A FRAMEWORK TO ANALYZE DATA GOVERNANCE OF SWISS POPULATION REGISTERS

Olivier Glassey[1]

In June 2006 the Swiss Parliament adopted a new law on population registers' harmonization in order to simplify statistical data collection and data exchange from around 4'000 decentralized registers. Besides there are more than 2'000 administrative services delivered to Swiss citizens and businesses, of which hundreds could potentially use data from population registers. The law is rather vague about the implementation of this harmonization and even though many projects are currently being undertaken in this domain, most of them are quite technical. We believe there is a need for analysis tools and therefore in this paper we propose a conceptual framework to analyse data governance of these populations registers, with a strong focus on information requirements and identity management. In order to develop this framework we built on existing approaches to define its building blocks: data consumers, data sources, identity in a given context, requirements, and data sets.

Keywords: governance, data, identity, population registers, modelling, framework.

1. Introduction

In June 2006 the Swiss Parliament adopted a new law on population registers' harmonization in order to simplify statistical data collection and data exchange. Indeed public registers are very fragmented in Switzerland. Until 2004 vital records (births, deaths, weddings, and adoptions) were held on paper registers by 1'750 Cantonal offices throughout Switzerland. Since 2004 the Federal Department of Justice and Police provides Cantons with a centralized database called Infostar. Furthermore there are around 2'500 resident registers (localisation of citizens), generally maintained by the 2'700 Swiss communes with some exceptions such as Geneva where they are operated at the cantonal level. In the Canton of Bern alone, there are 396 communes that use 26 different software solutions to manage residents' data. In addition to these "stricto sensu" population registers, there are several other registers that store data on citizens: fiscal register, foreigners' register, building and housing register, and so on. There are furthermore several databases that are not directly considered as registers but that are connected to these registers and store records on military, scholarships, old age and survivors' insurance, etc. A Federal Census will take place in 2010 and it is planned that data should be provided by these various population registers. Previous censuses were based on questionnaires sent to Swiss households and interviews. In order to collect and aggregate data automatically, a personal identification number had to be created. A first project called EPID (Unique Federal Personal ID) was rejected, mainly for data protection issues. It was then planned to have six sector ID numbers (population, social security, tax, defence, justice,

[1] Swiss Public Administration Network (SPAN) and Swiss Graduate School of Public Administration (IDHEAP), Route de la Maladière 21, 1022 Lausanne, Switzerland, olivier.glassey@idheap.unil.ch

statistics) in order to insulate data and to protect citizen's privacy. However this has not come to realization yet, and in order to implement the harmonization law, Switzerland has chosen to use the existing old-age and survivors' insurance number. However this number was not unique and not anonymous, therefore a new one is being introduced, starting July 2008.

2. Theoretical Framework

2.1 Data Governance

Although there is currently no consensus on the definition of governance, there are two domains that are well defined and widely accepted: political governance (whether global, local or territorial) and corporate governance. Both share a common approach based on decision processes and stakeholders' participation (shareholders, executives, political leaders, citizens, interest groups or any other organisation). This paper is discussing the governance of population registers, so it is clearly not corporate governance we are studying. What we want to survey is the elaboration of rules and decision processes that govern the management of public registers' data. Before 2004 there was no clear need for such a governance approach, as with more than 4'000 disconnected or paper-based registers it was not such a big issue. However, information and communication technologies (ICTs) are currently transforming these processes and we want to analyse how they are impacting our society, under the umbrella of what has been called eGovernance or digital governance.

[1] believe that it is necessary to establish reference principles in order to measure these changes and their impacts. They mention *accessibility*, *transparency* and *accountability*. We agree with them but we thought some additional dimensions were needed. In its policy paper on governance for sustainable human development, [2] defines governance as the exercise of economic, political and administrative authority to manage a country's affairs at all levels and proposes the following characteristics for governance: *Participation, Rule of law, Transparency, Responsiveness, Consensus orientation, Equity, Effectiveness and efficiency, Accountability, Strategic vision.*

At this point and before going further on into our data governance model, we think it is useful to provide a quick reference model on eGovernment and eGovernance and we will summarize their definitions by [3]:

- eGovernment covers three main areas: electronic services provision, electronic workflow and electronic voting (in terms of infrastructure, security and trust).
- eGovernance consists of electronic consultation (public administrations, citizens, businesses, interest groups, etc. give their input during the rule-creation process) and electronic participation (the afore-mentioned actors have not only the ability to give their opinion during decision processes but also to initiate them by using ICTs); the "Holy Grail" of eGovernance is a networked society where actors can follow decision processes (accessibility) and provide feedback and control (transparency) in order to make officials and elected responsible (accountability).

