Article

Metadata For Identity Management of Population Registers

Olivier Glassey

Swiss Public Administration Network (SPAN) and Swiss Graduate School of Public Administration (IDHEAP), Quartier Mouline UNIL, Lausanne 1015, Switzerland; E-Mail: olivier.glassey@idheap.unil.ch; Tel.: +41-21-557-40-20; Fax: +41-21-557-40-09.

Received: 21 January 2011; in revised form: 2 April 2011 / Accepted: 6 April 2011 / Published: 18 April 2011

Abstract: A population register is an inventory of residents within a country, with their characteristics (date of birth, sex, marital status, etc.) and other socio-economic data, such as occupation or education. However, data on population are also stored in numerous other public registers such as tax, land, building and housing, military, foreigners, vehicles, etc. Altogether they contain vast amounts of personal and sensitive information. Access to public information is granted by law in many countries, but this transparency is generally subject to tensions with data protection laws. This paper proposes a framework to analyze data access (or protection) requirements, as well as a model of metadata for data exchange.

Keywords: public; administration; metadata; information; sharing; transparency; privacy; framework; access

1. Introduction

The 2009 Ministerial Declaration on eGovernment of Malmö [1] recognized eIdentity as one of the key enablers for eGovernment, the latter being defined as “the use of Information and Communication Technologies in public administrations combined with organizational change and new skills in order to improve public services”.

In this paper I argue that identity management tools are central to the domain of population registers and that the latter are the pillars of many administrative services. Section 1 introduces the thematic of population registers and discusses the tensions between data protection requirements and data sharing requirements. Section 2 describes a conceptual model and a method in order to model requirements in terms of data access, data protection and identity. The metadata model described in Section 3 is built
upon this conceptual model. Finally, Section 4 examines the potential contributions of this approach and gives a few directions for future research.

1.1. Population Registers

A population register is an inventory of residents within a country, with their characteristics (date of birth, sex, marital status, etc.) and other socio-economic data, for instance occupation or education. According to the United Nations Statistics Division’s definition [2], "the main administrative functions of population registers are to provide reliable information for the various purposes of government, particularly for program planning, budgeting and taxation; for issuing unique personal identification numbers; for establishing the eligibility of individuals for voting, education, health, military service, social insurance and welfare and the pension system; and for police and judicial references. Population registers are also useful for population estimation, census planning, and census evaluation and for sampling frame of household surveys." In addition to stricto sensu population registers, there are many other public registers or official sources that store data on citizens: Tax, land, building and housing, military, foreigners, insurance, debt collection, welfare, courts, retirement, vehicles, and so on.

All over Europe public registers are changing from paper-based files to digitalized registers and from decentralized registers to interoperable registers, although situations differ strongly from country to country. Moreover in the last decades many European countries have adopted laws on transparency and access to public information on the one side and laws preserving the privacy rights of individuals on the other. These laws apply to public register data as well, but at the moment there is limited questioning regarding the strategies to be developed: are data sharing and data protection compatible, should tension mitigation processes be set up to find a balance between data sharing and data protection in a given context and should data sharing or data protection prevail? Data from population registers are subject to tensions between (notably) security, privacy, interoperability, usability and intrusiveness, and experience in using such data is still lacking in most European countries.

The situation of public registers and privacy policies varies considerably from country to country. It is very difficult to present a global view of the situation in Europe, but let us briefly present some representative examples.

Finland has a long history in the field [3] and was amongst the first European countries to collect data on population, during the sixteenth century; census lists (local population registers) have been maintained in Finland since 1634. Finland moreover created a centralized population register by law in 1969 and the computer-based register was introduced in 1971, also pioneering in that domain. Finnish citizens may have access to the central population registers and can request the authority to correct their data. Last the Finnish eID card, introduced in 1999, was the first ever operational national eID scheme.

In the United Kingdom, mandatory civil registration of births, marriages and deaths was first introduced in 1837 [4]. The United Kingdom still has decentralized paper-based civil registers; the administration of individual registration districts is the responsibility of registrars in the relevant local authority. Official registers are not directly accessible by the general public; however indexes are made available that can be used to find relevant register entries. Certified copies of the entries are issued for genealogical research or for administrative purposes such as passport applications. Some local indexes
are published on the Internet. The UK does not have identity cards (although they are being piloted on a voluntary basis) and the proposal to introduce them is politically highly controversial.

