/// INTEROPERABILITY BETWEEN METADATA STANDARDS: A REFERENCE IMPLEMENTATION FOR METADATA CATALOGUES

/// Keywords: Public administration, Metadata, Geographical Information, Geospatial data, Open Data, ISO, DCAT-AP, GeoDCAT-AP, CKAN, Open Standards for Linked Organizations

1   ABSTRACT

To enable the public and private sector to discover, adopt and reuse government information, administrations publish their data on data portals. The data is accompanied by structural metadata, providing information about the datasets. Governments publish information from different domains, including Geospatial Data, Open Data, Statistical Data, Archival Information, which is causing a wide variety of metadata standards. As these metadata standards often are not interoperable, it is a complex task for government administrations to publish their data in line with the regulations in the different data domains. This position paper reviews a potential strategy to simplify the management and reduce the costs of metadata portals.

2   INTRODUCTION

How can data providers keep it simple to describe and publish their data, with the 'once only' principle in mind, reaching the widest possible audience, and being at the same time in line with all regulations applicable to their data domain?

I There are plenty of data domains: Geospatial data, Open government data, Statistical data, Archival information, Documents, ... All those 'niches' have their directives, regulations, best practices and guidelines. When data is covering more than one of those niches, it is very hard to be compliant with the regulations of all the crossed domains.

Organic evolution of eGov initiatives in Europe has led to the present situation where data portals co-exist: a geoportal (for geographic datasets, dataset series, services and applications) and a portal for Open

government data, services and applications. So what about Geodata with a regulated open data licence or Open data with a geo component?

There are a wide variety of metadata standards and profiles, so how can the semantic differences be solved in transcending the silos of data domains, in order to keep it simple for data providers to describe their metadata only once, and reaching at the same time the widest possible audience by publishing on various portals?

On short terms, the minimal solution is to develop mappings, translations, from one metadata profile into another. In general, the mapping is uni-directional from the more specific niche metadata profile into the more basic / core metadata standard. Taking into account the loss of metadata elements by adding the core elements of the niche profile, translated into the language of the simple profile, in its storage format, ...

On long terms, the better solution would be to reconsider, possibly even reshape one (or both) standards, so that the differences which data providers encounter are converging one another. But therefore the standardisation organisations / regulatory bodies have to come together, each with the experts in its niche, and try to gather shared insights on semantics and standardisation.

3 OBJECTIVES

To keep it simple and 'once only' for data providers to describe and publish their public sector information on the web, we need to solve the semantic differences in the diverse metadata standards and profiles. This enables data providers to reach a vast audience, and at the same time aligning to all regulations and guidelines applicable to their data domain, describing information once only.

To find out how this could be achieved, a comparative study was initiated by the Flemish Government. Its aim is to compare both the Geographic and the General Open government data approach towards their respective metadata solutions. This research focuses on three levels: standards, metadata management systems and portals.

The results of the study are crucial with regards to the simplification of the current solutions towards a more cost effective and simplified future implementation. Also, we hope that by closing the metadata gap between the Geographic and the Open government data world, we will clear the road for incorporating other data domains, e.g. statistical data or documents and archival information.

Since simplification is an outcome of the study, the next step is to implement those in a pilot. Because only that way, the real impact, effectiveness, automation, ... will be proven applicable.

The results can help administrations in Europe in more efficiently, cost saving, more straightforward, … dealing with the same metadata standards, systems and portals.

This proposal helps data providers to publish their information not only conform their 'niche' metadata profiles (e.g. ISO and INSPIRE), but as well into the frequently used standards (e.g. DCAT) used in data portals.

Semantic differences will be solved, e.g. the ISA Core Vocabularies (including Core Public Service Vocabulary, Core Public Organization Vocabulary, Core Location Vocabulary, …) that are used to complete the metadata elements.

//

But to really close the gap, the semantic differences should be picked up to solve by the standardization bodies itself (W3C, OGC, JRC, ISA, ISO). All possible standards have the same basic semantics: the Dublin Core. Every silo created their own specialisms to describe their data into. And we really need to find the common denominator, in order to simplify; and to come to a harmonized strategy on metadata catalogues and reusable reference implementation. This means not to create another standard, but trying to solve the semantic differences in at least the exchange between standards; but preferable by changing those elements that are semantically mismatched in the standards (e.g. ISO, DCAT, …) themselves.

