Exploring the Notion of a National Data Infrastructure and the Governance Issues Surrounding It

Alessia C. NEURONI\textsuperscript{a}, Marianne FRAEFEL\textsuperscript{a}, Beat ESTERMANN\textsuperscript{a}, Thomas JARCHOW\textsuperscript{a} and André GOLLIEZ\textsuperscript{b}
\textsuperscript{a} E-Government Institute, Bern University of Applied Sciences
\textsuperscript{b} Opendata Switzerland

Abstract. A national data infrastructure (NDI) provides data, data-related services and guidelines for the re-use of data as an easily accessible service to citizens as well as public and private organizations. As such, it allows the efficient sharing of data between providers and consumers, supports new business models, and is thus a key enabler for the digital economy, societal collaboration and political processes. The paper relates to an ongoing project, discusses prevailing concepts on (data) infrastructure development and proposes a classification scheme for conceptualizing national data infrastructures in a given context. The discussion in particular focuses on governance issues related to establishing and maintaining a national data infrastructure that goes beyond the focus on open government data.

Keywords. E-government, government data, open data, infrastructure development, governance of networks, role of the state.

1. Introduction

In order to develop their full potential for the digital economy and society, data need to be provided extensively and systematically. As the OECD study on data-driven innovation observes, 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 nonmarket goods that generate positive externalities [1]. Managing infrastructure resources in an openly accessible manner may be socially desirable when they facilitate such downstream activities [2]. This principle has been recognized by the application of the "open data" principles to government data and research data.

In the era of big data, opening up datasets is however not enough: in order to be able to effectively extract value by gaining new insights through recombining data, data need to be enhanced in a way that they can easily be connected to data from various sources. Both the process of data publication and stewardship as well as data enhancement are costly undertakings, which potentially benefit a large number of downstream users. Data governance understood as the guiding of collective action therefore not only needs to address the question of who gets access to what data for what purpose under what conditions, but also to assign responsibilities and retribution.
mechanisms for data maintenance and enhancement in order to ensure the sustainability of the common infrastructure.

Building shared (open) data infrastructures has become a priority of governments around the world. Just as electricity, streets, and water are core infrastructures that serve citizens, companies, and governments alike, so too can a data infrastructure be understood as a community-wide need respectively a public good, similar to education, human resources, healthcare, and public services [3].

The present paper relates to an ongoing research project aimed at fostering the debate on the establishment and governance of a national data infrastructure (NDI) in Switzerland. Since the project is still in its initial phase, the goal of the present paper is to present basic considerations on data infrastructure development. Its main contribution is a preliminary framework for characterizing NDIs based on a literature review.

The paper is structured as follows: In the next section we provide a brief outline of the project. In section 3 we present a frame of reference for a NDI by discussing relevant literature in the field. Section 4 builds upon the literature review and provides a classification scheme for clarifying the notion and characteristics of a NDI. The paper closes with a discussion of the main considerations and an outlook to the next activities.

2. Description of the Research Project

The considerations presented in this paper relate to the ongoing project «Governance for a National Data Infrastructure in Switzerland» ¹. The goal of the project is to identify and address governance issues related to establishing and maintaining a national data infrastructure. While several countries have made first experiences regarding the establishment of a NDI, a corresponding project in Switzerland has still to gain contour. The first step of the project is therefore to foster a common vision of a national data infrastructure and to sketch out a draft model for its realization. The research questions to be tackled are:

- What are the elements of a national data infrastructure?
- Who are the key stakeholders and what is their expected role in this context?
- What are the main challenges and important governance issues?
- Which activities should a roadmap for creating an NDI address?

The methodological approach of the project is based on an analysis of literature in relevant areas (infrastructure resources, prevailing concepts of (data) infrastructures, developing and governing of shared infrastructures), three case studies on existing initiatives in other countries (UK, Denmark and the Netherlands) and around 15 semi-structured interviews with (potential) key stakeholders (public administration, the private sector, civil society, and academia). The selection of stakeholders follows an ecosystem approach [4] and has been guided by the concept of stakeholder salience, i.e. power, legitimacy, and urgency [5].

