Core vocabularies and their role in promoting semantic interoperability

Athanasios Karalopoulos

European Commission

DIGIT.B6 – Interoperability Solutions for European Public Administrations (ISA)
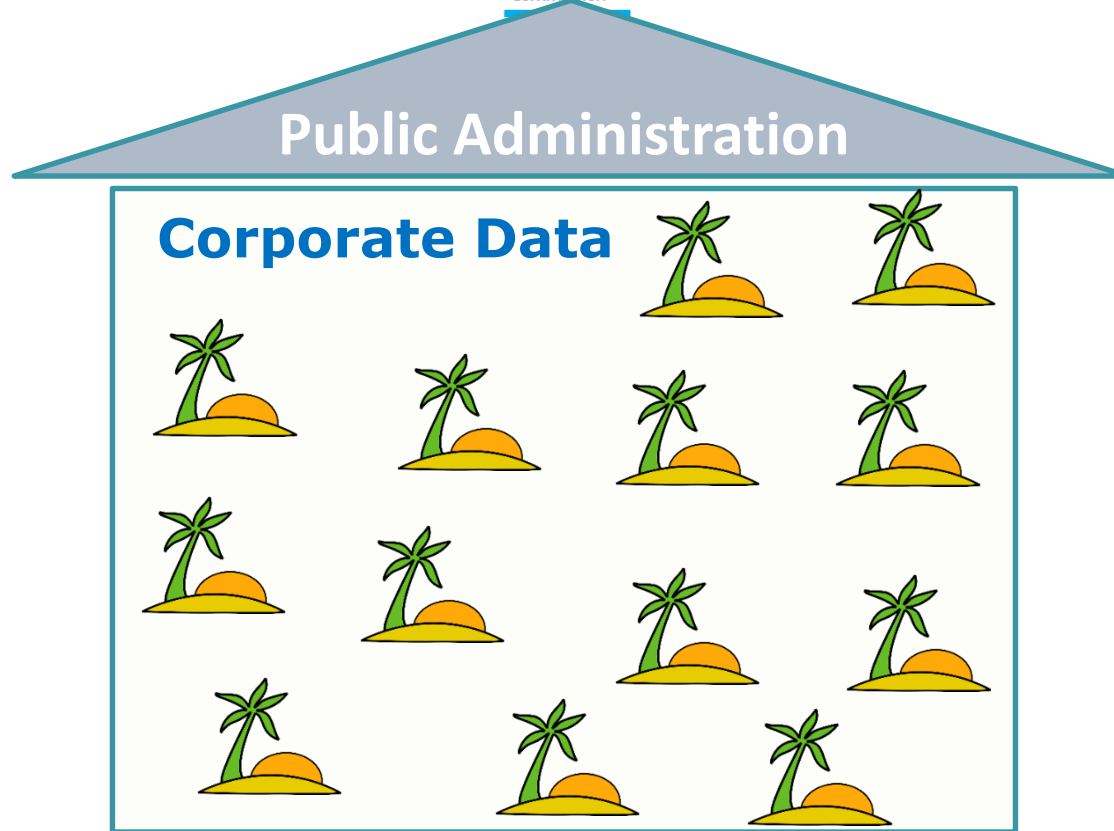
European Commission

ISA

- Every 2 days we create as much information as we did from the beginning of time until 2003

- The number of Bits of information stored in the digital universe is thought to have exceeded the number of stars in the physical universe in 2007
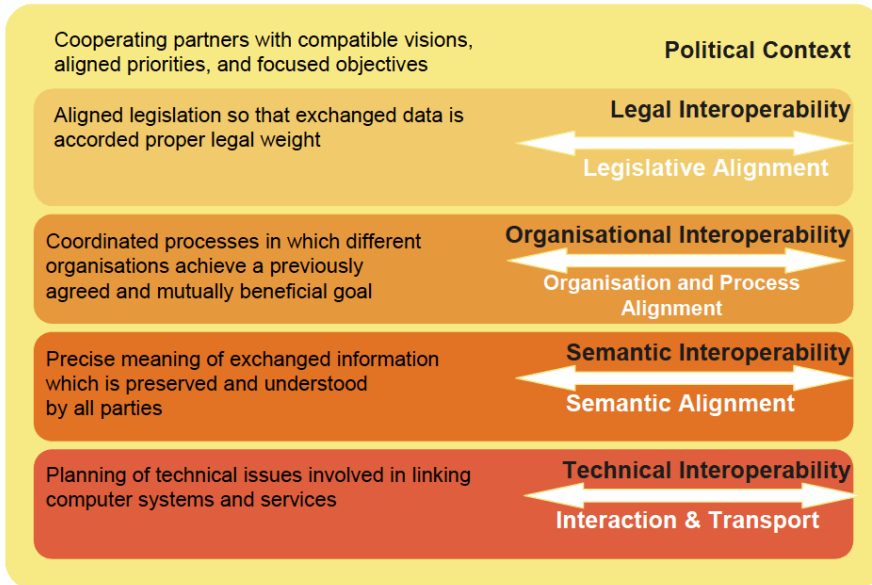
But…