So far we have defined the general dimensions of governance and we have explained the particularities of eGovernance or digital governance, but we still need to define a specific set of requirements in terms of data governance. To that effect we will use the information criteria of CoBIT [4]. This framework for IT governance consists of a set of good practices ensuring that IT is aligned with the business and enables business processes. It also provides resources for risk and performance management. The data requirements (or information criteria) defined by CoBIT are:

- *Effectiveness*: relevant, correct, consistent, usable and timely information.
- *Efficiency*: provision of information through the optimal use of resources.
- *Confidentiality*: protection of sensitive information from unauthorised disclosure.
- *Integrity*: accuracy, validity and completeness of information.
- *Availability*: information is available when required.
- *Compliance*: information use is complying with the laws, regulations and internal policies.
- *Reliability*: information can be trusted.

2.2 Identity Management

In section 2.1 we defined a set of requirements in terms of governance and data governance; we believe it is necessary to integrate them in our framework to study population registers' harmonization, but we also think that this is not sufficient and we argue that an identity layer must also be added to our framework. Indeed, citizens have multiple identities when dealing with different parts of the public sector: as an individual, as a family, as a taxpayer, as a soldier, as a patient, possibly even as a prisoner, and so on. [5] write that personal identification resides at the heart of many forms of government service delivery and that historically and archetypically such identification was based on manual form filling 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 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 not sufficient anymore. We will cite to Canton of Geneva eGovernment strategy [6] to illustrate this: it is stated that no Cantonal office should ask citizens for information that the State of Geneva already has somewhere.

However this paper is not about a technical approach of eIdentity management (user name, password, personal identification numbers, smart cards, PKI, fingerprint readers, mobile phone, and so on). We are interested in the legal-normative approach of identity management, answering to questions such as how does the law on public registers' harmonization fit together with data protection legislation? Furthermore the Swiss Confederation and many Cantons have a law on transparency stating that all public documents and information must be publicly available, with exceptions for national security, trade secrets and citizen's privacy. Our goal is not to provide legal answers to these questions; it is rather to develop a framework that allows its users to describe requirements in terms of data, processes and identities.

[7] defines digital identity as a set of claims made by one digital subject about itself of another digital subject, with a digital subject being a person or a thing and a claim being an assertion of the truth of something, e.g. "I am Paul and I am over 18", "I am Mary and I am married to John". [7] furthermore defines seven laws of identity that are now widely used:

- *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 which discloses the least 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.
- *Directed identity*: a universal ID system must support both "omnidirectional" identifiers for use by public entities and "unidirectional" identifiers for private entities, thus facilitating discovery while preventing unnecessary release of correlation handles.
- *Pluralism of operators*: a universal ID system must channel and enable the interworking of multiple identity technologies run by multiple identity providers.
- *Human integration*: the universal ID metasystem must define the human user as a component integrated through protected and unambiguous human-machine communications.
- *Consistent experience* across borders: a unifying ID metasystem must guarantee its users a simple, consistent experience while enabling separation of contexts through multiple operators and technologies.

According to his own words, [7] defined these laws with the goal of "giving Internet users a deep sense of safety, privacy and certainty about who they are relating to in cyberspace". The context of population registers managed by the public sector is different, but these laws are still very relevant, even if they need to be adapted in some cases.

To conclude this section on identity management we will mention several other important issues: authenticity, trust, accountability, privacy, usability and data life-cycle [8], locality (identities are situated in particular contexts and one may have different or overlapping identities attached to given contexts), reciprocity, mutual understanding principles [9]. We will not deal with all of them neither integrate all the laws of identities in our framework (details will be given in section 3), but we think it was important to at least mention them to give readers a broader view on the topic.

