Core vocabularies and their role in promoting semantic interoperability

Dr. Vassilios Peristeras

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

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
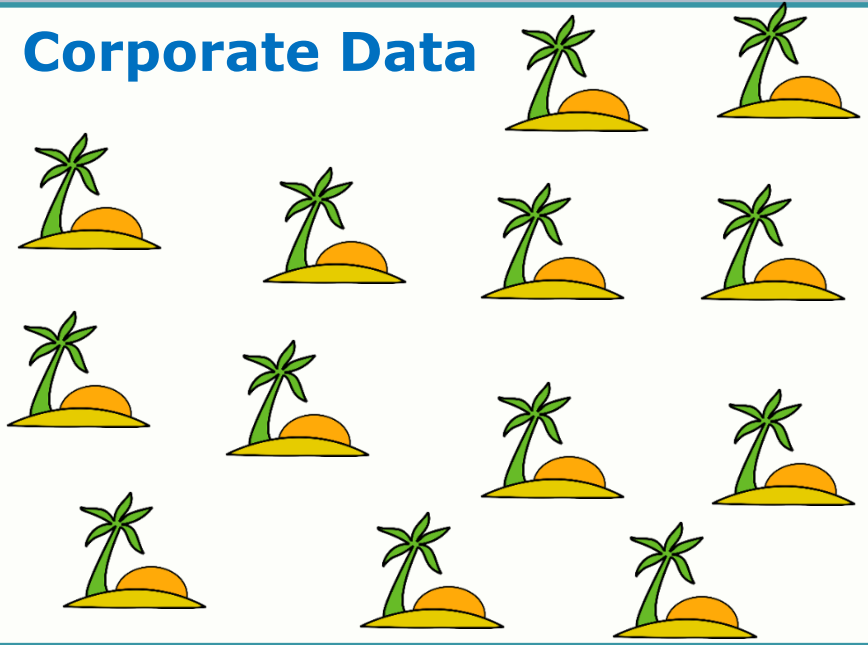

But…

- 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

Public Administration

Corporate Data

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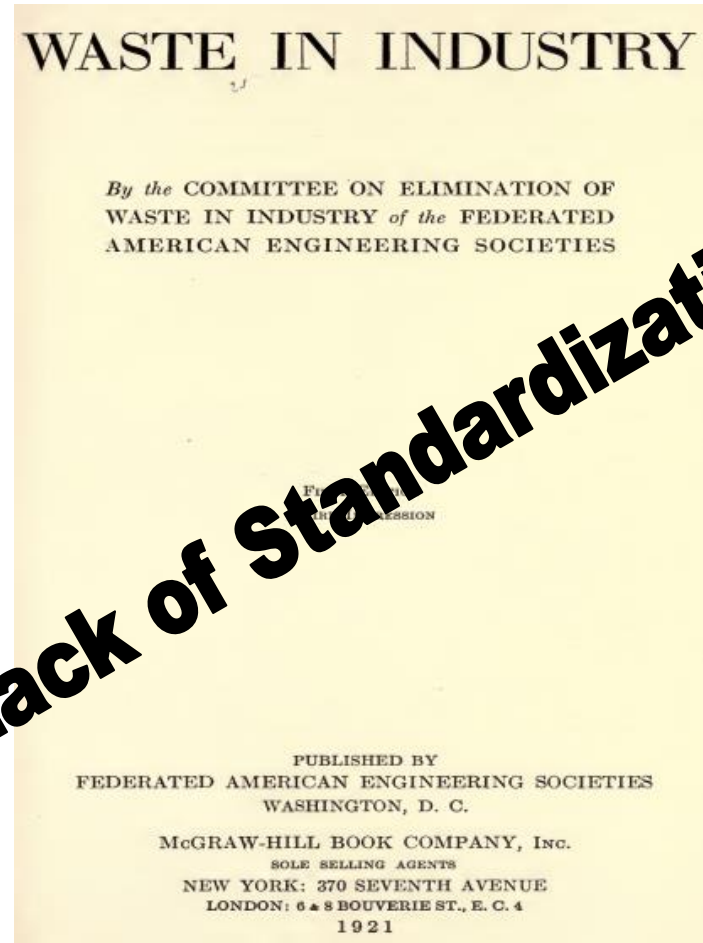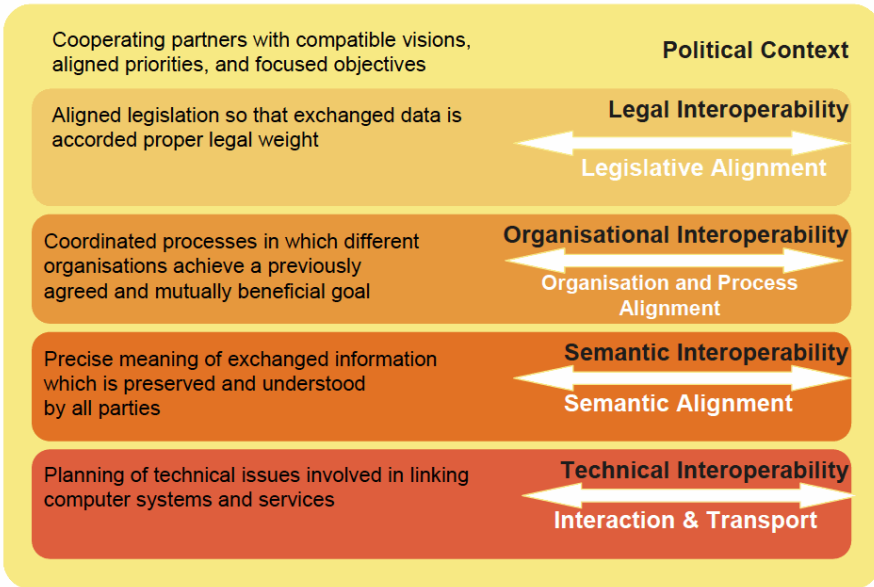- How to achieve global integration with local responsiveness

