Promoting semantic interoperability for open public data

15 May 2014, Paris
Vassilios.Peristeras@ec.europa.eu

ISA

From national close documents… to European linked open data

The ISA work on promoting open data standards

Feedback on the questions received

European
Commission

Modernisation of public administrations in the EU political agenda

- **Extract from Council Conclusions, October 2013:**

" *Open data is an untapped resource with a huge potential for building stronger, more interconnected societies that better meet the needs of the citizens and allow innovation and prosperity to flourish.* **Interoperability and the re-use of public sector information shall be promoted actively**. **"**

European Commission

from National Closed Documents...

...to European Open Data

4

Public Sector Information

- **Governments are producers, collectors, owners and distributors of huge amount of information**
- **Historically, restricted and problematic access for several reasons**

European
Commission

- **"Closed by default"**
- **Accessibility**
- **Discoverability**
- **Cross-referencing**
- **Handling**
- **Costs**

From Documents…

… to data

Name	XXXXX XXXX
Date of Birth	31/10/1924
Place of Birth	Garrett, Maryland

National Closed Data Documents...

Open Data initiatives…

- promote transparency
  - promoting accountability by providing the public with information about what the Government is doing
- allow the creation of new, innovative, added-value services
  - enabling the mash up of existing data in unforeseen ways
- accelerate scientific progress
  - to meet the challenges of the 21st century in scientific discovery and learning
- improve the quality of decision-making
  - providing the means for evidence-based policy development
- foster collaboration across government and beyond
  - improving effectiveness by encouraging partnerships and cooperation across levels of government, and between the Government and private institutions
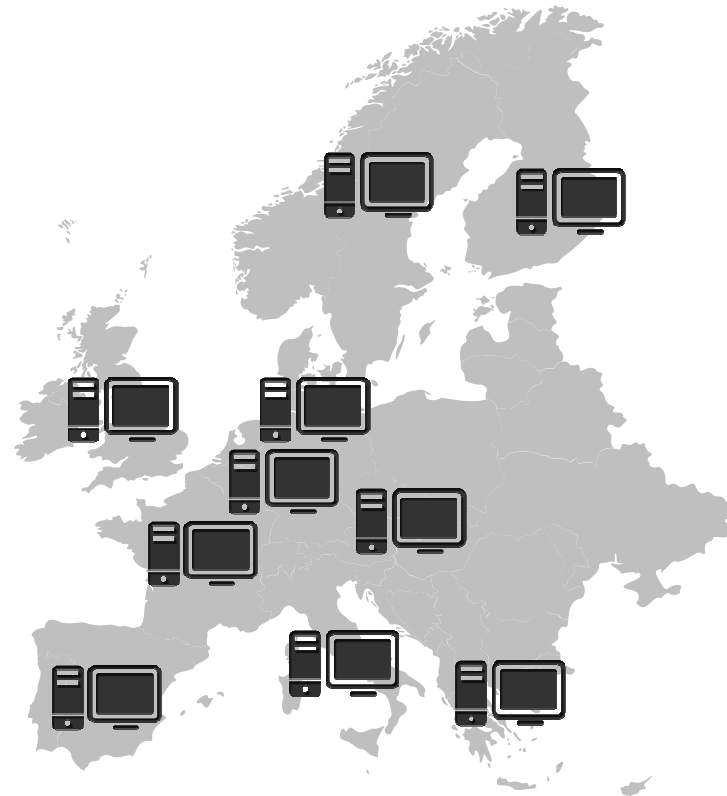
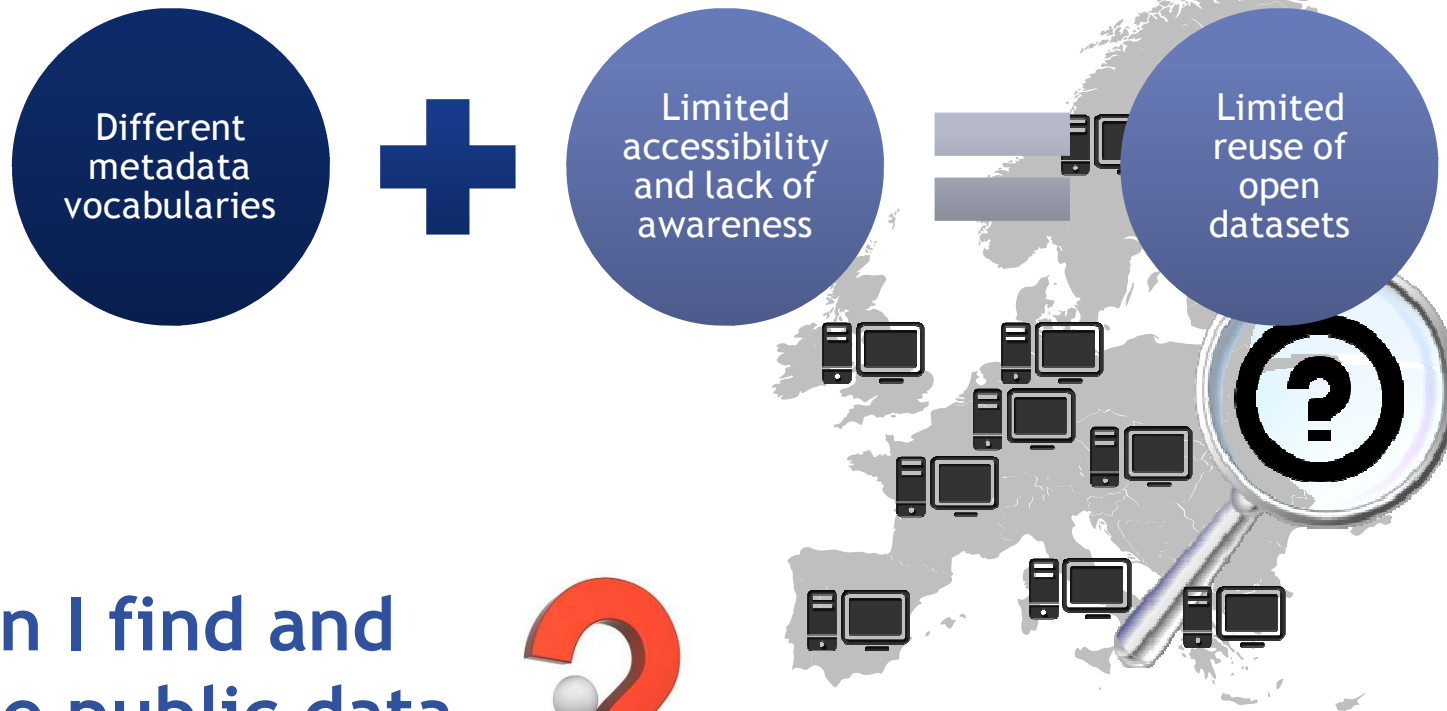European
Commission