Finally the outcome is a harmonized strategy on Metadata Catalogues and a reusable reference implementation.

4 SCOPE

Initiatives on metadata: How to solve the semantics in transcending the silos:

1. Comparative study to close the metadata gap between the Geographic and the General Open government data world
2. Implement a pilot to prove the effectiveness, automation, … of the proposed solution
3. Sharing knowledge with other member states
4. Gather feedback on the uptake and effectiveness of the implementation
5. Continuation of the work in the pilot with other data domains (e.g. Archival information)
6. A harmonized strategy on Metadata Catalogues and a reusable reference implementation
7. Publish and promote a Best Practices paper
8. Promote the results at meetings and partner events (e.g. INSPIRE, ISA², W3C, OGC, Semantic Web Conference, …)

5 PROBLEM STATEMENT

How can you keep it simple for data providers, the 'once only' principle in mind, reaching a vast audience, and at the same time aligning to all regulations and guidelines applicable for their data domain, describing information once only.

In the present situation there are plenty of data domains: Geospatial data, General Open government data, Statistical data, Archival information, Documents, ... All those niches have their directives, regulations, best practices and guidelines. When data is covering more than one of those niches, it is very hard to be compliant to the regulations of all the crossed domains.

There are so many metadata standards and profiles, how to solve the semantics in transcending the silos (see figure 1 below)?

This problem must be tackled at the highest possible level, by the regulatory bodies (ISA, INSPIRE, OGC, W3C, …) towards a harmonized strategy on Metadata Catalogues and a reusable reference implementation. So that

//

A HARMONIZED STRATEGY ON METADATA CATALOGUES AND A REUSABLE REFERENCE IMPLEMENTATION /// 7.10.16 3 /// 6

all countries can work in accordance to the outcome of this proposal, to be interoperable at EU and even global level, conform the EIRA principles (see figures 2 and 3).