Public Administration

Corporate Data

Information strategy

– Principles

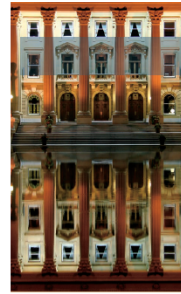– Governance (who CRUD info)

– Data standards

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?

Lack of Standardization

WASTE IN INDUSTRY

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

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

European Commission

ISA

SOCIAL AGREEMENTS
ON DATA STANDARDS

Open Semantic 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 engine to improve production in industry, as well as to cut costs and waste.

Scientific work started by the French revolution and the metric system paved the way for the metric standard. It was discussed and promoted in several countries. Standards were first perceived as key in the network industries such as railways. Interestingly, the need to establish standards across different industries as identified quite early.

Just as industry standards have contributed to the industrial revolution, information standards have become an important engine in the information revolution and information society. Standardization in information technology was a central concern for industrial standardization, and it started at the technical 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 prepared for reuse, we shall suffer from a substantial waste in information and communication technology investments. This is particularly true for governments, because they remain heavily involved and the major ICT investor. Semantics are also relevant both for the development of enterprise systems and for open data and interoperable platforms. Semantic standards could also affect the availability and quality of linked open government data.[3]

Semantic Standards Worldwide

Although not always systematic, several semantic standardization activities are in process worldwide, including efforts that aim to create semantic-related 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
  - I know all my clients and their information requirements
    - Island IT design

towards...

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

European Commission

Increasing visibility of existing data standards

ADMS & Joinup Catalogue of data standards

Establishing core standards for master data (base registries)

Core Vocabularies & Access to Base Registries

Improving interoperability of open data

DCAT-AP

Promoting good practices: metadata, big data and eParticipation

Case Studies and Recommendations

isa

| Pilots | Communities | Studies | Visits | Events |

Making visible existing solutions

ADMS & Catalogue of semantic standards

isa
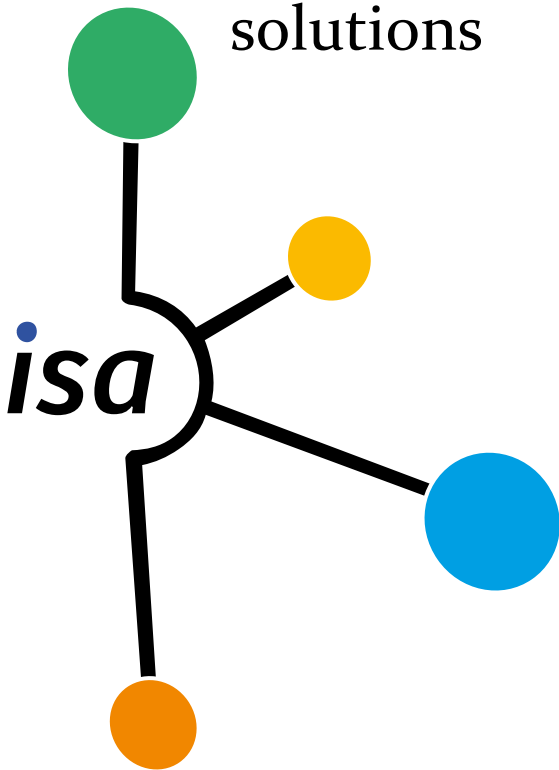
ADMS

- Using the same data (semantic) standards promotes interoperability

- By reusing what is available you save resources, you reduce risks, and you become more interoperable with others

