A strategy to gradual implementation of data interoperability

João Baptista Gonçalves, Binomial

joao.baptista.goncalves@gmail.com

Luisa Domingues, ISTAR, ISCTE-IUL

luisa.domingues@iscte.pt

Abstract.
Data interoperability is a major concern on e-government, both from the point of view of service offering and from the point of view of public administration efficiency. This paper purposes an incremental, pragmatic approach to data interoperability. It is argued that integration with minor required initial efforts from institutions is feasible, may provide useful solutions and is a solid ground basis for subsequent evolution. This paper presents general guidelines and model solutions to support this approach. Also, presents a demo implementation that proves feasibility of the purposed models and delivers useful solutions on a specific business e-government scenario. Although still limited in range and demonstrated on a quite specific business environment, it is expected that the analysis and the proposed strategies, solutions and models be of interest on a larger spectrum of data interoperability problems.

Keywords: Data interoperability, Taxonomy, Open data, E-government, DaaS, Data Services Catalogue

1 Introduction /Context

Access of citizens and enterprises to public administration services often involves the need to relate to various institutions or departments. This leads, frequently, to problems of redundancy, duplication of required data, etc. It is expected that electronic services contributes, quite significantly, to attenuate this problems. However, a lot of the traditional resilience of public administration services regarding coordination, normalization and integration of data and processes obstinately survives in electronic solutions.

Traditionally cooperation across public entities is an issue for (i) cultural reasons, i.e., lack of a Public Administration (PA) holistic view as a single entity which provides services to citizens. Instead each public entity performs a range of services in line with his mission acting as a stand-alone; (ii) organizational factors, i.e., bureaucratic organizations sometimes with highly politicized leaderships; and (iii) autonomous management.

Companies and citizens are, often, still forced to recurrently deliver the same information to different PA entities, and even in the same entity, to start different processes. The autonomous decision centers in each PA entity and the lack of a single ICT strategy for PA led to a technological landscape characterized by the existence of multiple and isolated information systems. Therefore, these information silos encumber data sharing between different public entities and even between organic units of the same entity. Moreover, the same concepts could have distinct meanings among different public entities, leading to different data.

This issue has been addressed over the last few years at policy and technical levels.

States and major public ICT institutions conduct efforts for that purpose. For instance, Europe encourages initiatives on data interoperability [1] [2] and prompt research projects in common taxonomy and ontologies, entitled as semantic web [3].

Data interoperability is major question on service integration. Services tend to define and treat data in their own terms. Therefore, often the same data appears differently across institutions, compromising exchange and reuse. Even only slightly differences in format, coding etc. are a major headache for data interoperability.

Besides meaning, data interoperability also traceability (who accesses to what information), data quality (validated by the competent authorities) and access profile. While in an internal circuit, the workflow allows to define which data the users can access, according to their working context. However it is hard to ensure this authentication when the access to the data is outside the boundaries of the workflow.

Whereas the intra-organizational data sharing has already taken significant steps with investments aiming IT integration [4], the lack of data sharing among public entities, raises a greater organizational, management and technology challenge. There are a widespread agreement that data sharing among public agencies would bring improvements in effectiveness (i.e., the level of service provided to economic agents and citizens) and efficiency (i.e., reducing the cost and time related with control and data validation).

The current state of technology offers several technical solutions to overcome data interoperability data (semantic web, XML, ..) [3]. Therefore the biggest challenge is not the technical view but the definition of cooperation and data management models.

This paper addresses the problem of data interoperability in PA. The main objective is to contribute to enforce cooperation between PA entities in order provide better service and improve efficiency.

PA entities often deal with limited resources. Projects oriented to data integration often require big effort and provide only long term results. This circumstances often discourage initiatives on this area.

This paper purposes a simple scalable model to data interoperability. It is argued that simple solutions, requiring minor initial effort might, in some circumstances, be a path to overcome this challenges.

A simple interoperability model is purposed for this matter.

Also, it is presented a system implementation targeted to specific actual business environment scenario where all the referred questions apply. The targeted scenario includes two main functionalities: (i) an enterprise that directs to a service (filling an application form, for example) might access and, if possible, reuse similar data available in the same public entity as well as in different public entities that relate themselves to provide this facility; and (ii) a public entity that is allowed to read information from another public entity about an enterprise.

Specific conditions that apply to this problem. The data involved is similar in concept, tough – not surprisingly - quite different in record organisation, fields, formats and codification.