- We still have no corporate information strategy in place

- We still don't know what data is stored in the unit next to us

- We still don't use data standards in our systems

- We still need to manually update hundreds of systems and applications when a new country enters the EU

ISA

European Commission

Public Administration
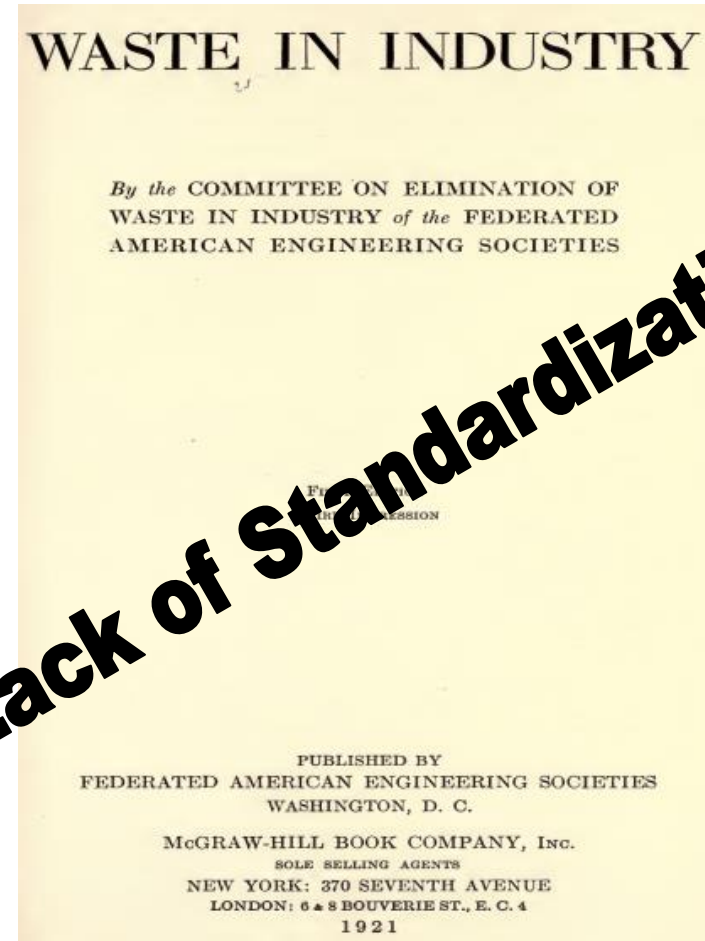
Corporate Data

- Who cares about the global optimum
- How to ensure horizontal oversight & coordination
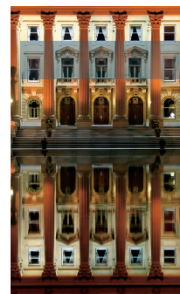- How to achieve global integration with local responsiveness

ISA

European
Commission

Cooperating partners with compatible visions, aligned priorities, and focused objectives
Political Context

Aligned legislation so that exchanged data is accorded proper legal weight
Legal Interoperability
Legislative Alignment

Coordinated processes in which different organisations achieve a previously agreed and mutually beneficial goal
Organisational Interoperability
Organisation and Process Alignment

Precise meaning of exchanged information which is preserved and understood by all parties
Semantic Interoperability
Semantic Alignment

Planning of technical issues involved in linking computer systems and services
Technical Interoperability
Interaction & Transport

How do we achieve technical interoperability?