- A large number of data (semantic) standards already exists

ISA

- The existing solutions are scattered in numerous places and are very difficult to find

  - Several national initiatives to create repositories/libraries/catalogues of semantic standards (e.g. Germany, Denmark, Finland, Estonia…)

  - Standardization bodies and third party initiatives generate valuable and highly reusable specifications (e.g. OASIS, W3C, UN/CEFACT…)

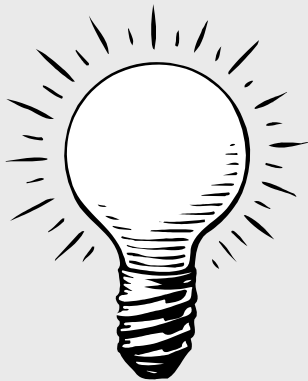 - Independent projects make available semantic standards to their own websites

ISA

How could we promote the visibility and reuse of existing data (semantic) standards at the European level?

How could we promote the reuse of existing data (semantic) standards at the European level?

... by agreeing on a common language (template) to describe semantic standards
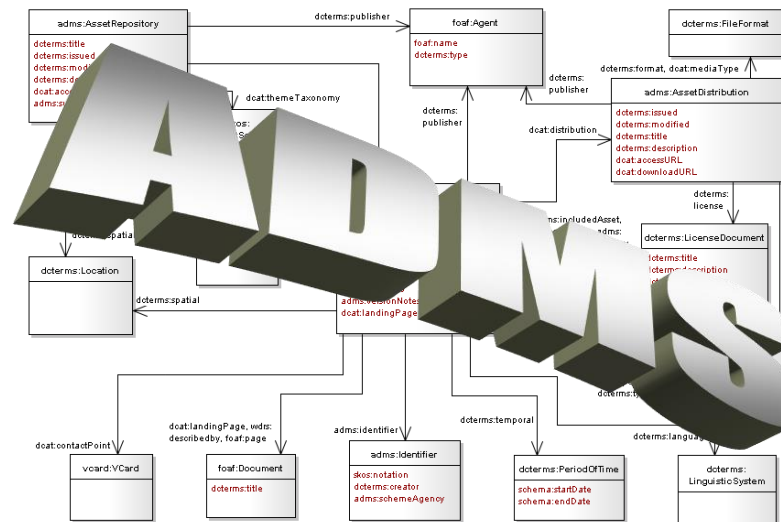
... creating a yellow page infrastructure with standards descriptions and links to the actual standards

Common template (metadata) for describing semantic standards

Asset Description Metadata Schema (ADMS)

May 2012: ADMS endorsed by the EU member states (ISA Coordination Group)

European Federated Interoperability Repository (EFIR)

ADMS-based repository of any type of interoperability solution

Catalogue of interoperability solutions

joinup

- Interoperability solutions are described using ADMS

- Features simple and advanced search of interoperability solutions

- Semantic and technical standards, open source software, reusable services

Participant Profiles

Who is already sharing?

42 partner organisations

15 Member States

8 European Institutions

4 standardisation organisations

8 other stakeholders

9 Federated Forges

+ many individual contributors

= more than 2000 interoperability solutions

Implementation of ADMS

1. Joinup

The repository aims to facilitate the sharing and reuse of interoperability solutions made for public administrations. Although it reaches interoperability professionals around the world, the repository itself focusses on European solutions.

The solutions hosted on Joinup are described according to ADMS specifications. Joinup provides ADMS-compliant RDF export of metadata describing the solutions.

- Project participants: European Commission, ISA Programme

Implementation of ADMS

2. The Metadata Registry of the Publications Office of EU

The Metadata Registry (MDR) registers and maintains definition data (metadata elements, named authority lists, schemas, etc.) used by the different European Institutions.

MDR provides ADMS-compliant RDF export of Named Authority Lists.