The involved data is not public data, the problem of authorization must be considered. Every access must be authorized by the owner (i.e., the enterprise) and accesses must be tracked.

The paper is organized as follows: Section 2 describes the research methodology. The approach used is based on a lot or published work concerning the general problem of "data interoperability" and related areas and technological concepts, this is reviewed in section 3. Section 4 presents strategic guidelines to deal with the problem and an architectural operational solution to implement those same guidelines. The section 5 presents a developed demo, currently implemented for evaluation. Finally, Section 6 presents conclusions and proposes work evolution.

2 Research Methodology

The Design Science Research Methodology (DSRM) [5] was considered suitable in order to support this research. DSRM incorporates principles, practices, and procedures required to carry out research in information systems. This research methodology meets three objectives: it is consistent with prior literature, it provides a nominal process model for doing design science (DS) research, and it provides a mental model for presenting and evaluating DS research in information systems [5]. Selecting DSRM ensures the existence of a set of activities that underpin the coherence between the practical application and the principles and strategies defined for the proposal interoperability model. The DSRM includes six activities are: problem identification and motivation, definition of the objectives for a solution, design and development, demonstration, evaluation, and communication.

We define these activities in the context of our research as follows:

1. Problem identification and motivation. This research is driven by the need to overcoming of so-called eBarriers in the exchange of documents and information among public entities in order to deliver efficient and integrated electronic public services.
2. Objectives of the solution. The objectives of this research are:
   (a) To propose a model following a pragmatic approach which enables a smooth integration and semantic interconnection among public entities, without requiring a large initial effort or a disruption in the semantic data or technologies used; and

(b) To demonstrate in a very restricted but real environment, the feasibility of the proposed model principles.

The objectives will be realized through:

(a) Identification of the strategies and principles to support an incremental model;
(b) The definition of an interoperability model;
(c) The definition of an ontology applied to a specific and narrow environment for demonstration purpose.

3. Design and development. We first analyze the existing initiatives to overcome the challenge of interoperability in general, and in the particular context of public administration (Section 3). In a second step we defined the strategy for the model implementation. Based on these strategy guidelines were specified the structure of business and usability metadata to achieve machine-readable representations of data contents (Section 4). Also were defined levels of service for each data component, in order to assure an incremental adherence.
4. Demonstration. We defined an ontology, named *"dip"* (data interoperability protocol) focused in a very narrow and specific usage context, i.e. the process to support co-financing projects submitted by enterprises, to demonstrate the model implementation.
5. Evaluation. The demo system was tested in the real context of a public entity, although in a test environment. The solution demonstrated versatility in terms of adaptation to different types of contents and does not represent a relevant change or efforts in the solution implementation.
6. Communication. We publish this paper to share our experience.

3 Conceptual research background - Issues and approaches

In the last years, many efforts were directed towards enabling interoperable information systems through consistency and uniformity in the way that information is described, stored and retrieved, especially in complex organisations such as governments [6]. Therefore, many approaches, architectures, and protocols were proposed in order to make open data more machine-readable, and interoperable [7].

The Service Oriented Architectures (SOA) approach and the widespread use of Web Services, brought flexibility and interoperability to data integration through a class of web services, called Data Services [8], that access and query data sources. In a basic data service usage scenario, the owners of a data store enable web clients and other applications to access their, otherwise externally inaccessible, data by publishing a set of data services [8].

Defining the semantics of data services is a significant driver to enable data interoperability. An interesting approach to define the semantics of data services is by describing them as semantic views over a domain ontology [7].

A semantic interoperability asset is as a collection of reference metadata elements that sharing them among governments would contribute to increased interoperability across organizational and geographic boundaries [6]. Metadata, defined as information about data, identify the structure and meaning of data. According to Ralph Kimball [9] metadata elements can be classified into technical and business metadata. The technical metadata specifies how exactly the data is structured and stored in files or databases, in order to allow applications and tools access and manipulate it. The business data, expressed through business requirements, time-lines, business metrics, business process flows, and business terminology, help to understand the data and their usage requires extensive and in depth understanding of business entities, tasks, rules and the environment.