WASTE IN INDUSTRY

By the COMMITTEE ON ELIMINATION OF WASTE IN INDUSTRY of the FEDERATED AMERICAN ENGINEERING SOCIETIES

Lack of Standardization

PUBLISHED BY
FEDERATED AMERICAN ENGINEERING SOCIETIES
WASHINGTON, D. C.

McGRAW-HILL BOOK COMPANY, Inc.
SOLE SELLING AGENTS
NEW YORK: 370 SEVENTH AVENUE
LONDON: 6 & 8 BOUVERIE ST., E. C. 4
1921

ISA

European Commission

Open Semantic Standards

SOCIAL AGREEMENTS ON DATA STANDARDS

AI AND GOVERNMENT

Editor: **Vassilios Peristeras**, European Commission, vassilios.peristeras@ec.europa.eu

Semantic Standards: Preventing Waste in the Information Industry

Vassilios Peristeras, *European Commission*

It is not sufficient to attempt to standardize the product of a given industry, for almost every industry is so dependent upon others that they too must co-operate.

—Herbert Hoover, 1921

In 1921, Commerce Secretary Herbert Hoover—who then became president of the US—published a report under the title "Waste in Industry." The report, at the peak of the industrial era, standardization identified as a primary enabler to improve production in industry, as well as to cut costs.

Scientific research pushed by the French revolution and the industrial era paved the way for the modern standards, discussed and promoted in several instances and were first perceived as key enablers in network-based industries such as railways. Interestingly, establishing standards across different industries is also quite early.

Just as industrial standards have contributed to the industrial revolution, information standards have become an essential part of the information revolution and information industry. Standardization in information technology was a continuation of industrial standardization, and it operated at a new level. The main goal and advantage of standardization in information technology has been to achieve interoperability for hardware and software.

Barriers at the technical level are only one aspect of the interoperability problem. As widely acknowledged nowadays, for example, in the European Interoperability Framework, interoperability conflicts can appear at the technical, semantic, and/or organizational level. Technical standardization has largely contributed towards truly interoperable networks, devices, and communication protocols. With this progress at the technical level, semantic interoperability is perceived as the next challenging barrier for information exchange, especially in eGovernment environments. However, systematic standardization efforts in the area of semantics are rather rare, and even the term *semantic standard* remains weakly defined. The more general term *standard* varies greatly, depending on the context, and can refer to anything from a screw thread, a unit of measurement, or a way of looking at the world.[2] Semantic standards are related to world interpretations: they represent "a way of looking at the world."

Unless semantic standards and specifications are identified, aligned, documented, managed, and promoted for reuse, we shall suffer from a suboptimal situation in information and communication technology investments. This is particularly true for governments, because they remain heavily invested and the major ICT investor. Semantics stay relevant both for the development of enterprise systems and for open data and interoperable platforms. Semantic standards could also support the availability and quality of linked open government data.[3]

Semantic Standards Worldwide

Although not always systematic, several semantic standardization initiatives are in process worldwide, with two main efforts that aim to create semantic standards libraries or catalogues of what already exists. In US and Europe, government projects are creating repositories of semantic standards and promoting their reuse, with third parties cataloguing existing semantic standards and standardization bodies' initiatives related to semantics.

In the US, the National Information Exchange Model (NIEM, www.niem.gov) has extended its initial coverage, which was restricted to the judicial domain to engage stakeholders from a wide spectrum

Moving from...