- Project participants: Publications Office

Implementation of ADMS

3. Poolparty Semantic Suite of the Semantic Web Company

PoolParty is a semantic technology suite that offers solutions for knowledge management. As a semantic middleware, PoolParty can enrich business information with valuable metadata and links content assets automatically. The PoolParty Thesaurus & Taxonomy Manager is one of the products offered by the Poolparty suite which allows the publication of projects according to ADMS.

- Project participants: The Semantic Web Company

Making visible existing solutions ▶▶ ADMS & Catalogue of semantic standards
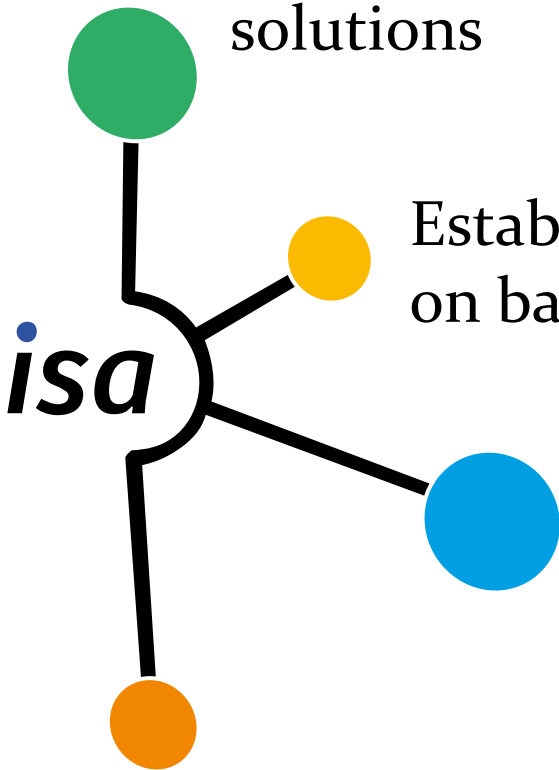
Establishing agreements on basic semantics ▶▶ Core Vocabularies

"...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

Value

- **New systems:** Core Vocabularies to be used as default starting points for data modeling
- **Existing systems:** Core Vocabularies to be used as reference data models to allow semantic mapping and as the basis to create exchange protocols
- **Open Data:** As export specifications for publishing open data
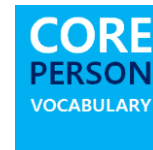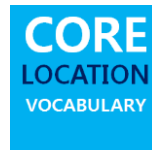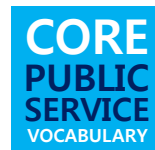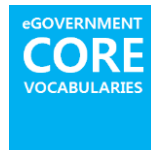
Core vocabularies

Simplified, re-usable, generic and extensible data models that capture the fundamental characteristics of a data entity in a context-neutral fashion.

CORE PUBLIC SERVICE VOCABULARY

CORE PERSON VOCABULARY

CORE BUSINESS VOCABULARY

CORE LOCATION VOCABULARY

DCAT APPLICATION PROFILE FOR EUROPEAN DATA PORTALS
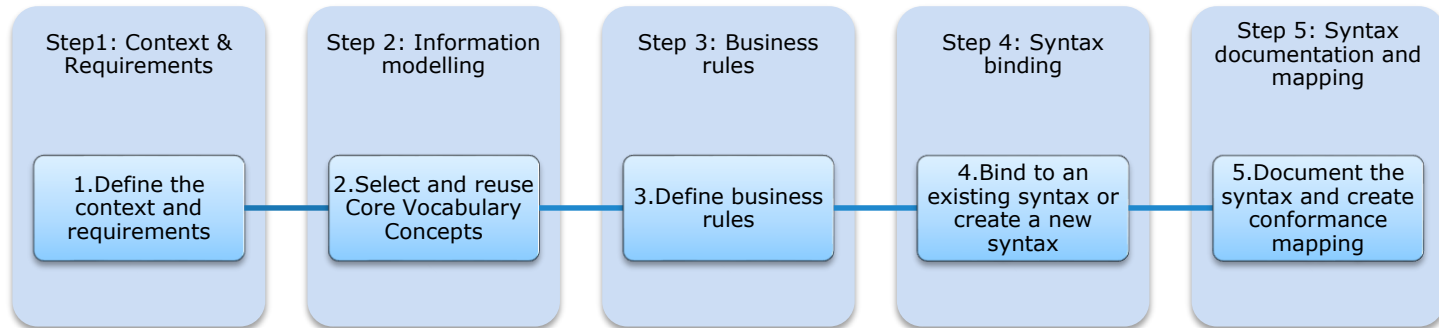
D3.1 – PROCESS AND METHODOLOGY FOR CORE VOCABULARIES

Deliverable

JOINING UP GOVERNMENTS

Handbook for using the Core Vocabularies
How to use the Core Vocabularies?