On the other hand, more and more portals arising on the web to provide information to the end-user. Sometimes arrangements in harvesting are made from one portal into the other, but often the information is collected multiple times with doubles as a consequence.
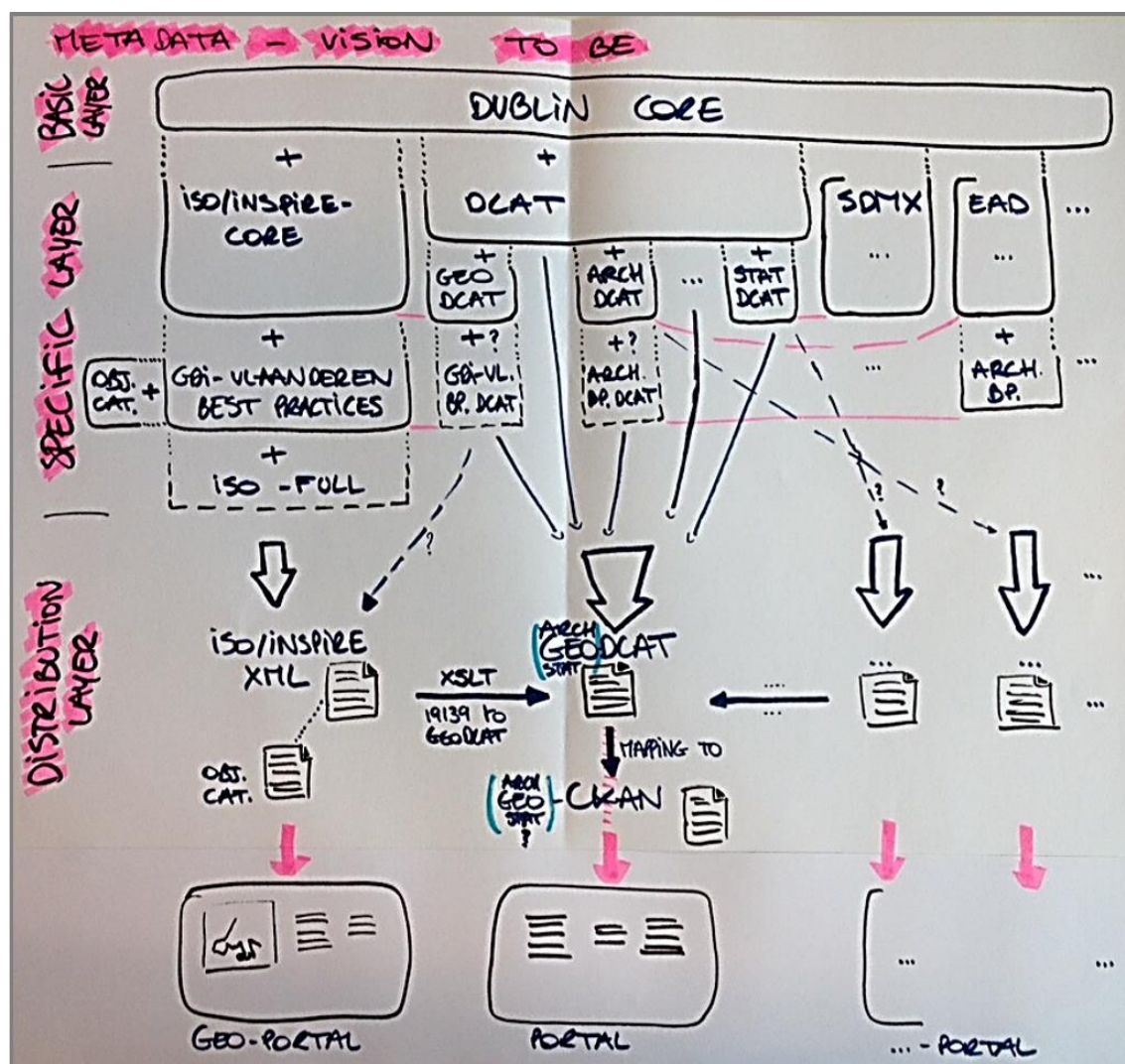

Fig 1. Metadata Standards for Geospatial, Open government, Statistical and Archival Information (Geraldine Nolf, Informatie Vlaanderen)
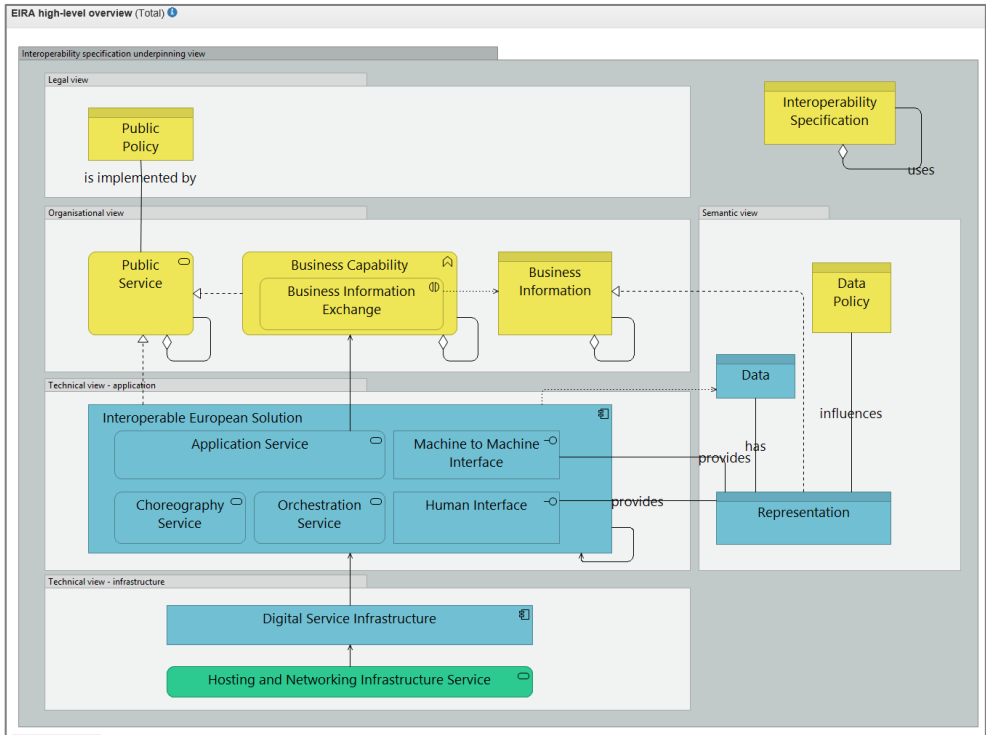
//