The Kingdom of Belgium conducted a first general population census in 1846. Detailed population registers are operated at the local level and the Belgian National Register worked on the basis of voluntary co-operation of municipalities from 1969 until 1983, when it became mandatory [5]. The National Register contains a strictly limited data set of local registers. Citizens cannot directly consult the local registers, except that each individual has the right to request a copy of his/her own data. In 2003 Belgium was also one of the first countries to deploy an e-identity card nationwide [6] and by the end of 2009 all Belgians should have an eID card [7].

In June 2006 the Swiss Parliament adopted a new law on population registers’ harmonization in order to simplify statistical data collection and data exchange. Until 2004 vital records (births, deaths, weddings and adoptions) were held on paper registers by 1750 cantonal offices throughout Switzerland, but since 2004 there has been a federal centralized database called Infostar [8]. There are furthermore around 2500 resident registers, generally maintained by municipalities. For administrative purposes, Swiss citizens routinely request certified copies of their resident register entry or so-called vital records extracts. In 2009 Swiss citizens barely accepted (50.1%) the introduction of a biometric passport [9] and one of the hottest topics of the campaign was the possible use of a chip on the future identity card.

In Romania the first census was conducted in 1838 [10]. For the time being the population registers are operating at the local level, at the place of residence of each citizen by the Ministry of Administration and Interior, through the National Inspectorate for Population Register, which issues a traditional identity card. However, the Personal Identity Number [11] is generated and administered by the National Centre for Managing Databases of Population Register. Citizens have limited access only to their own private data, subject to laborious formal requests. In July 2010 the Romanian National Registry of Population was centralized and a pilot platform for issuing electronic identity cards should be introduced at the beginning of 2011.

Preliminary research by Pollitt [13], who studies civil registers in the UK, Finland and Belgium using a public management perspective, indicates that population registers have previously attracted very few research projects or scholarly papers within the field of public administration/public management. Public registers have been studied from a historical (or even genealogical), statistical and demographical perspective (dozens of such demographical studies are available at Popline [14]), but not in terms of public management.

1.2. Data Sharing vs. Data Protection

This paper does not specifically target the core components of knowledge management and information sharing. Indeed Pardo et al. [15] observed that research on knowledge sharing focuses on knowledge forms (tacit/explicit) and on the processes and technology that support (or prevent) knowledge sharing. It rather reflects on the sharing of public information (the supply side from the public administrations) and on the access to public information (the demand side from citizens, groups, lobbies and other public administrations).
Although I will not enter into an in-depth discussion of the concepts of access to information and transparency, let me mention a few key issues to be taken into account when developing the analysis framework. Most scholars and practitioners do not contend the idea of public data sharing, and some state that this is essential for a working democracy, along with Kierkegaard [16], who called upon Thomas Jefferson and his “information as currency of democracy” to declare that “the public must know what information is available from which government body, and how and where this can be located…” This is also the case in many national regulations around the world. In 2006, 70 countries had passed Freedom of Information Acts or access to information laws and 50 additional countries were in the process of doing so [17]. However the existence of a legal framework does not guarantee access to information per se: The Open Society Justice Initiative [18] shows that in countries with freedom of information laws, only 33% of the requests for information were fulfilled (with another 38% receiving no answer at all). Furthermore the existence of laws on access to information does not “necessarily contribute to the perceptions of transparency of government policymaking” [19]. This might be changing slowly, as Combe [20] emphasized that “data sharing … has required a shift in cultural norms in public sector organizations” and noted that “the mindset of government is for data sharing to take precedence over privacy”.

On the other hand, many countries also have laws on data protection or on privacy. Although the origin of the concept of privacy dates back to Aristotle, thinkers such as Locke [21] and Mill [22] further defined the concept. Later on Westin [23] developed his theory on the control of information. For him, privacy is “the claim of individuals, groups or institutions to determine for themselves when, how and to what extent information about them is communicated to others”. This approach has become the basis of many pieces of data protection legislation, although it is still relatively new; according to Hornung and Schnabel [24], the German Federal Constitutional Court “invented the new basic right of informational self-determination” in 1984. The authors furthermore argue that this right constitutes the legal basis for data protection in Germany and explain that it is grounded in the work of sociological systems theorist Niklas Luhmann. Nonetheless data protection and privacy are perceived rather differently in different regions of the world (e.g., in China, in the United States of America and in Europe). Regarding the dimensions of personal data protection, this framework is focused on European practices; they might be quite heterogeneous but there is a directive of the European Union that sets the stage. In a survey on the implementation of this European Union Directive 95/46/EC on data protection [25], Otjacques et al. discuss various legal issues pertaining to the processing of personal data, among which are:

- The obligation to notify a given authority when personal data are processed and the categories of exceptions, e.g., when specifically mentioned in the legislation;
- The rights of the person to be notified and to access his/her own data;
- Data sharing between administrations using an identifier, with or without a supervisory authority.