- Close-world systems design
  - o I know all my clients and their information requirements
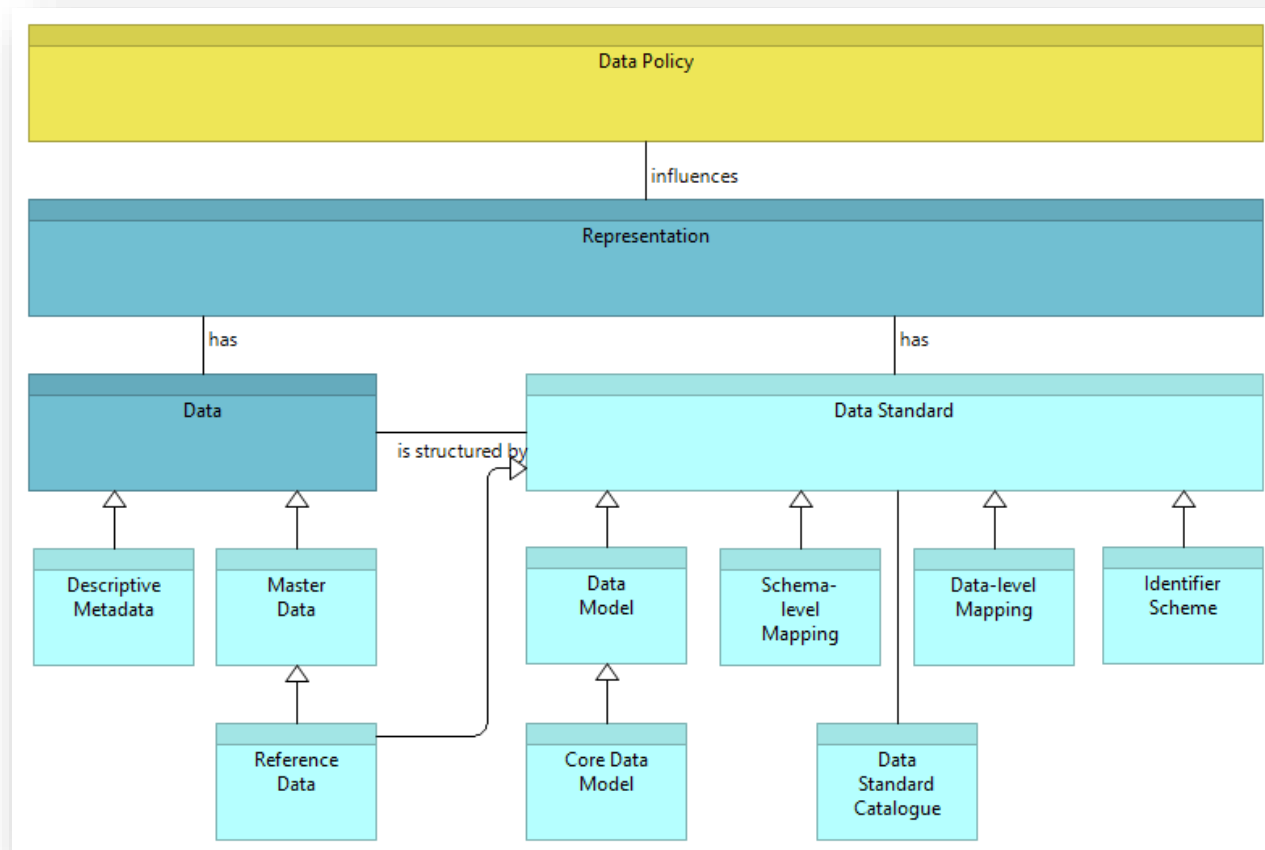   - Island IT design

towards...

- Open-world, data-driven systems design
  - o I don't know who and how in the future will use my data
    - Design based on data standards

Agreeing on common terminology & definitions

Source:https://joinup.ec.europa.eu/site/eia/EIRA/EIRA_beta_dev

Data standard

" A data standard is a structural metadata specification that describes or defines other data [ISO111179]. Structural metadata indicates how compound objects are put together [NISO]. "

European Interoperability Reference Architecture
Semantic View

The two main types of data standards are:

- data models (including Core Vocabularies)
- reference data

European
Commission

Data models

66 A data model is a collection of entities, their properties and the relationships among them, which aims at formally representing a domain, a concept or a real-world thing. 99

European Interoperability Reference Architecture
Semantic View

66 The use of standardised data models as common building blocks for developing information systems guarantees a minimum of semantic consistency and facilitates information exchange. 99

ISA

European
Commission

Reference data

❝ Reference data is small, discrete sets of values that are not updated as part of business transactions, but are usually used to impose consistent classification. Reference data normally has a low update frequency. Reference data is relevant across more than one business systems belonging to different organisations and sectors. ❞

European Interoperability Reference Architecture

Semantic View

→ Examples: lists of status codes, currency codes, country abbreviations, demographic fields, genders, file types.