The project adopts a holistic, multi-disciplinary view on the issues at stake (technical, semantic, economic, societal, and legal aspects), promotes a shared understanding of infrastructure development and provides the basis for concretizing and coordinating activities in that respect. It is based on the assumptions that the

¹ The project is founded by the Hasler Foundation as a pre-project for a submission in the National Research Programme «Big Data» at the Swiss National Science Foundation.
provision and the realization of the benefits of a national data infrastructure relies on collective action [6], [7] and that an open, co-productive approach to its governance will foster sustainability (cf. [8], [9]).

The project is in its initial phase. The interviews have been conducted; their analysis and the development of the draft model for conceptualizing the Swiss NDI are however still outstanding. The paper thus presents results from the literature analysis.

3. Reference Frame for a National Data Infrastructure

3.1. Data as an Infrastructure Resource

Data and data analytics have become an essential driver of innovation, and it has been argued that data should be considered as one of our society’s central infrastructure resources [1]. From an economic point of view, infrastructure resources are fundamental resources that don’t 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 is often socially desirable to manage them in an openly accessible manner [2], [10]. This has for instance been recognized by the application of the “open data” principles to government data and research data.

According to a classification provided by Frischmann [2], data relate to non-traditional infrastructures (information resources, internet resources) that – just as traditional infrastructures – have the potential to generate positive externalities and result in social gains. Data meet the following characteristics of infrastructural resources: 1) they may be consumed in a non-rival fashion for some appreciable range of demand; 2) the social demand for data as resource is driven primarily by downstream productive activities that require data as an input and 3) they may be used as an input into a wide range of (private, public or social) goods and services.

3.2. Interrelations Between Prevailing Concepts of (Data) Infrastructures

The notion of a national data infrastructure is not straightforward, but bears connections to and overlaps with other concepts dedicated to infrastructure development in a digital environment. This includes concepts on developing e-government infrastructures, national information infrastructures, or open data infrastructures.

Irrespective of the given focus of interest, the different concepts have in common that there is usually no common understanding of what an infrastructure comprises (cf. [11], [12]). Research in the field stresses that infrastructures comprise both technical elements (hardware, networks, services, etc.) and social elements (management, governance, standards, agreements etc.) [13], [11]. In that respect, Jetzek distinguishes between an IT infrastructure and a regulatory infrastructure [14]. Typically, infrastructures are or should be flexible [13] and evolve over time in accordance with the needs of their multiple users [15].

3.2.1. E-Government Infrastructures

Infrastructure development is a core issue for improving public service-provision in the context of e-government. The focus is on shared infrastructures for enabling inter-
agency collaboration [13], [12]. A common e-government infrastructure lays the 
ground for interoperable and re-usable solutions that allow for providing public 
services seamlessly. Among other requirements, this necessitates the exchange and re-
use of data that are often stored in multiple information systems held by different 
actors. Janssen et al. propose the following hierarchy of interoperability requirements 
in e-government, ensuring the interoperability of data being the most basic task to be 
accomplished [16]:
- organizational interoperability (collaborating, designing cross-agency 
  processes and supply chains);
- interoperability of services (sharing, re-using services or components);
- interoperability of applications (integrating applications);
- interoperability of data (sharing information from heterogeneous systems).

The relevance of interoperable data and information sharing is also stressed in 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. The focus is on base registries that are legally controlled and maintained by 
public administrations and provide authentic sources of information on items such as 
persons, companies, vehicles, licenses, buildings, locations or roads. The European 
Commission recommends making this information available for re-use while installing 
appropriate security and privacy measures for managing access and control [17]. With 
view to e-government development, authentic sources are a key enabler for enhanced 

service provision as they are a necessity for improving user experience and 
administrative efficiency [18], [19].