2.3 Modelling Tools

Our goal is to use a conceptual framework to analyse governance processes of population registers' data, i.e. a set of methods, techniques and tools allowing us to model stakeholders and processes, and to formalise relationships between actors, processes and data. In previous work we developed a framework to model and share knowledge within large organizations or public administrations [10]. Called MIMIK (Method and Instruments for Modelling Integrated Knowledge), it provides a method (a series of steps taken to reach a certain objective) and instruments in the sense of concrete or abstract tools intended to assist users in fulfilling a given objective. These instruments consist mainly of abstract models that represent reality symbolically, in terms of concepts and relationships, and that are implemented as diagrams, i.e. simplified and structured visual representations of concepts and relations. The starting point of MIMIK was the model theory approach developed by [11] who aimed at integrating process modelling and knowledge management, as well as the work of [12] on process modelling and workflow models. We later analyzed several modelling techniques with very different backgrounds in order to provide a practical approach: commercial methods that are tightly integrated with a modelling environment, such as ARIS Toolset, MEGA Process / MEGA Designer or ADONIS; open and standard modelling methods such as OSSAD or OPEN; methods for knowledge-based engineering such as MOKA or CommonKADS; language and notation for object-oriented software-intensive systems such as UML.

MIMIK consists of 8 types of diagrams, most of them being inspired or directly taken from existing modelling techniques. *Concept maps* are the top-level diagrams and show the strategic goals of an organization in terms of functions or processes and their relationships. *Context diagrams* are almost exactly the same as use cases in UML, with the addition of the concept of knowledge packet, i.e. an abstract representation of a set of knowledge components. These components encapsulate documents, databases, files, implicit knowledge, etc. Knowledge packets show at the abstract level what type of knowledge is necessary in order to complete a process and which knowledge is relevant in a given context. MIMIK also provides two knowledge matrices, in order to show formal links between knowledge and processes (*Knowledge-interaction matrix*) and knowledge and actors (*Knowledge-actor matrix*). *Actor-role diagrams* can be classical organizational charts, graphical actors-roles diagrams, or matrices that formally link actors and roles in cases where the organization is too complex to be shown graphically in an intelligible way. Finally MIMIK uses *RDF schemas* to represent know-why, know-what and know-how, as well as *UML interaction* and *activity diagrams* to describe processes and activities. We decided to adapt this framework for our needs in terms of data governance, notably by integrating the requirements defined in section 2.1 and 2.2.

3. Building a Conceptual Model

We adapted the MIMIK framework by adding the following concepts or building blocks to model identities and data used when accessing population registers:

- *Data consumer*: actor requiring data from a register.
- *Data source*: register or database containing information on citizens.
- *Identity*: digital identity of the citizen in a given context.
- *Requirement*: expression of consumer's needs in terms of data on citizens.
- *Data set*: digital data on citizens provided by a data source.
- *Knowledge component*: meta-data encapsulating the content of context diagrams (Fig. 1) and formalizing the context of data exchange with the above-mentioned building blocks.

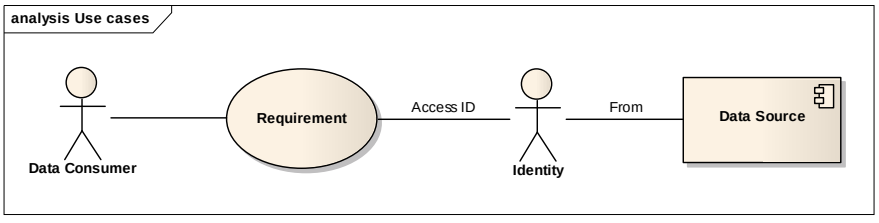

Figure 1: Generic Context Diagram

In the following sections we describe our building blocks; however this is a generic representation of reality, as Cantons might have different operational modes.

3.1 Data Consumers

Data consumers are the stakeholders that have a potential access to population registers; apart from some exceptions all of them are public or para-public offices, at various administrative levels and functional positions. Fig. 2 shows the main ones but is not exhaustive. Let us mention that not all actors listed here currently use that access, but they could potentially do it

once registers are harmonized and publish data. If we take the example of public kindergartens, where parents pay a fee that is proportional to their income, citizens bring a paper copy of their taxation decision and must justify their residence place. Although not all kindergartens have the same requirements, one could imagine that they could have a direct access to the fiscal register and the resident register in order to calculate fees.