Many countries worldwide have defined metadata frameworks as part of their national e-Government strategies [4] [6]. Also, in Europe as part of the Digital Agenda for Europe, one of the seven flagship initiatives in the Europe 2020 Strategy for smart, sustainable and inclusive economic growth, the European Commission has adopted the Communication 'Towards interoperability for European public services', which aims to establish a common approach to effective interoperability among European public administrations [1]. In this context governments as well as the European Commission are sharing their metadata on the Web to encourage their reusability and consequently facilitate interoperability. This has led to a new kind of repositories focusing primarily on semantic interoperability assets, such as Digitaliser.dk in Denmark, the ESD toolkit standards lists in the UK and the European Union repository SEMIC.EU [6].

The SEMIC.EU platform promotes semantic interoperability among European Member states by collecting, evaluating, indexing and making available a large number of semantic assets from a single point of access. In this way, developers can easily discover and reuse assets like data models, taxonomies, codelists and vocabularies developed by others facing similar use case.

The Asset Description Metadata Schema (ADMS) is a standardized metadata vocabulary created by the EU's Interoperability Solutions for European Public Administrations (ISA) Programme of the European Commission to help publishers of standards document what their standards are about (their name, their status, theme, version, etc.) and where they can be found on the Web. ADMS descriptions can then be published on different websites while the standard itself remains on the website of its publisher (i.e. syndication of content). ADMS embraces the multi-publisher environment and, at the same time, it provides the means for the creation of aggregated catalogues of standards and single points of access to them based on ADMS descriptions.

Following a similar trend to vocabularies used as metadata on the web, definitions should first be agreed on fundamental concepts, where diverged and/or conflicting views can be handled. These concepts are defined as Core Concepts [10]. A Core Concept is a simplified data model that captures the minimal, global characteristics/attributes of an entity in a generic, country and domain neutral fashion. It can be represented as Core Vocabulary using different formalisms (e.g. XML, RDF, JSON). The Core Vocabularies are general semantic building blocks that can be extended into context-specific data models. Four Core Vocabularies were created: (i) Core Person which captures the fundamental characteristics of a person, e.g. the name, the gender, the date of birth, the location; (ii) Registered organisation:, that captures the fundamental characteristics of a legal entity, e.g. its identifier, activities, which is created through a formal registration process, typically in a national or regional register; (iii) Core Location which captures the fundamental characteristics of a location, represented as an address, a geographic name or geometry; and (iv) Core Public service that captures the fundamental characteristics of a service offered by public administration.

A platform of e-Government Core Vocabularies hosted by European Commission are supported by W3C [11] promoting enormous visibility worldwide.

Nevertheless, all these efforts include only generic properties that are not sufficient for completely fulfilling the needs of the diverse audience of government semantic interoperability assets, e.g., developers interested in ontologies or codelists or project managers interested in UML diagrams or reference datasets. RDF-based models use the Resource Description Framework (RDF) as a data model. Semantic Web [3]and Linked Data technologies have been applied to many e-Government catalogues and repositories to achieve machine-readable representations of their content metadata using RDF. The adoption of such technologies has several benefits like decentralized publishing and Web accessibility.

4 Proposed guidelines and solutions – incremental model

4.1 Strategy

Interoperability requires common semantic and formats. An ideal approach to develop an interoperability solution is to promote some kind of consortium where one or more entities engaged themselves on a project to develop common models, adapt their data to these models and develop methods to data exchange.

However, this ideal approach may suffer a few drawbacks. It requires organisations commitment in the project, involving budget, resources and the willingness to change and adapt, which is often difficult to mobilize all together. It also requires a huge initial effort before results arise, which may be discouraging. In addition it closes the door to entities not engaged in the project, thus compromising the adhesion of new entities.

When these drawbacks prevails, a more gradually based approach can introduce some features in order to leverage an interoperable solution. The keynote of this approach is the principle that each entity contribution is fitted to its constraints and capabilities. Therefore, the data provider organisations offer data in the best possible way, regarding interoperability. As data client, organisations are willing to receive data in any available format. In such context each organisation defines its own pace depending on its know-how and available resources.

In order to support this pragmatic and incremental approach a set of guidelines are presented:

- Definition of interoperability service level. A low level indicates that the provider organisation is only able to provide data in hermetic formats. While a high level demonstrates the ability to provide structured data according to semantic annotations, enabling a large potential of interoperability.

- An evolutionary path background. The data may be provided at a low interoperability level in a first stage, and later, may evolve to higher levels of interoperability.