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

Step 1: Context and requirements
The aim of this initial step is to define the context and elicit a set of technology-neutral requirements for the data model to be designed.

Step 2: Information modelling
When documenting information requirements, the conceptual data model of the Core Vocabularies is used as a starting point that can be customised.

Step 3: Business rules
The outcome of the third step is an enhanced data model with the cardinalities and constraints and the lists of sets of values that restrict the possible values for coded elements.

Step 4: Bind to an existing syntax or create a new syntax
When there is a standard syntax supporting a conceptual data model, it is recommended to maximally use the existing standard syntax. If no standard syntax is available, then a new syntax element can be created.

Step 5: Syntax documentation and mapping
The aim of this step is to create documentation of the syntax that allows users to implement it, and at the same time allows the owner to claim conformance of the data model to the Core Vocabularies.

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

Implementation of Core Vocabularies

Pilot projects developed by ISA together with participants from the Member States

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

Implementation of Core Vocabularies

Projects developed by third parties

1. OSLO - Open Standards for Linked Administrations in Flanders

OSLO 1.1 extends the ISA Core Vocabularies to include properties and relationships needed by local administrations in Flanders, like family composition or persons acting on behalf of a registered organization.

OSLO 2.0 proposes guidelines for implementing web services and introduces the Software Catalogus, a repository of open standards, software packages, and service providers.

- Core Vocabularies used: Core Location Vocabulary, Core Person Vocabulary, Core Business Vocabulary, Core Public Service Vocabulary.
- Project participants: The Flemish ICT Organisation (V-ICT-OR).

Implementation of Core Vocabularies

Projects developed by third parties

2. Estonian Ministry of Economic Affairs – Integrated portfolio management of public services

The project aimed at increasing the quality of state planned public services providing (1) a machine-usable service description language, (2) a method for measurements of benefits and costs of public services delivered through various channels, (3) study feasibility of the „portfolio" management of public services, and (4) developing an architectural vision.

- Core Vocabularies used: Core Public Service Vocabulary
- Project participants: Estonian Information Systems Authority, Ministry of Economic Affairs and Communications, IT and Development, Centre of the Ministry of Interior, Estonian Association of Information Technology and Telecommunications.

Implementation of Core Vocabularies

Projects developed by third parties

3. 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 Vocabularies.

- Core Vocabularies used: Core Business Vocabulary.
- Project participants: DG DIGIT, DG JUST.

Implementation of Core Vocabularies

Projects developed by third parties

4. 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 various formats such as JSON, XML and RDF.

- Core Vocabularies used: Core Business Vocabulary.
- Project participants: Chrinon Ltd.

Implementation of Core Vocabularies

Projects developed by third parties

5. 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.

- Core Vocabularies used: Core Business Vocabulary, Core Location Vocabulary, and Core Person Vocabulary.