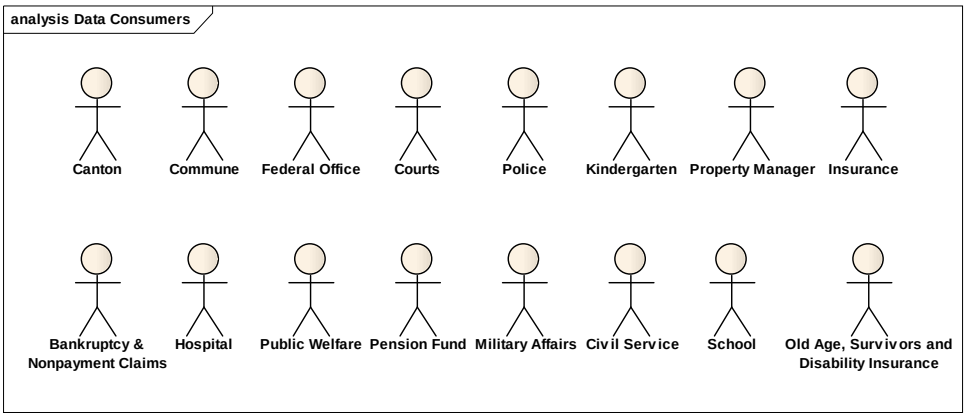

Figure 2: Data Consumers

3.2 Data Sources

As mentioned in the introduction there are official population registers as well as many other sources storing information on citizens. We list nine of them as potential data sources (Fig. 3), meaning that not all of them might currently be used as such, and even that some of them might never be opened to external access (e.g. military data).

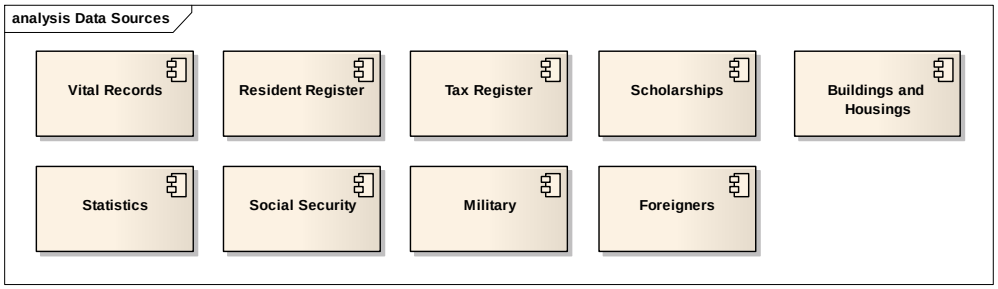

Figure 3: Data Sources

3.3 Digital Identities

Citizens have many roles in regard to public administrations and to the public and para-public sector: they pay taxes, they elect their representatives, they are required to take a basic health insurance, and they might benefit from social help or even go to prison, and so on. Once again the purpose of our identity model (Fig. 4) is not to be exhaustive or to be completely accurate, it is rather to show the variety of situations in which a citizen might find himself and the vast amounts of personal data that could be stored or used during these interactions.

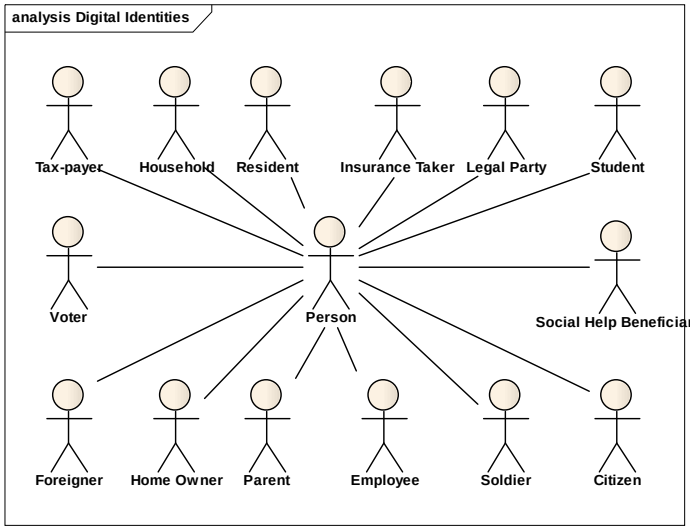

Figure 4: Digital Identities

3.4 Requirements

In Switzerland there are more than 2'000 administrative services delivered to citizens and businesses, from fishing and hunting licences to insurance subsidies or work permits. Of course not all of them rely on the use of data from population registers, but one can estimate than dozens if not hundreds of them use some kind of personal data. In order to describe them and to identify needs and requirements, we suggest the utilization of diagrams inspired by UML use cases.