❝ Standardised reference data is key to data integration and interoperability, and facilitate the sharing and reporting of information. ❞

ISA

Master data

" Master Data is the authoritative, most accurate data that is available about key business entities, used to establish the context for business transactions and transactional data. "

European Interoperability Reference Architecture
Semantic View

- Describe basic information involved in an organisation, e.g. information about persons, companies, vehicles, buildings, locations, roads
- Used by multiple business processes and different public services
- Stored in base registries

" Standardising master data formats and values are critical for successful system integration and information exchanges. "

ISA

European Commission

" What has been discovered over the years is that there are a number of (information) structures that are universal and applicable to all kinds of organizations, both private and public. There are four fundamental categories: People and Organizations, Geography, Physical Resources and Activities and Events. "

David Hay, Describing the World: Data Patterns

Agreeing on the representation of core data entities

The ISA Core Vocabularies

> " The Core Vocabularies are simplified, re-usable and extensible data models that capture the fundamental characteristics of a data entity in a context-neutral and syntax-neutral fashion [1] "

CORE **PERSON** VOCABULARY   CORE **BUSINESS** VOCABULARY   CORE **LOCATION** VOCABULARY   CORE **PUBLIC SERVICE** VOCABULARY

[1] https://joinup.ec.europa.eu/asset/core_vocabularies/description

13

CORE
PERSON
VOCABULARY

CORE
BUSINESS
VOCABULARY

CORE
LOCATION
VOCABULARY

CORE
PUBLIC
SERVICE
VOCABULARY

European
Commission

OSLO - Open Standards for Linked Administrations in Flanders

OSLO, an extension of the Core Vocabularies in a local context, is a simplified, reusable and extensible data model that captures the fundamental characteristics of information exchanged by Flemish public administration in the domains: contact information, localisation and public services.

- OSLO extends the Core Vocabularies to include properties and relationships, like family composition, contact details, or persons acting on behalf of a registered organization.
- Responsible organisation: The Flemish ICT Organisation (V-ICT-OR).

European
Commission

Integrated portfolio management of public services - Estonia

Estonia used and adapted the Core Public Service Vocabulary in order to document and manage public services in service portfolios.

- Estonia extended the CPSV to include information related to benefits and costs of public services delivered through various channels.

- Estonia implemented a Web form for creating descriptions of public services based on the CPSV(-AP).

- Estonia and ISA piloted a metadata reference collection architecture for public services using the CPSV(-AP).

- Responsible organisations: Estonian Information Systems Authority, Ministry of Economic Affairs and Communications

ISA

European
Commission

The Italian Application Profile of the Core Public Service Vocabulary

Italy extended the CPSV(-AP) in order to specify a minimum set of metadata required to   document public services made   available to citizens and enterprises   by public administrations.

- Italy extended the CPSV(-AP) with a number of classes such as licence, role, contact point, required for describing public service in the Italian context.
- Responsible organisations: Agency for Digital Italy

Business Registers Interconnection System (BRIS)

BRIS facilitates the distribution of information from each of the Member States' business registers to the registers of other Member States in a standard message format and in the relevant language version.

- The BRIS XML schemata makes basic company information searchable and exchangeable, and is based on the XML syntax of the Core Business Vocabulary.
- Responsible organisations: DG DIGIT, DG JUST.

European
Commission

OpenCorporates

OpenCorporates makes available information for more than 43 million companies in 50 jurisdictions and is the largest openly licensed database of companies in the world.

- Companies' metadata are available via an open API in RDF based on the Registered Organization Vocabulary, which is the RDF distribution of the Core Business Vocabulary.
- Responsible organisation: Chrinon Ltd.

ISA

European
Commission

Inforegister - Register App Engine Graph API

This is a RESTful service providing access to continuously maintained, linked open government and commercial data for business applications. The data includes debts, unsubmitted tax declarations and submitted annual reports, contact details, VAT registrations and board members of all companies in Estonia.

- Companies' metadata are available in RDF via an API based on the Registered Organization Vocabulary, which is the RDF distribution of the Core Business Vocabulary.
- Responsible organisation: Inforegister