3.2.2. OGD Infrastructures

Infrastructure development is also a major issue for fostering open government data 
initiatives. The focus is on shared infrastructures, which allow third parties to make 
use of OGD. OGD initiatives – with a focus on the development of policies and central 
data portals or data catalogues – are usually considered as a subset or an extension of e-
government [20]. While a clear demarcation between e-government infrastructures and 
OGD infrastructures is not always possible, distinguishing features typically relate to 
the type of government data (public data vs. open data) and the associated goals of data 
sharing (improving public service provision vs. stimulating service innovation by third 
parties).

Most contributions on open data infrastructure development are concerned with 
guiding strategies and the existence and functionalities of open (government) data 
portals. OGD benchmarks often assess data provision against the widely acknowledged 
open data principles [21] or the G8 open data charter [22]. Besides measuring data 
availability (range of data) and accessibility (data formats, licensing, costs, etc.), user 
support and functionalities for stakeholder engagement are receiving increased 
attention (e.g. [23], [24]). Availability of data mostly relates to coverage of sectoral 
data (education, health, finance, etc.) [25], while the provision of basic government 
data (key registries) is hardly a topic. One reason for this might be that OGD strategies’ 
focus on the open data principles tends to foreclose the integration of government data 
that are unlikely to be governed by these principles (e.g. in terms of licensing or access 
control). Thus, basic e-government data and open government data tend to be dealt 
with separately. This is also reflected in existing governance structures: As a recent
study shows, the vast majority of national open data portals have been developed by governmental actors, but implemented independently from existing national government portals [23].

3.2.3. National Data or Information Infrastructures

Compared to the common understanding of open data infrastructures, the notion of a national information infrastructure (NII) – or similarly – a national data infrastructure (NDI) is more open with regard to data, implementation options, fields of application and goals. Data infrastructures can comprise data that is owned by governments, businesses or non-profit organizations, the data can be openly licensed, it can be made available for re-use by specific stakeholders or be closed [26]. The goal of establishing a data infrastructure is to make data available and re-usable as far as possible in order to realize social, environmental or economic value generation. To this end, relevant data should be identified under a strategic framework in order to improve data governance [26].

Several countries have adopted the concept of a national data or information infrastructure in order to effectively share core government data sets within and outside government and stimulate their use across boundaries [26]. Thereby it can be observed that the adopted initiatives or policies conceptually strengthen connections between government data held in base registries and OGD. This is the case for instance in the UK [27], [28], [29], in Denmark [30], [31], [14] and the Netherlands [19].

3.2.4. Consolidated View on Data Related Infrastructure Development

The terms "information infrastructure" and "data infrastructure" tend to be used synonymously. We propose to use the term "national data infrastructure", as it is closer to the terminology used in the OGD context and more elementary in terms of an information hierarchy [32]. Regardless of the terminology used, the type of infrastructure at hand is characterized by a range of components and the involvement of various actors with different requirements. Accordingly, the governance of such infrastructures needs to be developed by the stakeholders collaboratively [33].

3.3. Infrastructure Governance and the Role of the State

Seizing the benefits from data driven innovation requires collective action and the willingness of collaboration in order to create economic and public value [34]. While company-wide data governance frameworks have been a topic in information science for decades [37], the big data era confronts us with the same task, but worldwide and in a setting where power structures are less regulated. Against this background, political actors are confronted with a need to think about the roles the state should play in the data economy and how to concretely fill them in. The development of shared data infrastructures in which the state is likely to be involved can therefore be defined as a governance challenge.

Generally, governance can be described as the process of horizontal coordination in which heterogeneous actors are involved in creating a shared understanding and definition of the problems they are confronted with and of the measures to be taken to resolve them [13]. A governance framework needs to focus on the key elements that are relevant for a cross-boundary common view of the reality [38] and should support a vision that satisfies all relevant stakeholders [39] who may act according to different
rationalities, i.e. a legal, economic and/or technological one [13]. In particular, it should be noted that data driven value creation may heavily rely on activities by non-governmental actors.