- 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

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)
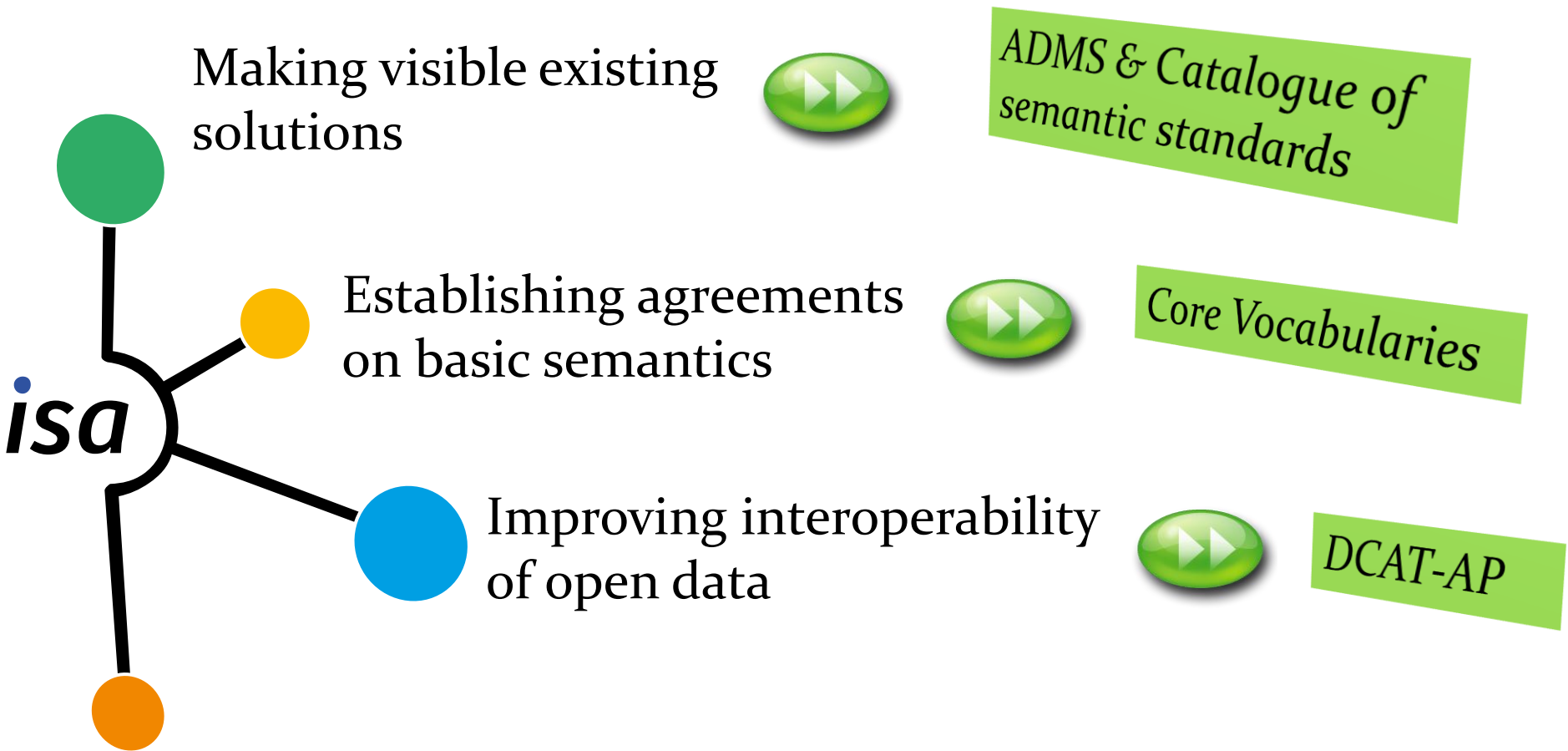
Could extend the Core Public Service Vocabulary
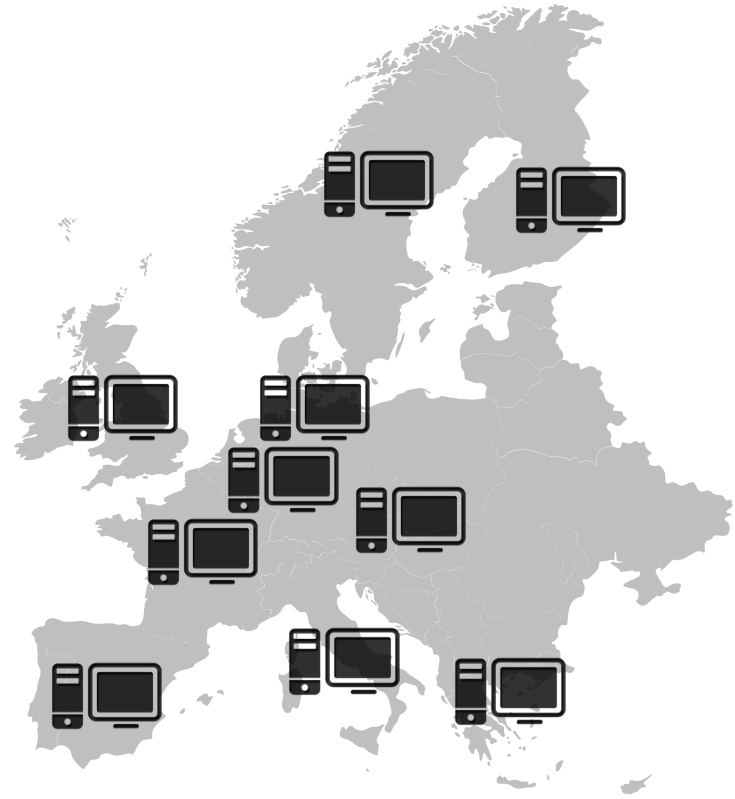
2) Core Mandates and Powers
   o To be used by eIDAS for eIDs

3) Core Public Organisation
   o How to model a public agency, extension of the Core Business Vocabulary