Pilot implementations
developed by ISA with participants from the Member States

- Core Location Pilot: interconnecting Belgian National and Regional Address Registers.
    - Participating MS: Belgium.
- Core Public Service Pilot: describe public services only once.
    - Participating MS: Belgium, Ireland.
- Creating an integrated view of public sector organisation data.
    - Participating MS: Greece, Italy
- Maritime surveillance data pilot: Integrating disparate sources of maritime surveillance data.
    - Participating MS: Spain.
- Linked data pilot on plant protection products.
    - Participating MS: Austria, Belgium, Germany, Greece, Hungary, The Netherlands, Poland, Sweden.
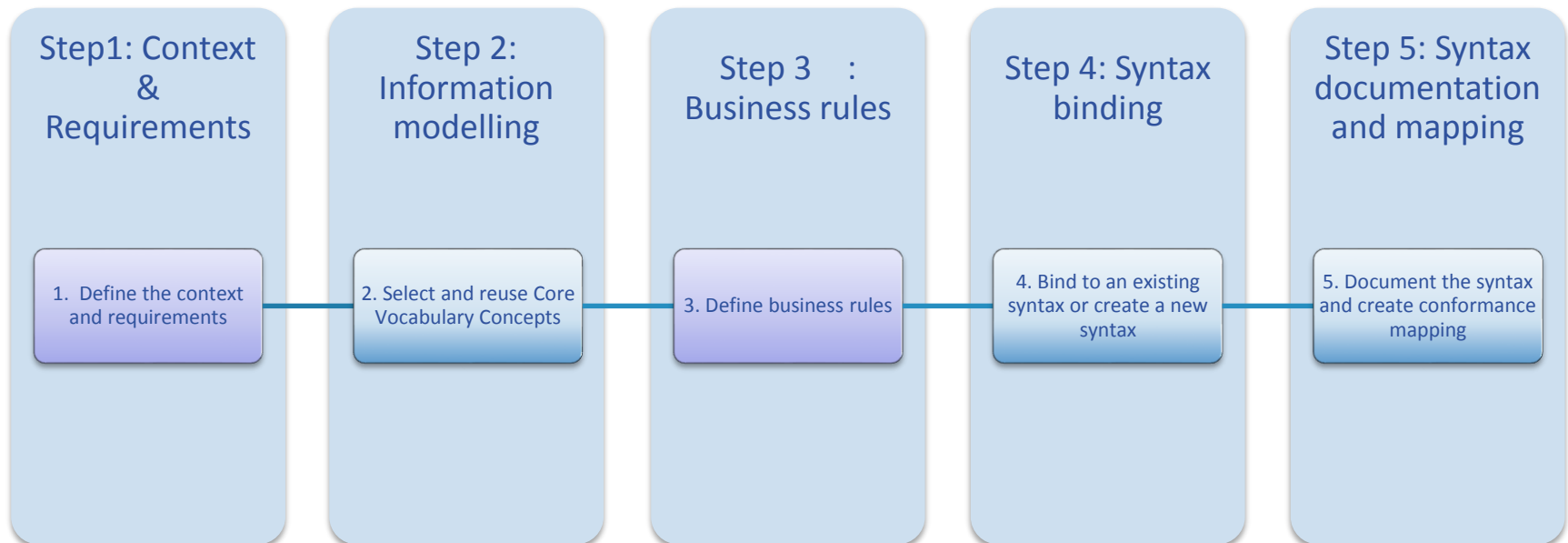
SmeSpire Database – pilot

The SmeSpire database is a public, searchable and structured repository of products, services, projects, tools, procedures, methods and experience of the Geo-ICT SMEs in Europe.

- Project participants: Epsilon Italia, Gist, Cenia, Slovenská agentúra životného prostredia, JRC, Epsilon International, Paragon Europe, Epsilon Consulting LTD, GISIG, PSU, GraphiTec, AGI, Tracasa, KUL, INFO-LOGICA OOD

Handbook for using the Core Vocabularies
How to use the Core Vocabularies for design and alignment?

Step1: Context & Requirements	Step 2: Information modelling	Step 3 : Business rules	Step 4: Syntax binding	Step 5: Syntax documentation and mapping
1. Define the context and requirements	2. Select and reuse Core Vocabulary Concepts	3. Define business rules	4. Bind to an existing syntax or create a new syntax	5. Document the syntax and create conformance mapping

Source: https://joinup.ec.europa.eu/site/core_vocabularies/Core_Vocabularies_user_handbook/Handbook-for-using-the-Core-Vocabularies_v0.50.pdf

Core Vocabularies: what's next

1. **Core Criterion and Core Evidence**

- o Criterion: business/logical rule to be fulfilled to allow the execution of a service (e.g. Age>18)
- o Evidence: (alternative) administrative proof that a criterion is fulfilled (e.g. birth certificate, access to the population registry to retrieve DoB)

Could extend the Core Public Service Vocabulary

2. **Core Public Organisation**

- o How to model a public agency, extension of the Core Business Vocabulary

Work starts soon…

Community of Practice
Scope

A network of representatives from (or working for) public administration organisations aiming to:

- Share knowledge, experiences and lessons-learnt on core data models and reference data
- Harmonise existing initiatives
- Develop together guidelines for core data models and reference data
- Define together areas of future collaboration in ISA Action 1.1

Discover more at:
https://joinup.ec.europa.eu/community/semic/og_page/cop-data-standards

European
Commission

Data standards: guidelines
Objectives

- **Share knowledge** with the members of the Community of Practice.