European Commission

Open data strategy

- Communication on Open Data ([COM(2011)882](#))
- A revision of the Decision governing the re-use of Commission's own information ([2011/833/EU](#))
- Revision of the Directive on the re-use of public sector information ([2013/37/EU](#))
- Open data-portals
  - [EU Open Data portal](#)
  - [Pan-European portal](#)
- Support for R&D

Open
National ~~Closed~~ Data...

150+
Existing OGD Portal

European
Commission

Different
metadata
vocabularies

+

Limited
accessibility
and lack of
awareness

=

Limited
reuse of
open
datasets

?

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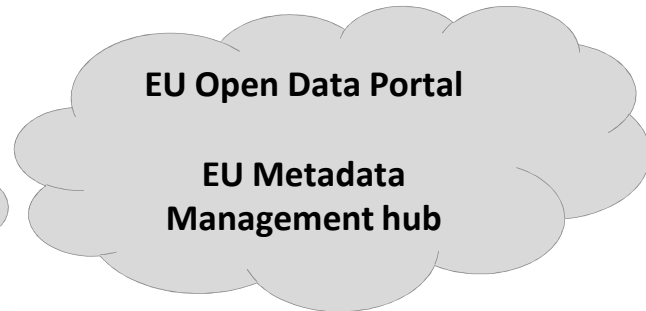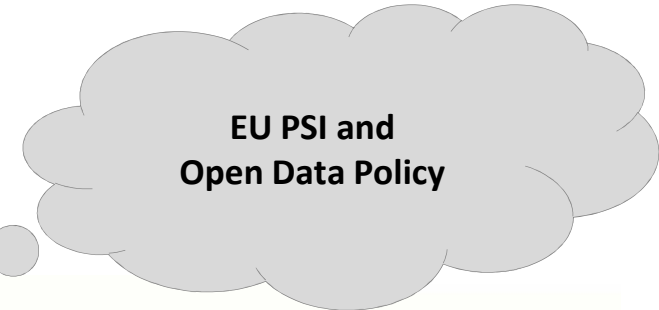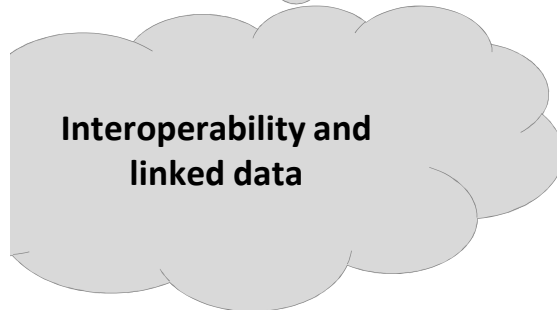combine public data
from various sources?**