4 **///** 6   7.10.16 **///** A HARMONIZED STRATEGY ON METADATA CATALOGUES AND A REUSABLE REFERENCE IMPLEMENTATION

Fig 2. EIRA high-level overview[1]

//
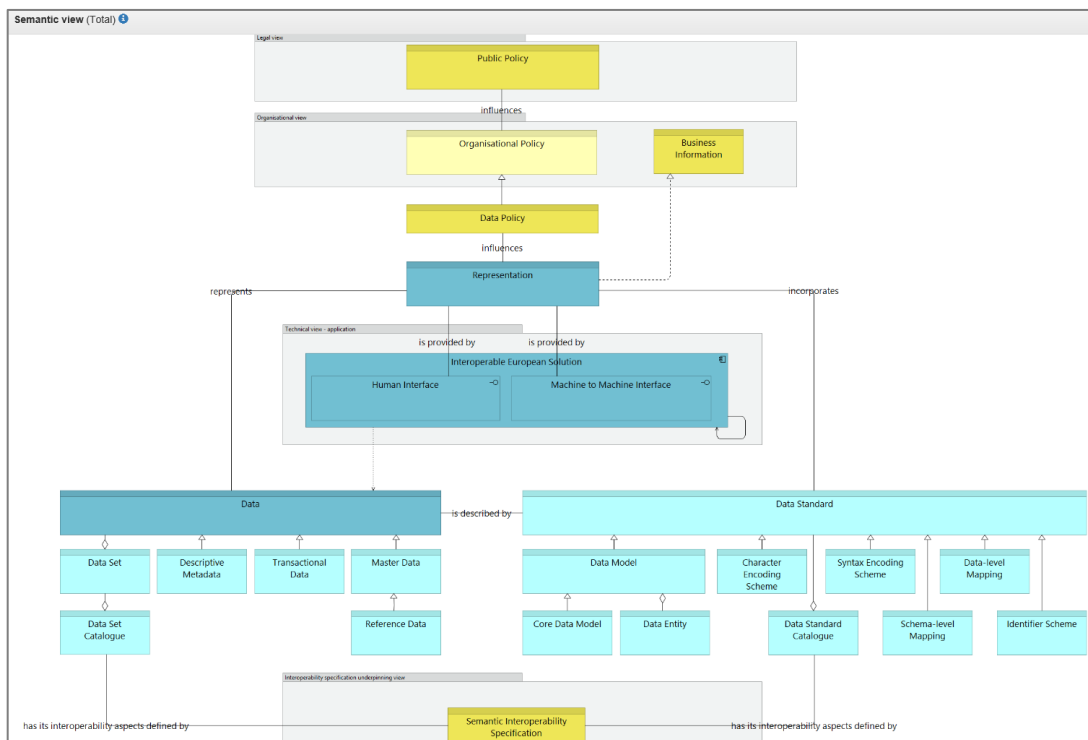
Fig 3. EIRA Semantic view[2]

6 IDENTIFIED USER GROUPS

The target group are the public administrations in Europe (including the EU, federal, regional and local administrations) promoting their public services and information.

510.1 million citizens and entrepreneurs in 28 member states, will benefit from better discoverable information and services.

Private actors, already scraping the web for public services, businesses, addresses etc. will have access to authoritative qualitative data.

7 AUTHOR

Geraldine Nolf, Informatie Vlaanderen

---

[2] https://joinup.ec.europa.eu/asset/eia/description

///