Here again, there might very well be a directive, but the question is then how can data protection be implemented practically? The classical approach has long been to “seek ways to prevent information from escaping beyond appropriate boundaries” [26]. However the authors argue that this is no longer sufficient and they propose an alternative approach based on accountability, where “the use of information should be transparent so it is possible to determine whether a particular use is appropriate
under a given set of rules and that the system enables individuals and institutions to be held accountable for misuse”. This is supported by a survey on the determinants of trust in government cross-boundary information sharing conducted by Gil-Garcia et al. [27], who shows that clearly defined roles and responsibilities support trust and thus information sharing.

More generally Dawes [28] proposes two principles for transparency policies of public information:

- **Stewardship** focuses on the accuracy, integrity and preservation of information holdings.
- **Usefulness** recognizes that government information is a valuable asset that can generate social and economic benefits through active use and innovation.

## 2. Conceptual Model

Formal techniques are used to model data sharing and data protection requirements, and to translate them into machine-readable metadata. In previous work [29] I discussed the concepts of knowledge representation and formalization in detail. These ideas are not new; mathematical tables, graphs or set theory had been used to structure knowledge long before information systems were developed. In this particular domain there are several accepted methods for formalizing knowledge, e.g., rules, semantic networks or concept diagrams. Several techniques are furthermore available for graphical knowledge representation or visualization, from ad-hoc drawings, visual metaphors or animations to conceptual diagrams or scientific charts [30]. Current technology goes further than representation as it supports the translation of graphical knowledge into machine-readable semantics [31]. Although there are several initiatives that support the visualization and formalization of basic identities and social networks, such as FOAF (Friend of a Friend), which uses RDF (Resource Description Framework) to describe persons, their activities and their relations to other people and objects, there is to my knowledge no tool to achieve this with a specific focus on privacy.

The conceptual framework defined in [29] is used as a starting point to define identity and privacy management processes. This set of methods, techniques and tools will allow us to model stakeholders and processes and to formalize relationships between actors, processes and data. They consist mainly of eight abstract models that represent reality symbolically, in terms of concepts and relationships, and are implemented as diagrams, *i.e.*, simplified and structured visual representations of concepts and relations. This framework, called MIMIK, furthermore uses RDF schemas for knowledge modeling. RDF can also be used to build data models for e-Government [32]. In his paper discussing identity issues brought up by the new Swiss law on population registers’ harmonization, Glassey [33] proposes a conceptual framework to analyze the data governance of these population registers, with a strong focus on information requirements and identity management. Such a framework is surely useful for analysts and domain experts, but it is not meant for citizens and end-users. The same applies to standardized approaches to data exchange, which are mostly understandable by people with a technical background. Identity and privacy management tools should provide very intuitive approaches to knowledge visualization, inspired by the latest Web 2.0 trends and social network approaches that many people seem to adopt very easily, even with very limited knowledge of the Internet.

As mentioned above, this framework was quite useful for analyzing the data governance of population registers. However, it was based on information criteria (amongst others effectiveness, confidentiality and reliability) and on citizens’ identities in a given context. It comprised the following
Building blocks: Data consumers, data sources, identity, requirements and data sets. It thus needed to be extended to include data sharing and data protection dimensions. The detailed selection process of relevant elements will not be explained here; they are listed in Table 1.

**Table 1.** Description of data sharing and data protection dimensions.

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Metadata</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data sharing</td>
<td>- Public source</td>
</tr>
<tr>
<td></td>
<td>- Legal basis</td>
</tr>
<tr>
<td></td>
<td>- Access rules</td>
</tr>
<tr>
<td></td>
<td>- Access log</td>
</tr>
<tr>
<td>Data protection</td>
<td>- Notification to authority</td>
</tr>
<tr>
<td></td>
<td>- Notification to person</td>
</tr>
<tr>
<td></td>
<td>- Use of identifier</td>
</tr>
<tr>
<td></td>
<td>- Access to own data</td>
</tr>
</tbody>
</table>

In order to complete the data sharing/data protection official policies a typology of restrictions to transparency was adapted; Pasquier and Villeneuve [34] propose a typology of organizational behaviors tending to prevent or restrict access to information. These behaviors should also be taken into account in the model (see Table 2). These are “grey” policies that might be difficult to formalize but are however relevant as contextual information for this modeling approach.