The higher levels of interoperability requires the usage of common semantics. A wide range of ontologies already exist, from those that apply to a generic scope (example Dublin Core [12]), to those which are focused into particular businesses or institutions requirements (example ISA – e-Government Core Vocabulary [10]). However, specialization of ontologies in particular business areas, is encouraged [13]. Therefore it is reasonable accept that in a near future will arise shelf-ready semantics with metadata definition to cover business specific requirements. Nevertheless, the use of the most widespread ontologies whenever possible should be appreciated even if mixed with more specific ontologies.

In the following sections a model is proposed to fulfill these guidelines.

The model defines a few rules to system interconnection and data exchange, starting from scratch and leaving a ground basis for evolution.

4.2 Interoperability model

Considerer two systems A and B. System B possesses data that might be useful to activity A. A and B operate in the same business area, therefore manipulating similar business data. However, they do not have previous effort on data integration or data standardization. Supposing A is a client which, for his activities, is interested to get data that B can supply in usable form. The model describes (i) the way B supply data and (ii) the way A acts in order to discover available data and retrieve it in usable forms.

For the sake of reasoning, we considered a targeted business scenario where queries concerns available data on a given entity.

In order to get data from system B, A issues a query identifying the entity, and system B responds presenting a catalog of available data on that entity. The catalogue of available data must include pertinent information for automatic data processing.

Information is organized in *chuncks* or *"records"*. Each record is a piece of information identified with a unique invariant *id*, issued by the server. The record can be retrieved using that identification. Despite the *id* be invariable, the information obtained based on it depends on the information available at the moment it is requested.

Concerning automatic processing purposes, each record should be enriched with business and technical metadata [9].

The business metadata includes: (i) basic information, such as description, classification type, and contents identification relevant to the business; and (ii) data management information, covering data life cycle features like retrieval date, origin/source, and validation status (assuring whether validated by public entities).

The technical metadata, referred as usability metadata, should include relevant information for automatic processing of retrieved data. For this purposes three service levels were defined:

- Level 1 – document;
- Level 2 – structured data;

— Level 3 – semantic data.

Level 1 applies if data is available as a document, a pdf for example.

The notation *"document"* in this context means access to data without possibility of content automatic treatment. This is a very low level interoperability stage, nevertheless, not absolutely useless. It allows download a document with a requested information. The same interoperability level can be reached if data is available in a manageable format retrieved by a third party application. This means that client, although not capable of access and manipulate data by itself, knows an application that may be used to that purpose. Associated metadata identifies the specific document format, allowing manipulation at application level.

Level 2 applies if data is retrieved *"as raw data"*, say as a XML file.

This allows a client application to reuse this data, if meaning is known, thus requesting specific development effort to cope with origin formats.

Associated metadata identifies formats, including data models if available. Typical scenario is a record formatted as XML, along with a XSD specification. This allows the client system to develop specific tools to handle data, say a HTML page to view data or a program to extract data into a database.

Level 3 applies if data is available along with pertinent semantic information.

For this purpose, the server system informs ontologies that apply to retrievable data. This may be proprietary, defined by a public entity, as well as standard, widely known ontologies. The key point is that no specific ontology is adopted. The data provider system just inform what ontologies it can apply to facilitate data reuse. This allows a client system to automatically extract data applying the used ontology. Whereas specific development effort to process non-standard data structures would be required if the data were provided according to the service level 2, in this level automatic processing is possible.

Naturally, the use of a standard and widely known ontologies is strongly advised. It seems consensual [1] [6] that using consolidated, independent and widespread ontologies may be a real step towards data interoperability.

The incremental approach proposed has as its ultimate goal, to achieve the provision of data using the existing and widespread ontologies aiming to increase the availability and usability of interoperable data.

5 System description

5.1 Functional description

A demo system, named DIP (data interoperability protocol) was developed to implement as proof of concept for the model.

The system is targeted to the following actual problem. The public administration implements several initiatives on co-financing of projects submitted by enterprises. The implementation of this mechanisms often involves the need for submission of forms by enterprises along the several stages of the process, carrying information like identification, contacts, enterprise activity data, and so on. Often, this same information is repeated on applications on different institutions, if not on different departments of the same institution. Support for data interoperability facilitates the implementation of data reuse mechanisms, minimizing duplication and redundancy.

On the other hand, institutions themselves are interested, and this right is granted by law, on accessing information that the enterprises provided to other public entity, for auditing and general information purposes.

Within this context, DIP targets two main application contexts. An institutional context that implements data access between public entities. In this context it is allowed that users from one institution, say A, accesses information that another institution B possesses about a given enterprise.

