, H. (2016). Privacy as a Service in Digital Health. arXiv preprint arXiv:1605.00833.

[68] Sweeney, L. (2000). Simple Demographics often Identify People Uniquely, Carnegie Mellon University. Carnegie Mellon University, Data Privacy Working Paper, 3, 1-34.

[69] Thapa, B. (2016). Big Data in der Verwaltung: Chance mit Fallstricken, SGVW. Available from: http://www.sgvw.ch/2016/01/14/bigdata_thapa/.

[70] The Danish Agency for Digitization (2012). Good basic data for everyone. A driver for growth and efficiency. Danish Government, Local Government Denmark.

[71] UNECE (2016). Big Data in Official Statistics. United Nations Economic Commission for Europe. Available from: http://www1.unece.org/stat/platform/display/bigdata/Big+Data+in+Official+Statistics.

[72] Warren, S., & Brandeis, L. (1890). The Right to Privacy. Harvard Law Review, 4(5).

[73] Washington, A.L. (2014). Government Information Policy in the Era of Big Data. Review of Policy Research, 31, 319-325.

[74] Weber, K., Otto, B., & Österle, H. (2009). One Size Does Not Fit All – A Contingency Approach to Data Governance. ACM Journal of Data and Information Quality, 1(1), 1-27.

[75] Weber, R.H. (2016). Transparenz und Open Data. Jusletter IT, 25 May 2016.

[76] Westin, A. (1967). Privacy and Freedom. New York: Ateneum.

[77] Williamson, O.E. (1999). Strategy Research: Governance and Competence Perspectives. Strategic Management Journal, 20, 1087-1108.

[78] Wylot, M., Cudré-Mauroux, P., & Groth. P. (2014). Tripleprov: Efficient Processing of Lineage Queries in a Native RDF store. In: Proceedings of the 23rd International Conference on World Wide Web, ACM, pp. 455-466.

[79] Zuiderwijk, A., Janssen, M., Choenni, S., Meijer, R., & Alibaks, R.S. (2012). Socio-technical Impediments of Open Data. Electronic Journal of e-Government, 10(2).