Work starts soon…
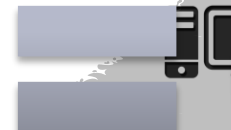
Making visible existing solutions

ADMS & Catalogue of semantic standards

Establishing agreements on basic semantics

Core Vocabularies

Improving interoperability of open data

DCAT-AP

isa

150+

Existing Open Government Data Portal

European Commission

Different metadata vocabularies

+

Limited accessibility and lack of awareness

=

Limited reuse of open datasets
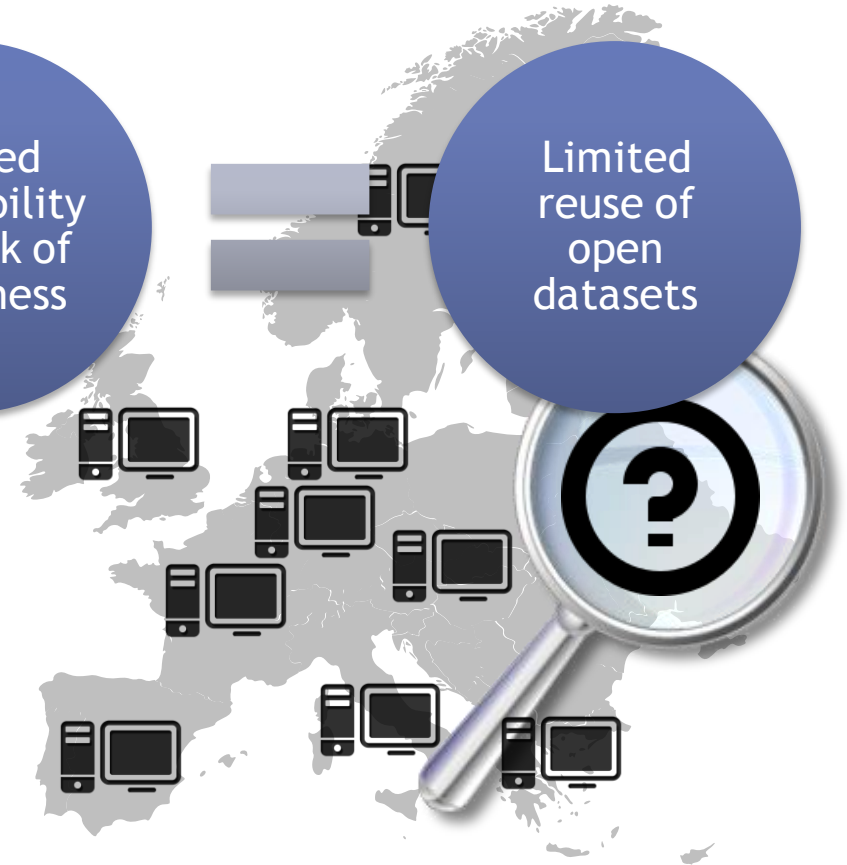
How can I find and combine public data from various sources?

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 Q1 2016

Implementation of DCAT-AP

1. European Data Portal

The European Data Portal harvests the metadata of Public Sector Information available on public data portals across European countries.

There are over 258,000 data sets harvested from 67 web portals. The metadata are structured according to established categories that follow the revision of the DCAT-AP.