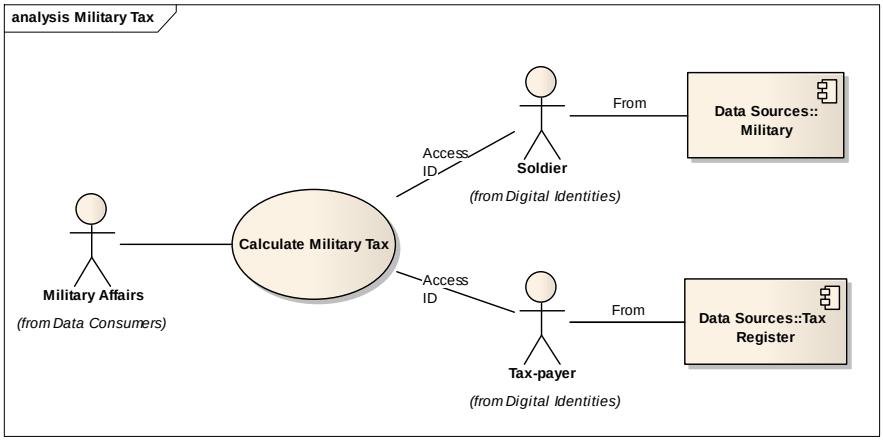

Figure 5: Military Tax

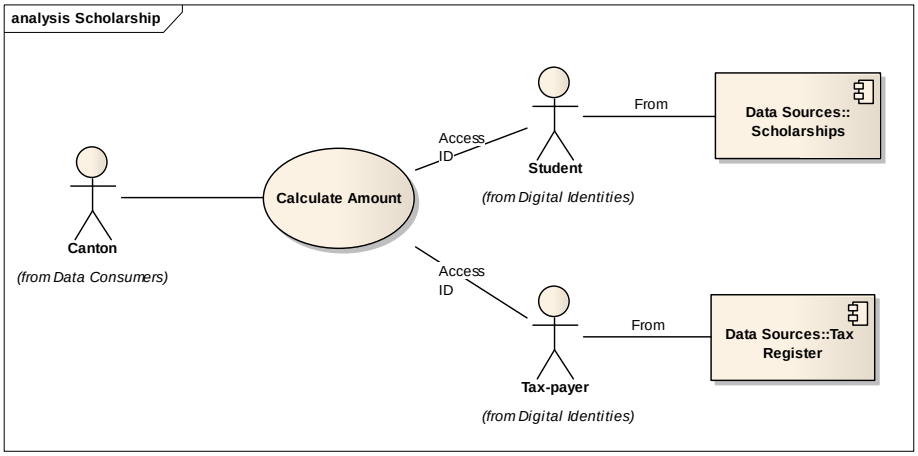

Figure 6: Scholarship's Amount

Fig. 5 shows an example where the military affairs department accesses two different registers in order to calculate the military tax. Indeed all Swiss male citizens are required to serve 300 days in the army and those who do not must pay a tax. This annual tax is calculated on the basis of the numbers of days that a citizen spend in the army (from 0 to 300), on the numbers of days he should have done at a given age and on his income. In order to do so it is necessary to match one's identity as a citizen, a soldier and a tax-payer.

Fig. 6 shows another example that is interesting in terms of identity management. Students have the right to request a grant for their studies under given conditions, such as the income of their parents, their age, etc. Here the notion of identity claim is central, as it is necessary to link a student to his/her parents, with some complex implications (adopted children, divorced parents, and so on).

3.5 Data Sets and Knowledge Components

Fig. 7 shows simplified data sets provided by data sources according to contexts and for given identities.