A second context implements a granted authorization by the enterprise user, to access own data in a public entity system. Thereby, on behalf of an enterprise user, is implemented the possibility of a public entity, owner of the system A, gather information from an external system B, for reuse in its own operation. For example filling an application form, on system A which requires information from external system B. It should be noted that this information is not public data. So each access must be supported on proper authorization. Institutional access is supported by law, however individual usage must be granted by the data owner.

The current versions of the system conveys the following record types:

- Entity description: name, address, activity, contacts, etc.;
- Balance sheet organized by year, both actual and predicted, in pdf file or structured data);
- Company history, in text format;
- Products and sales figures coded.

Although limited in extension, this set of record types include a wide range of data formats, including documents and structured data. Notice that similar data may occur in different formats, and therefore provided using different levels of service. For example to provide balance sheets, data may be available as pdf document or as XML data.

5.2 System interconnection

System interconnection for entity crossover is straightforward. This type of interconnection requires previous setup at administrative level. Following this procedures, a specific cross connection is implemented between the systems.

On the other hand, individual reuse of data, should be possible with minor previous integration effort between organisations. It should be possible that a user at system A retrieves and reuses information at system B with no previous knowledge of the foreign system, taken that both convey to the basic DIP protocol. In this circumstances it should not only be possible to retrieve information in raw sense but also, if systems agree on common ontologies, process this information automatically.

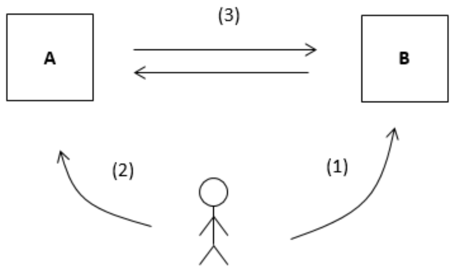

Fig. 1. System interconnection

For this purpose it is used a *"loosely coupled"* model for the system interconnection suggested in Fig. 1. Say a user in system A wants to access data in system B. The protocol goes like this:

(1) User issues a "credential" at system B.

(2) User registers this same "credential" at system A.

(3) System A issues a data request to system B, using the credential obtained in the step (2).

5.3 Query model

An issued query takes the format of a URI: *http://(systemB)/dip/entity-list#id*, meaning "what information is available about entity Id".

The query access is certified by the credential, which was previously issued by system B. The data provider system responds with a catalog of available records on that entity. Each record is identified by a unique *id*. Retrieval for a specific record is available on an URI *http://(systemB)/dip/entity-record#rec-id.*

5.4 Business metadata

Business metadata describes essential functional information. We developed a small utility business ontology and taxonomy to business domain, named *"dan"* from *"data annotation"*, including key aspects of business functionality.

The contents classification is carried out under the previously defined record types (topic 5.1). Each record is classified according to the following structure:

```
<dip:record-description about=’’id’’>
<dan:type>dan:Identification</dan:type>
```

Where *dip* tags the general name space of the interoperability proposed model and *dan* tags the business specific model. The *dan:type* tags record types value, which includes BalanceSheet, History and so on. In order to specify particular information, additional properties are defined depending of the record type. For instance if the record type is BalanceSheet the property year will be defined.

The *dan* namespace include fields for: origin identification, which describes how institution acquired data; date acquisition, representing when data was retrieved by organisation; and date validity, when data applies. Data origin is qualified as: “EntityInformation”- submitted by the entity as information; “EntityDeclaration” - submitted by the entity as formal declaration imposed by legal obligation; “PublicAdmin”- delivered from public administration services; or “ThirdPart”. The data origin field has an additional property defining the “administrative quality” of data. The possible values for this property are: “Information” – the entity delivered data to public administration without commitment; “Declarative” – the entity delivered the information by legal imposition; or “Verified” – the data is endorsed by public administration.

This is an extension of record classification, intended only for demonstration purposes:

```
<dip:record-description about=’’id’’>
<dip:classification>dan</>
<dan:type>dan:Identification</dan:type>
<dan:origin>dan:EntitieDeclaration</dan:type>
<dan:date>2016-03-23</dan:date>
<dan:local-metadata>
  <dan:local-metadata-property>origin</>
  <dan:local-metadata-property-value>(application form on
call X)</>
  <dan:local-metadata-property-descritpion>(text descrip-
tion of property)</>
</dan:local-metadata>
<dan:local-metadata>
  <dan:local-metadata-property>contents</>
  <dan:local-metadata-property-value>partners </>
  <dan:local-metadata-property-descritpion>(includes en-
tities partners and group relationships)</>
</dan:local-metadata>
```

The use of *dip* ontology is required for protocol usage. Though the use of a specific business ontology is not required. Nevertheless, it is not expected that ontologies exists for every specific business environment and, on the other hand, it is difficult that general and abstract ontologies fulfil all the requirements of specific business areas.