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
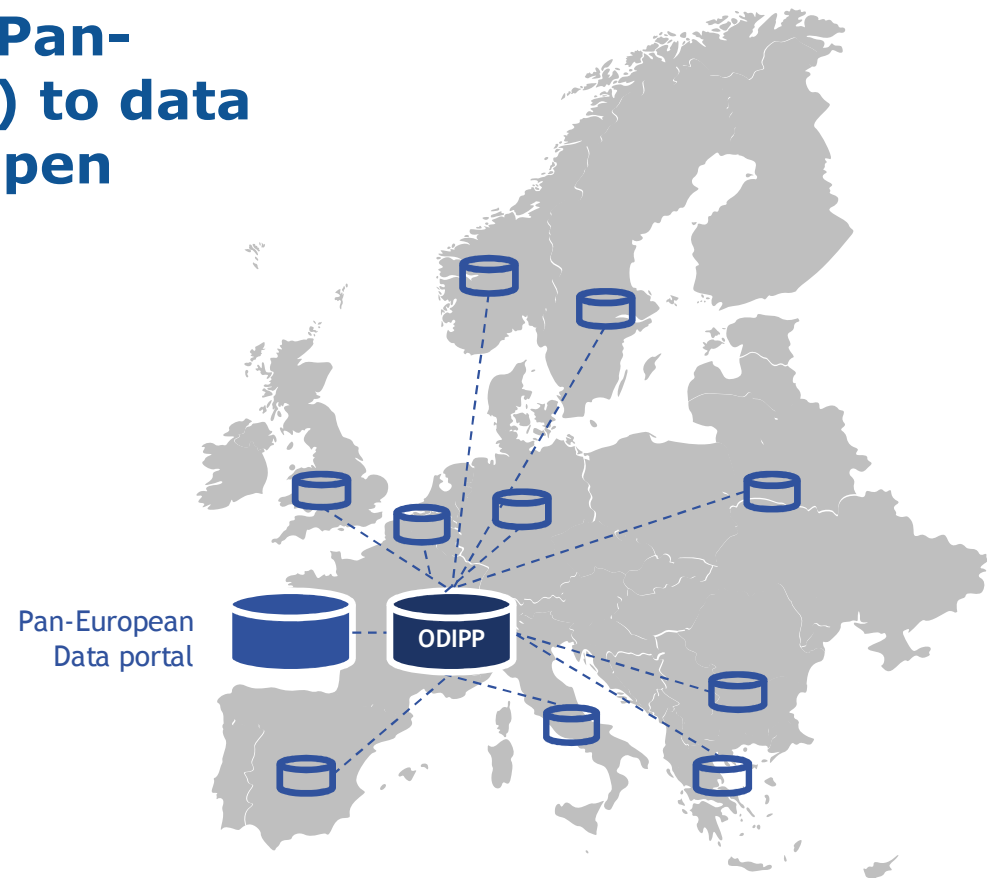
European
Commission

Shared initiative ...