**Table 2.** Types of restrictions to transparency.

<table>
<thead>
<tr>
<th>Not Subjected to Transparency Regulations</th>
<th>Subjected to Transparency Regulations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transparency is <strong>not required</strong> but can be applied on a voluntary basis</td>
<td>- Transparency laws are simply rejected or bypassed</td>
</tr>
<tr>
<td></td>
<td>- All legal means are used to obstruct access to information</td>
</tr>
<tr>
<td></td>
<td>- Access to information is strained by such issues as a lack of resources or misunderstandings</td>
</tr>
<tr>
<td></td>
<td>- <strong>Full</strong> access to public information is proactively provided</td>
</tr>
</tbody>
</table>

Harmonized and/or digitalized public registers could deeply change existing administrative processes, as many public services use personal data that could be retrieved from population registers or other public data sources instead of citizens having to fill in dedicated forms. Some public services already undertake this where there is a legal basis to do so, but in other cases citizens could be empowered to open access to their data in order to simplify administrative procedures. Citizens have many roles (or identities) in regard to the public and para-public sector: They pay taxes, they elect their representatives, they might need permits (working, building, fishing, owning a dog) and in some cases they benefit from social help or even go to prison. There is a large number of potential administrative services (around 1200 of them have been identified in Switzerland [35]), but this shows the vast amounts of personal data that could be stored or used during these interactions.

Let us present a scenario for identity and privacy management of public registers’ data. By default citizens would have only one privacy profile, e.g., an empty profile, a profile containing only a
personal identification number or basic information (in accordance with national regulations). They would furthermore be offered the possibility to build different profiles according to their needs in a given context. Figure 1 shows basic profiles for an individual: One as a student with data to which he might want to give access to the university where he is registered, and another that he would use to vote and to fill in his tax declaration (possibly online).

**Figure 1.** Identities and privacy profiles.

![Diagram of identities and privacy profiles](image)

Individuals should not only be able to manage their privacy and identity profiles as such; they should also be able to decide when and to whom they allow access, and to which personal data within official public data sources, by using their privacy profiles (Figure 2). These privacy profiles would of course have to be strictly compliant with the legal rules, *i.e.*, the data sharing and data protection rules discussed above.

**Figure 2.** Access to a public data source granted by an individual.

![Diagram of access to a public data source](image)

In addition to *stricto sensu* population registers (mostly based on vital and residence data) there are many official data sources, with various levels of access. Figure 3 shows a selection of registers’ data that are used digitally in one way or another in Swiss administrative processes, through XML files, data interfaces, Webservices, *etc.* This is only shown as an example, but there are certainly similar computer-based public registers in many other European countries.
In addition to being able to manage their profile and give access to their data, individuals could also have tools allowing them to see who used what information in order to deliver an administrative service. One such example would be a tax decision, for which one could see where basic information (tax rate, taxpayer category, etc.) comes from and when it was retrieved, or what information from land registers is used for tax on property. This would support the auditability of administrative decisions.

If one wants to create a privacy profile and give access to personal data from various public registers, it would first be necessary to build on classical data management tools. These should at least support the following requirements in terms of data, amongst those defined by COBIT [36]:

- **Effectiveness**: Relevant, correct, consistent, usable and timely information is provided;
- **Confidentiality**: Sensitive information is protected from unauthorized disclosure;
- **Integrity**: Information is accurate, valid and complete;
- **Compliance**: Information use complies with the laws, regulations and internal policies.

Data management is indeed necessary but not sufficient: Such a scenario would also require personal identification mechanisms and identity management functionalities. These are inspired by Kim Cameron’s Seven Laws of Identity [37]:

- **User control and consent**: Digital ID systems must only reveal information identifying a user with the user’s consent;
- **Minimal disclosure for a constrained use**: The solution that discloses the smallest amount of identifying information and best limits its use is the most stable long-term solution;
- **Justifiable parties**: Digital ID systems must be designed so that the disclosure of identifying information is limited to parties having a necessary and justifiable place in a given identity relationship.