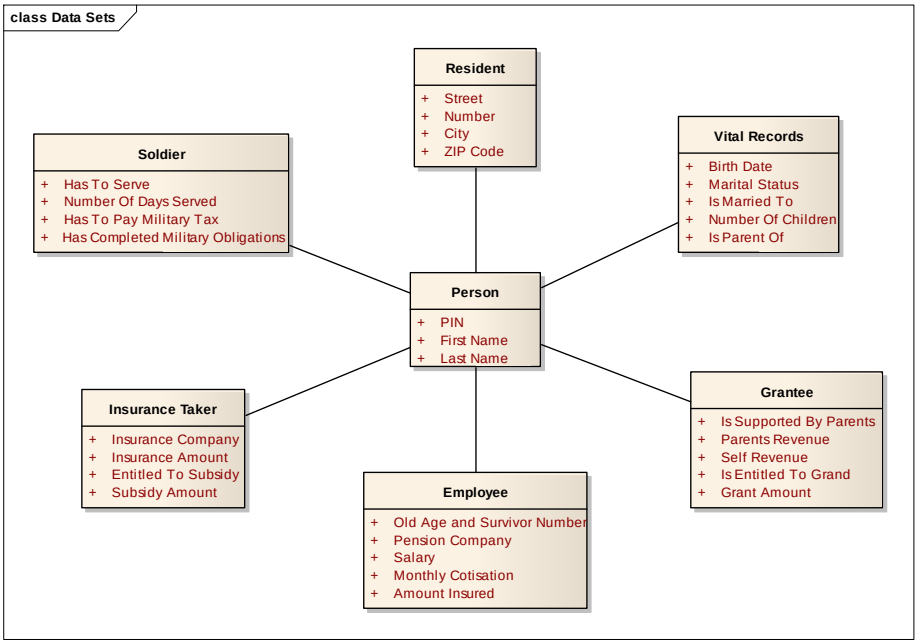

Figure 7: Selected Data Sets

Knowledge components are described in RDF (Resource Description Framework), a W3C standard for metadata description and machine-readable semantics in XML. Data components integrate information on the context (data consumers and sources), on identities as well as the data sets themselves. They also contain elements of data qualification regarding security and confidentiality (e.g. high/medium/low security requirements; confidential, restricted or public data).

4. Future Work and Conclusions

We are currently in discussion with various stakeholders (Cantons, communes, Swiss Federal Statistics Office, and other research centres) in order to launch a broad study on population registers, as well as on trade and company registers. We will use this framework to analyse needs and requirements of these various stakeholders in terms of data exchange and of registers' harmonization. We believe this project will offer a very interesting field of experimentation and validation for the approach described in this paper.

References

[1] BA Aubert, B Vandenbosch, M Mignerat, Towards the Measurement of Process Integration, CIRANO Working Paper, 2003, http://www.cirano.qc.ca/pdf/publication/2003s-04.pdf, last retrieved 18.02.2009

[2] UNDP, Governance for Sustainable Human Development, A UNDP Policy Paper, 1997, http://mirror.undp.org/magnet/policy/, last retrieved 18.02.2009

[3] TB Riley, E-Government vs. E-Governance: Examining the Difference in a Changing Public Sector Climate, The Commonwealth Secretariat and Government Telecommunications and Information Services, Public Works and Government Services, Ottawa, 2003

[4] IT Governance Institute, COBIT 4.1, 2007, http://itgi.org/cobit/, last retrieved 18.02.2009

[5] M Lips, JA Taylor, J Organ, Identity Management as Public Innovation: Looking Beyond ID Cards and Authentication systems, In: Bekkers, van Duivenboden, Thaens (Eds), ICT and Public Innovation: Assessing the Modernisation of Public Administration, Amsterdam: IOS Press: Amsterdam, 2006

[6] Genève, Projet de loi ouvrant un crédit d'investissement de 30 850 000 F pour le développement de l'administration en ligne, 2007, http://www.ge.ch/grandconseil/data/texte/PL10177.pdf, last retrieved 18.02.2009

[7] K Cameron, The Laws of Identity, 2005, http://www.identityblog.com/stories/2005/05/13/TheLawsOfIdentity.pdf, last retrieved 18.02.2009

[8] M Casassa Mont, P Bramhall, M Gittler, J Pato, O Rees, Identity Management: a Key e-Business Enabler, HP Laboratories Bristol, 2002, http://www.hpl.hp.com/techreports/2002/HPL-2002-164.pdf, last retrieved 18.02.2009

[9] G Roussos, D Peterson, U Patel, Mobile Identity Management: An Enacted View, International Journal of Electronic Commerce, 2003, Vol. 8, No. 1, pp. 81–100

[10] O Glassey, Method and Instruments for Modeling Integrated Knowledge, Knowledge and Process Management, 2008, Vol. 15 No 4, pp. 247-257

[11] B Wyssusek, M Schwartz, B Kremberg, F Baier, H Kralmann, Business Process Modeling as an Element of Knowledge Management - A Model Theory Approach. In Proc. Managing Knowledge 2001: Conversations and Critiques. Leicester University

[12] G. Mentzas, C. Halaris, S. Kavadias, Modelling business processes with workflow systems: an evaluation of alternative approaches, International Journal of Information Management, 2001, vol. 21