When seeking to establish a NDI, 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. Core elements to be addressed include considerations on data access and reuse, portability and interoperability, linkage and integration, quality and curation, “ownership” and control as well as data value and pricing (cf. [1], [35]). To facilitate the creation of public and economic value, incentive systems for collective action and collaboration are required, covering the entire data life-cycle [13], [34]. 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, including the protection of privacy [36].

As for the latter aspect, the state clearly plays a crucial role as regulator. Policies on the usage of data are however only one aspect of data politics. With regard to developing and maintaining a NDI, the state can potentially adopt a range of roles. Shin for instance distinguishes between the role of government as a direct intervener (strategist, builder, regulator, and producer) and the role as an indirect facilitator (guider, leader, and integrator) [3]. With view to the data value chain potential roles of governmental actors can further be differentiated and extended, e.g. as data collectors, data users, operators of a system or infrastructure, as service providers or administrators [1].

4. Conceptualizing a National Data Infrastructure

Based on relevant literature on e-government and OGD infrastructure development, we propose the following classification scheme for discussing the establishment of a NDI in the form of a morphological box. The goal is to provide a basis for developing and testing implementation scenarios and to structure possible policy elements during the iterative research process of the ongoing project [41]. The selection of variables is based on a team-internal discussion and has been guided by the idea of describing the main characteristics, instead of detailing all possible sub-characteristics (e.g. regarding data provision, cf. [42]).

The first cluster of characteristics relates to fundamental considerations on a NDI, i.e. its nature, value and scope. The second cluster relates to considerations related to governing infrastructure development, i.e. its basic elements, strategic foundation, architecture and governance. The third cluster focuses on the data to be made available for re-use through a shared infrastructure, i.e. the stakeholders involved in the data process and the type of data under consideration.

| Table 1. NDI Classification Scheme for Characterizing NDIs |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|
| Characteristics | Basic Notion of the NDI |
| General perspective | ideational / guiding | strategic / controlled | functional / operational | technical / physical |
| Value orientation | public-value-oriented | business-oriented | mixed |
| Scope & expected impact | local issues | national issues | global issues |
| Role of the state [3] | proactive intervener | facilitator |
The preliminary classification scheme serves as a basis for conceptualizing the Swiss national data infrastructure. It will be used for analyzing the stakeholder interviews, i.e. for identifying salient characteristics of a future data infrastructure and its governance. Conversely, the stakeholder input will help us concretize the classification as a basic model for strategy and governance decisions.

5. Conclusions and Outlook

Understanding data as a resource requires considerations on establishing shared infrastructures for facilitating re-use of data. The notion of a national data infrastructure serves to foster an integrated view on data-based value generation, thereby accounting for developments in the areas of data-sharing in e-government and through OGD-portals. It relates to the idea of making data available for re-use under a common framework that is generally open with respect to data ownership (state, private) or usage conditions (shared, open).

The concretization of such a framework is likely to differ across countries and is depending on existing structures and cultures. Based on a literature review, a preliminary classification scheme for characterizing NDI has been presented. It will
serve for the description of different stakeholders’ conceptions of a NDI and the governance issues at hand in the case of Switzerland. The classification scheme is meant to be generic and requires further validation. It could be used and tested as an instrument for analyzing and discussing the conceptualization and development of NDIs in other national contexts or for conducting cross-country comparative studies.

The goal of the ongoing project is to facilitate the development of a common vision of a NDI and the identification of challenges that need to be taken into account when developing a governance framework. The main issue to be addressed is the identification of stakeholder roles in the ecosystem. As for governmental actors, we are interested in clarifying the role of the state in two respects: as an enabler for the development of new data-based services by third parties and as a facilitator for modernizing public service provision. In both cases, the design of sustainable business models for data provision, enhancement, and stewardship, as well as the overcoming of collective action problems will be crucial.

6. References