Personal identification resides at the heart of many forms of government service delivery [38] and historically and archetypically such identification was based on manual form completion and paper-based authentication processes (the citizen shows his identity card or passport to an official). For each administrative procedure a paper form had to be filled in and it was then stored (and most of the time forgotten). Now, with registers being digitalized and used in order to provide the information required for an administrative procedure, these identification mechanisms are no longer sufficient. Indeed, many debates over privacy have been driven by concerns about identity abuse, whether related to technological developments or not. Private companies increasingly gather information about their
customers, but governments have far more power to collect data that could be much more sensitive. It is thus crucial to set up appropriate tools for privacy and identity management.

3. Metadata Definition

In order to formalize the metadata needed to implement the scenario described in Figure 1–3, a simple metadata model was developed (Figure 4). The central concept is that of a Privacy Profile, a combination of:

- **Personal Data**: This is the description of the actual data from public registers
- **Identity**: This is the profile of the user (student, taxpayer, etc.)
- **A Data Policy** (explained below in Figure 5), where users define their preferences regarding data sharing and/or data protection; controls are applied so as to respect relevant regulations.

![Figure 4. Privacy profile model.](image)

![Figure 5. Public data access policy model.](image)

There are several well-developed and sound approaches to **Personal Data** and **Identity** metadata (for example Windows CardSpace [39], OpenID [40] or the open-source Higgins [41]). Furthermore many governments have defined their own practices, such the eCH-0107 [42] standard created by the Swiss eCH standardization body. These various models and standards do not all cover the same ground but
let us take Higgins (Figure 6) in order to exemplify the concept: Personal data (name, address, etc.) are formalized, as well as relationships with other people (on the basis of FOAF—Friend of a Friend [43]). On top of that, an identity layer is added on the basis of a SKOS (Simple Knowledge Organization System [44]) classification and on the SPARQL [45] query language and inference mechanisms (SPIN–SPARQL Inferencing Notation [46]). SPIN is used to attach rules and constraints to the identity and data models (e.g., to check that the data match one’s identity in a given context).

The generic Higgins model is used for Personal Data and Identity metadata, but a specific Data Policy model is needed. There have indeed been efforts to develop privacy policy markup languages (for example P3P–Platform for Privacy Preferences [47]) or access control languages (e.g., XACML–eXtensible Access Control Markup Language [48]). These languages do provide a sound reference to develop a specific set of relevant metadata. They do not however completely suit the requirements of the framework and this is why it needed some adaptations. For the basic description of Data Policies (Figure 5), standards such as Dublin Core are used. A few concepts regarding the life cycle of a document are added: Creation, version, validation, publication and expiration. Data are accessible on the basis of given data access policies only between the time when they were officially validated and published until their expiration date, if they are compliant with data protection requirements. Again, SPIN capabilities are used in order to match policies, data and individuals.

Figure 6. Higgins personal data model [49].
4. Conclusions and Future Work

This framework to analyze access is grounded in previous work on knowledge modeling and data governance. Furthermore new dimensions or concepts are introduced as regards data sharing and data protection on the basis of a literature review. This framework should be quite useful for:

- Analyzing data sharing in a given context: What data are available, what rules are applicable, are there data protection requirements, and so on?
- Identifying potential issues, *i.e.*, if there is full transparency, why bother? If it is a case of strained transparency, one might take a different strategic approach to solve it from that used in a case of obstructed transparency.
- Describing data sharing processes, as well as stakeholders.
- Managing documents’ life cycle with regard to the context, the legal requirements and the stakeholders’ requests for data access.

Building on this analysis framework specific metadata can be used to describe public data access policies, which could be combined with existing metadata XML implementations.

Although existing technologies (RDF, SPARQL, *etc.*) should allow us to implement and to interpret access rules, I believe there is still a strong need for traceability and accountability in the context of public data. In order to investigate these issues further, I plan to work with credential-based techniques that support the management of exceptions (see for example [50]).

References

40. OpenID. Available online: http://openid.net/ (accessed on 4 April 2011).
44. SKOS Simple Knowledge Organization System. Available online: http://www.w3.org/2004/02/skos/ (accessed on 4 April 2011).
45. SPARQL Query Language for RDF. Available online: http://www.w3.org/TR/rdf-sparql-query/ (accessed on 4 April 2011).
47. Platform for Privacy Preferences (P3P). Available online: http://www.w3.org/P3P/ (accessed on 4 April 2011).


© 2011 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).