- Project participants: European Commission

ISA

Implementation of DCAT-AP

2. DCAT-AP validator developed in 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.
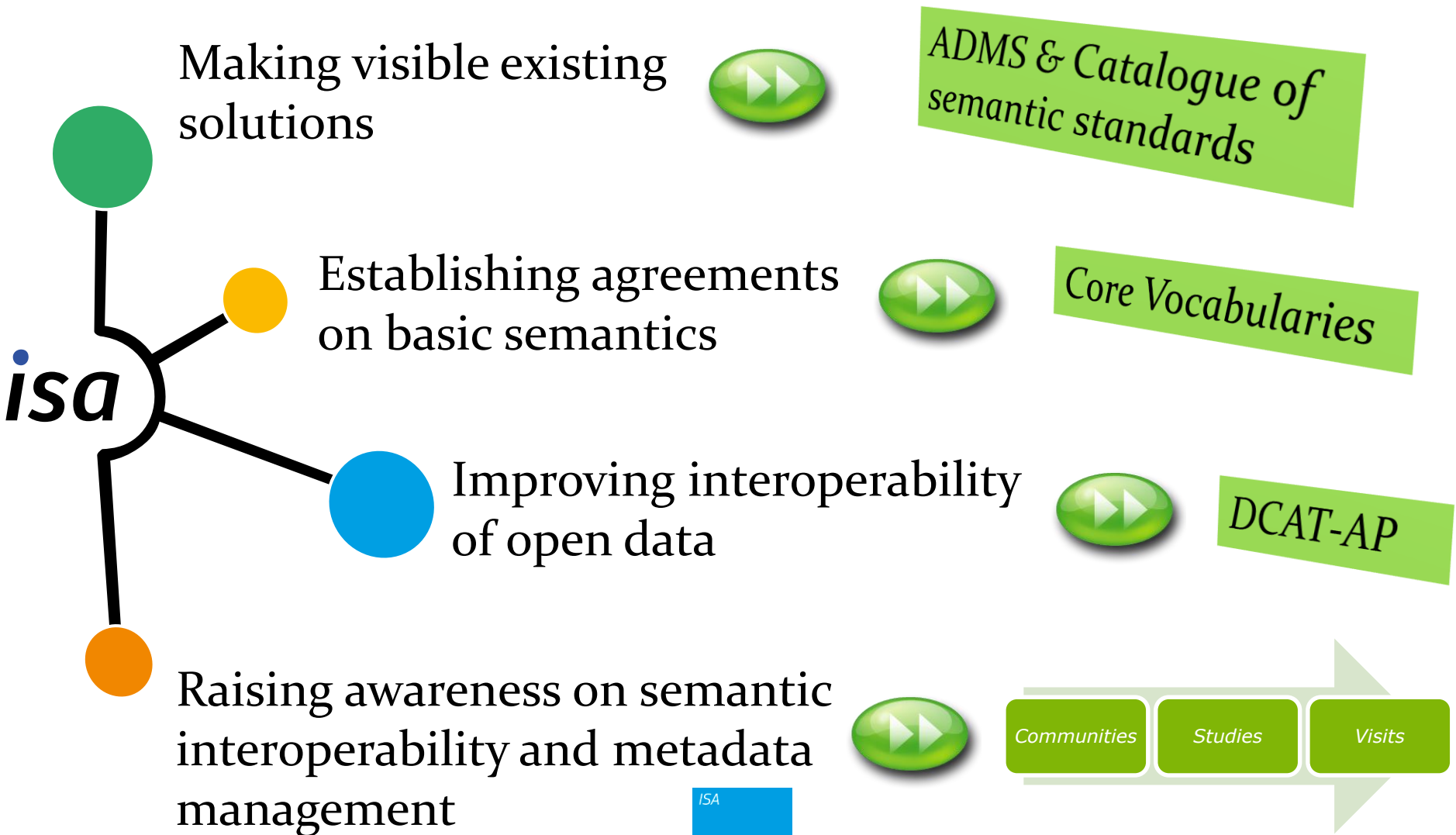
- Project participants: DG CONNECT

Implementation of DCAT-AP
In the member states

National data portals provide metadata export:

- The Dutch Open Data Portal (https://data.overheid.nl)
- The Brussels Open Data Portal (http://opendata.brussels.be)
- The Flemish Open Data Portal (http://opendataforum.info)
- Swedish national data portal (http://oppnadata.se/)
- Norwegian Open Data Portal (http://spec.dataportals.org/)
- The Spanish Open Data Portal (http://www.datos.gob.es)

Re-use

Do not reinvent

How Linked Data is transforming e-Government

https://joinup.ec.europa.eu/community/semic/document/case-study-how-linked-data-transforming-egovernment

10 Rules for Persistent URIs

https://joinup.ec.europa.eu/community/semic/document/10-rules-persistent-uris

Cookbook for translating relational data models to RDFs

https://joinup.ec.europa.eu/community/semic/document/cookbook-translating-data-models-rdf-schemas

SEMIC Conferences

https://joinup.ec.europa.eu/community/semic/event/semic-2014-semantic-interoperability-conference
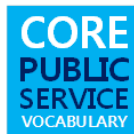
Core Vocabularies Handbook *(in progress)*

Business and cost models for Linked Government Data

https://joinup.ec.europa.eu/community/semic/document/study-business-models-linked-open-government-data-bm4logd

Visit our initiatives **joinup**

Get involved

Follow @SEMICeu on Twitter

Join the SEMIC group on LinkedIn

Join the SEMIC community on Joinup

ISA

Vassilios.Peristeras@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

ISA

Get involved

Follow **@SEMICeu** on Twitter

Join the **SEMIC** group on LinkedIn

Join the SEMIC community on Joinup