- **Support new and on-going initiatives** on data standards in public administrations by providing common guidelines, good practices and lessons learned concerning the management of data standards.

- **Identify alignment opportunities** amongst the various initiatives.
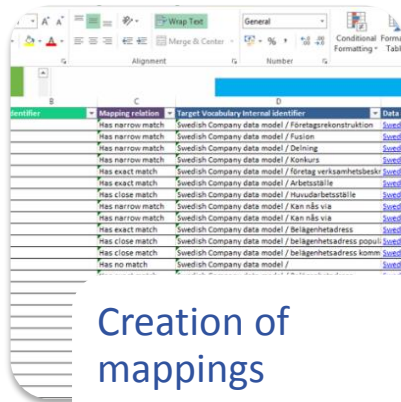
Data standards: case studies
Objectives

A case study is considered within the scope of the current project if it meets one or more of the following criteria:

- It relates directly to the **management and governance of data standards**.

- It is about **data standards** targeting the implementation of IT systems having regional, national or cross-border coverage, or concerns the **exchange of information** between systems within or across borders.

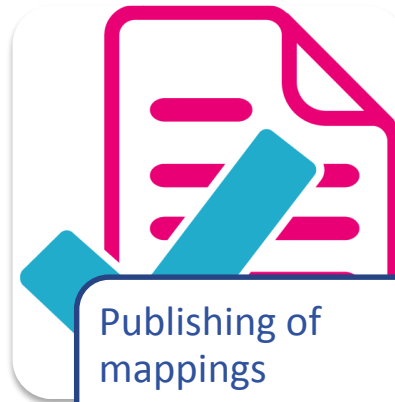- It relates to **data standards** covering one or more policy areas.

Harmonising core data models
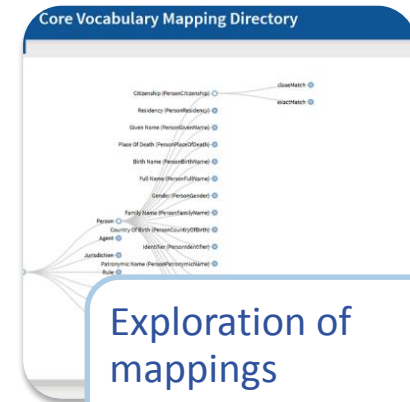## Core Data Models Mapping Directory

Creation of mappings

• Concept mappings
• Spreadsheet based

Publishing of mappings

• Quality control
• Transform to RDF

Exploration of mappings

• browsing
• Visual
• Machine processeable

Discover more at: http://mapping.semic.eu

Harmonising core data models
Creation of mappings between Core Data Models

Conceptual schema level mappings

• Express relationships between entities

• The existence of a relationship is prerequisite for an implementation

Syntax level mappings

• Express relationships between entities and how one representation can be transformed into another one

• Requires:

  o A physical representations of each Core Data Model (e.g. XML, DB, RDF)