However, public administration as a whole applies general rules of procedure and it might be possible that common vocabularies arise within consortiums of public entities, at national level or even at multinational levels [2]. If this happens, use of these ontologies is clearly preferable, as an alternative or complement.

The implemented system demonstrates the use of this business metadata. This is crucial mainly in the institutional context where users search for available information about a given entity. Metadata provides clues about each available records, and the user then proceeds, from that point, selecting the elements he finds of interest.

5.5 Level 1 implementation

As indicated in section 4, additionally to business metadata, each record is described in terms of interoperability, i.e. functional metadata. First item of this description is the “level of service” (described in topic 4.2.).

Service level 1 indicates that the information is available as a document. Additional required information states if the document is provided by an application, allowing the client system to use it in the best possible way. The specification has a more functional intent than usual “myme type”. At this point a simple taxonomy is used to qualify contents as “spreadsheet”, “pdf”, etc.

Actually, we considered this an area of future potential to work. It happens that in the context of the targeted specific business area there is a limited range of document formats. Although documents are generally closed to data extraction, they usually follow organizational standards. This might be exploited. There exists technological tools to deals with document contents: spread sheets, formatted text and even pdf. There is not, to our best knowledge, conceptual mechanisms to model internal contents. It seems possible to develop specific modeling of internal document contents thus depicting a way of extracting this information.

The implemented system uses level 1 services both at institutional and enterprise contexts.

At institutional level a general protocol for information viewing was implemented, using document classification type “html”. This means the server provides an html file that might be “mounted” in a bowser or java application, for instance. This allows a client system to get a record in html and use it to directly show it in browser applications.

Actually, this constitutes a really useful functionality in the context of targeted business. Institutional access mainly deals with remote data consultation. It happens that institutions manipulates a lots of structured data, similar in concepts, but in different formats, codification, and so on. Although useful, it is a long process to proper classify all this different data structures in order to make them available according to a semantic classification. The simple possibility of having this information available, even only for viewing, in the local system with minor development effort is major breakthrough.

Enterprise interface uses this application level just for demonstration purposes, allowing the attachment a remote downloaded document when filling a local application form.

5.6 Level 2 implementation

Level 2 implementation indicates that information is retrievable as structured data. The proposed *dip* classification indicates that further specification of data formats and data models are available. This earlier version of system implementation only uses XML and XSD files for that matter.

So, a record committed to services level 2 is enriched with the following properties:

```
<dip:record about=id>
<dip:service-level>2<>
```

```
<dip:service-level-description>stuctured-data<>
<dip:data-format>xml<>
<dip:data-model>xsd<>
```

This service level allows data to be processed in a model-dependent way. Client system must have previous knowledge of data semantics and develop specific procedures to benefit from data. Therefore, specific development effort is required.

System uses this model to implement a specific client/server interconnection between two systems, demonstrating import of remote data, available in XML, when filling a local application form. This requires specific development to convert and transfer data.

5.7 Level 3 implementation

Level 3 implementation indicates that information is retrievable as structured data enriched with semantic information. The proposed *dip* classification indicates the specific ontologies applied.

System implementation on level 3 was developed just for record identification, which is the most general and structured record type handled by the system. In order to illustrate and demonstrate the proposed ideas, two different semantic models are used, one specially developed for this purpose and other using standard, well established and close-oriented to the target business environment.

A typical identification record includes fields like:

```
<id>….</id>
<name>…..</name>
<address>….</address>
```

Actually, there are several variations of these fields: different field list, field names, and so on. Implementation uses a specific business model, named *"eid"* and a widespread and well established ontology. The "E-Government Core Vocabulary" [10] was the ontology selected because it is closed to business environment – Core registered organisation.