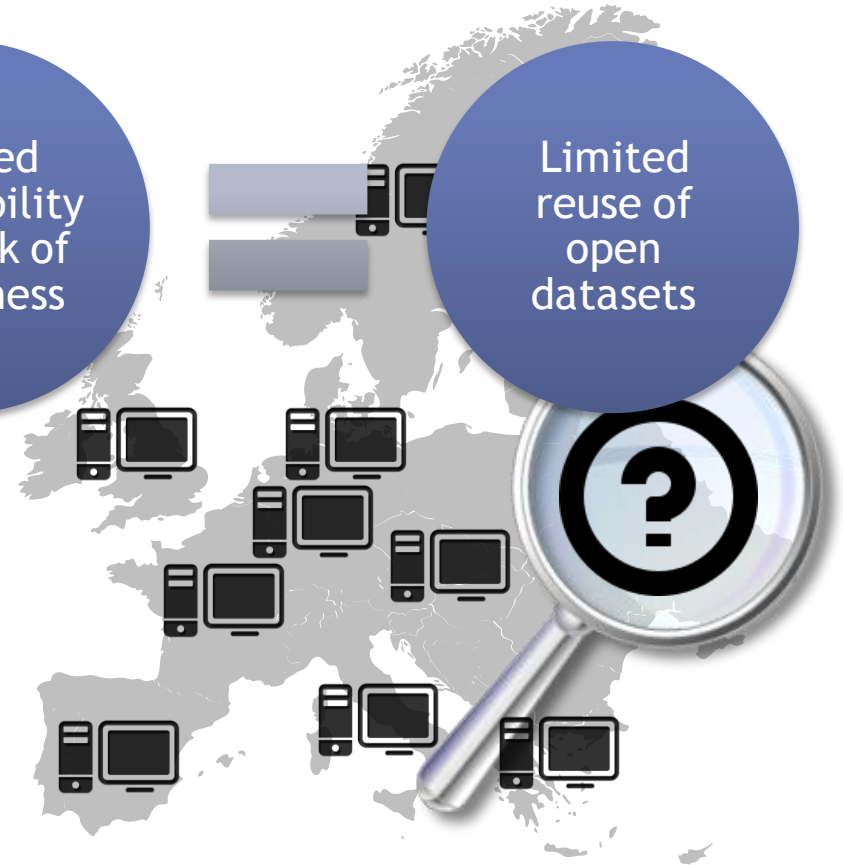 o A transformation engine. (e.g. XSLT, R2ML)

European Commission

Different metadata vocabularies

+

Limited accessibility and lack of awareness

=

Limited reuse of open datasets

How can I find and combine public data from various sources?

European
Commission

DCAT
APPLICATION
PROFILE FOR
EUROPEAN
DATA PORTALS

The DCAT Application profile (DCAT-AP) is a common template to describe public sector datasets and data catalogs

- DCAT-AP v1.1 published in October 2015
- GEO/DCAT-AP to be published in December 2015
- STAT/DCAT-AP is under development, expected Q2 2016

The European Data Portal implements the DCAT-AP

The European Data Portal harvests the metadata of datasets available on public data portals across European countries. There are over 258,000 datasets harvested from 67 web portals from 34 countries.

- The European Data Portal uses the DCAT-AP as the interoperability specification for harmonising and providing metadata of datasets from:
  - Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, France, Germany, Greece, Finland, Hungary, Iceland, Ireland, Italy, Latvia, Lichtenstein, Lithuania, Luxembourg, Malta, Moldova, The Netherlands, Norway, Poland, Portugal, Romania, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, United Kingdom
- Responsible organisation: DG CONNECT, European Data Portal project

ISA

DCAT-AP validator – Open Data Support

The DCAT-AP validator is a web application that checks metadata description of datasets for integrity and consistency against the DCAT-AP specification.

- Responsible organisations: DG CONNECT, PwC

CKAN DCAT-AP extension & plugins

CKAN is a data management system that makes data accessible – by providing tools to streamline publishing, sharing, finding and using data. CKAN is the most popular open-source software for implementing open data portals across the globe.

- This extension provides plugins that allow CKAN to expose and consume metadata from other catalogs using RDF documents serialized using DCAT(-AP).
- Responsible organisation: CKAN

OpenDataSoft solution

OpenDataSoft is a French SME developing open data portal solutions, which have been installed in a number of cities, including those of Paris and Brussels.

- The software of OpenDataSoft supports a DCAT-AP export.
- Responsible organisation: OpenDataSoft

ISA Action 1.1 - http//semic.eu

Vassilios.Peristeras@ec.europa.eu

Athanasios.Karalopoulos@ec.europa.eu

Visit our initiatives

ADMS
ASSET
DESCRIPTION
METADATA
SCHEMA

DCAT
APPLICATION
PROFILE FOR
EUROPEAN
DATA PORTALS

ADMS.SW

REGISTERED ORGANIZATION VOCABULARY

eGOVERNMENT
CORE
VOCABULARIES

CORE PUBLIC SERVICE VOCABULARY

CORE LOCATION VOCABULARY

CORE PERSON VOCABULARY

Get involved

Follow @SEMICeu on Twitter

Join the SEMIC group on LinkedIn

Join the SEMIC community on Joinup

contact@semic.eu

ISA