A record deliverable in service level 3 is thus enriched with the information suggested by the following:

```
<dip:record about=id>
<dip:service-level>3</dip:service-level>
<dip:service-level-description>semantic-data</>
<dip:data-format>xml</>
<dip:data-model>xsd</>
<dip:semantic-annotation>(system)/eid</>
<dip:semantic-annotation>(core)</dip:semantic-annotation>
```

The data semantic is available in proprietary XML as well as in the named ontologies. Correspondent data is then noted as suggest d in the following:

```
<eid:id>….</edi:id>
<eid:name>…..</eid:name>
<eid:address>….</eid:address>
<core:LegalEntityIdentifier >….</core:LegalEntityIdenti-
fier>
<core:LegalEntityLegalName >….</core: LegalEntityLegal-
Name>
<core:LegalEntityRegisteredAddress >….</core: LegalEnti-
tyRegisteredAddress >
```

This level of service is exploited in user interface. Within the process of filling a form application user may indicate an external system that is assumed to implement the *dip* protocol. The system is then queried for information available on a given entity. If data, specifically an identity record, is located with a recognized ontology classification, data is imported to the application form.

6 Conclusion

At time of paper writing the system described in section 5 is implemented for evaluation purposes. The implementations makes clear that these ideas are feasible. Also demonstrates that it is possible to achieve useful results in data interoperability with minor development effort and with quite common, open source, technologies.

Furthermore, even mainly oriented to concept proof, the developed system is targeted in real environments and ready to be deployed in production environment after evaluation.

Meanwhile, the developed application is an earlier implementation and a lot of work is still do be done, both at conceptual, modeling and implementation level.

A few questions were pointed throughout the paper. The classification of contents at document level exchange, may be a major point of interest. For this purpose it requires minor effort on behalf of organisations to deliver documents and it may prove to be useful to data discovery and high level data manipulation.

The important question, however, is the further development of the purposed strategy for data interoperability and the associated models. The proposed model, being an earlier formulation, requires actual development and consolidation. It requires generalizations to cope with a more wide range of interoperability mechanisms. It requires additional operational skills, like service and data catalog discovery. Additionally it requires more elaborated models to cope with evolution from lower to higher level of interoperability and from specific to common and widespread ontologies.

References

[1] Interoperability solutions for European public administrations (ISA) - European Commission, "European Interoperability Framework (EIF) - Towards Interoperability for European Public Services," 2011. [Online]. Available: http://ec.europa.eu/isa.

[2] Interoperability solutions for European public administrations (ISA) - European Commission, "Federation of Semantic Assets Repositories - The SEMIC.EU approach," May 2011. [Online]. Available: http://ec.europa.eu/isa.

[3] W3C, "Semantic Web," 2016. [Online]. Available: https://www.w3.org/standards/semanticweb/. [Accessed 2016].

[4] U. Nations, "United Nations e-Government Survey 2014 - E-Government for the future we want," United Nations Division for Public Administration and Development Management, New York, 2014.

[5] K. Peffers, T. Tuunanen, M. A. Rothenberger and S. Chatterjee, "A Design Science Research Methodology for Information Systems Research," *Journal of Management Information Systems,* vol. 24, no. 3, Winter 2007-8.

[6] G. Shukair, N. Loutas, V. Peristeras and S. Sklarß, "Towards semantically interoperable metadata repositories: The Asset Description Metadata Schema," *Computers in Industry,* vol. 64, no. 1, pp. 10-18, January, 2013.

[7] A. Malki, M. Barhamgi, S.-M. Benslimane, D. Benslimane and M. Malki, "Composing Data Services with Uncertain Semantics," *IEEE Transactions on Knowledge and Data Engineering,* vol. 27, no. 4, pp. 936-949, April, 2015.

[8] M. J. Carey, N. Onose and M. Petropoulos, "Data Services," *Communications of the ACM,* vol. 55, no. 6, pp. 86-97, June, 2012.

[9] R. Kimball and M. Ross, The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling (3rd ed.), Wiley, 2013.

[10] Interoperability solutions for European public administrations (ISA) - European Commission, "eGovernment Core Vocabularies," 5 May 2011. [Online]. Available: https://joinup.ec.europa.eu/sites/default/files/c1/23/4a/egovernment-core-vocabularies.pdf.

[11] W3C, "Registered Organization Vocabulary," W3C Working Group, 01 August 2013. [Online]. Available: https://www.w3.org/TR/vocab-regorg/.

[12] European Union, "e-Government Core Vocabularies handbook - Using horizontal data standards for promoting interoperability," 2015. [Online]. Available: http://ec.europa.eu/social/publications.