EU PSI and
Open Data Policy

Communications Networks, Content and Technology
European Commission Directorate General

DG CONNECT

isa
Interoperability Solutions
for European Public Administrations

EU Open Data Portal

EU Metadata
Management hub

**Interoperability and
linked data**

Publications Office

DCAT-AP enables the provision of a single point of access (aka Pan-European Open Data Portal) to data sets published in national open portals

Pan-European
Data portal

ODIPP

publicdata.eu

**2013:pilot
of a pan-European
Open Data Portal**

**2014-2020:
progressive
implementation
as one of the CEF**
(Connecting Europe Facility)
**Infrastructures
(Open Data)**

From national close documents… to European linked open data

The ISA work on promoting open data standards

Discussion on the questions received

Existing data standards should be used wherever possible when publishing open data

eGOVERNMENT
CORE
VOCABULARIES

DCAT
APPLICATION
PROFILE FOR
EUROPEAN
DATA PORTALS

ADMS
ASSET
DESCRIPTION
METADATA
SCHEMA

ADMS
F/OSS

CORE
PERSON
VOCABULARY

CORE
BUSINESS
VOCABULARY

CORE
LOCATION
VOCABULARY

CORE
PUBLIC
SERVICE
VOCABULARY

European Commission

Existing data standards should be used wherever possible when publishing open data:

- The DCAT Application Profile should be used to describe open data catalogue and datasets.
    - a native specification for new open data portals
    - as an export format for existing data portals to export the catalogue and datasets metadata

DCAT
APPLICATION
PROFILE FOR
EUROPEAN
DATA PORTALS

Existing data standards should be used wherever possible when publishing open data:

- The (Asset Description Metadata Schema) ADMS specification should be used to describe metadata and semantic standards.
    - ADMS based descriptions are already collected by the EC Joinup platform, which provides a federated, single point of access to over 2000 specifications from over 20 sources.

ADMS
ASSET
DESCRIPTION
METADATA
SCHEMA

European Commission
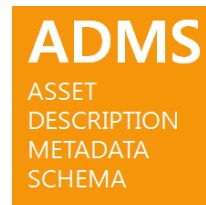
Multi-disciplinary
working group

57 people

16 Member States representatives

Wide range of backgrounds

Businesses

Governments Repository owners

Standardisation bodies

Academia Libraries

Statistics

7	Number of virtual meetings of ADMS Working Group
295	Number of messages on ADMS WG mailing list
> 2500	Number of accesses to ADMS versions on Joinup
15	Number of public comments
232	Number of JIRA issues tracked and resolved

ADMS as W3C Note

W3C Working Group Note

W3C®

Asset Description Metadata Schema (ADMS)

W3C Working Group Note 28 May 2013

This version:
http://www.w3.org/TR/2013/NOTE-vocab-adms-20130528/
Latest published version:
http://www.w3.org/TR/vocab-adms/
Previous version:
none
Editors:
Phil Archer , W3C / ERCIM
Gofran Shukair , DERI, NUIG
Author:
Makx Dekkers , AMI Consult (original document)

The vocabulary defined in this document is also available in these non-normative formats: RDF/XML and Turtle.

Copyright © 2013 W3C® (MIT, ERCIM, Keio, Beihang), All Rights Reserved. W3C liability, trademark and document use rules apply.

Abstract

ADMS is a profile of DCAT, used to describe *semantic assets* (or just 'Assets'), defined as highly reusable metadata (e.g. xml schemata, generic data models) and

European Commission
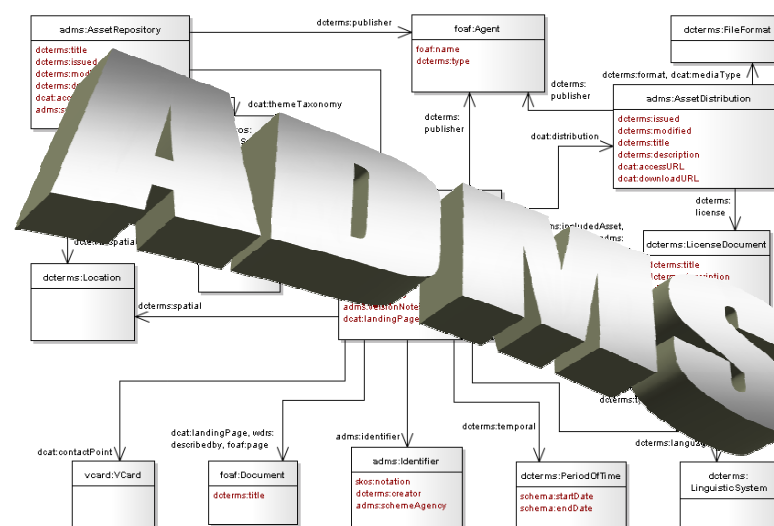
Existing data standards should be used wherever possible when publishing open data:

- The Core Location, Core Public Service, Registered Organization Vocabularies should be used to describe open data related to locations, public services, organizations and businesses respectively
    - These specifications are particularly suitable for publishing as open data, Public Sector Information which is stored in large base registries e.g. cadastral, one-stop public service portals, and business registries

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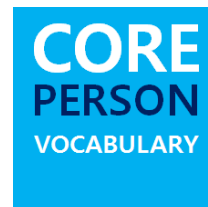
Core vocabularies

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

D3.1 – PROCESS AND METHODOLOGY FOR CORE VOCABULARIES

Deliverable

CORE PUBLIC SERVICE VOCABULARY

CORE PERSON VOCABULARY

CORE BUSINESS VOCABULARY

CORE LOCATION VOCABULARY
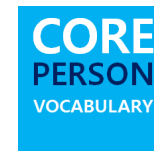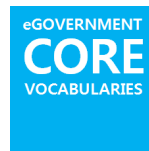
DCAT APPLICATION PROFILE FOR EUROPEAN DATA PORTALS

JOINING UP GOVERNMENTS

Usage

- New systems: As default starting points for data modeling

- Existing systems:

  - As reference data models for integration and information sharing (e.g. through mappings)

  - As export specifications for publishing open data

European Commission

Multi disciplinary working group

67 people

21 Member States and the US, South Africa and Croatia

+
EU institutions

Standardization bodies

External experts/academia

Core Vocabularies Working Group

W3C methodology

CORE LOCATION VOCABULARY
Core Location TF: chaired by EC Joint Research Centre/H6 (INSPIRE Directive)

CORE BUSINESS VOCABULARY
Core Business TF: chaired by DG MARKT (European Business Registry project)

CORE PERSON VOCABULARY
Core Person TF: chaired by EUROJUST

European Commission

Core Vocabularies are available through the Join.up platform

RDF   XML

All Core Vocabularies have been endorsed by the MSs in the context of the ISA Coordination Group

Open License

ISA Open Metadata License v1.1
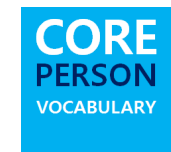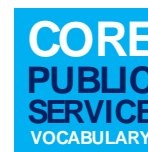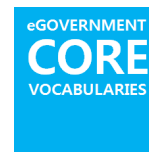
https://joinup.ec.europa.eu/category/licence/isa-open-metadata-licence-v11

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

Registered Organization Vocabulary

Core Business Vocabulary

W3C Note

W3C Working Group Note 28 May 2013

This version:
http://www.w3.org/TR/2013/NOTE-vocab-regorg-20130528/
Latest published version:
http://www.w3.org/TR/vocab-regorg/
Previous version:
http://www.w3.org/TR/2013/WD-vocab-regorg-20130108/
Editors:
Phil Archer, W3C / ERCIM
Marios Meimaris , National Technical University of Athens (Invited Expert) (from 1 March 2013)
Agisilaos Papantoniou , National Technical University of Athens (Invited Expert) (until 28 Feb 2013)

The vocabulary defined in this document is also available in these non-normative formats: RDF/XML and Turtle.

Abstract

The Registered Organization Vocabulary is a profile of the Organization Ontology for describing organizations that have a national or regional register.

European linked ~~National~~ Open Data…

Public Sector Information: The story so far…

- **National Closed Documents**
- **National Closed Data**
- **National Open Data**
- **European Linked Open Data**

Common Metadata and Semantic Standards

From national close documents… to European linked open data

The ISA work on promoting open data standards

Discussion on the questions received

a/ Gouvernance de la régulation du web et sa normalisation :
 - L'ISA (ou un organe de la Commission) participe-t-elle aux travaux du W3C?
- Avez-vous connaissance d'états membres (EM) qui siègent au W3C?
- Avez-vous des recommandations quant à l'opportunité que les EM siègent directement au W3C ou contribuent activement à certains groupes de travail?

The European Commission is not a member organization in any standardization organizations. W3C participated as a subcontractor to ISA work. The list of Members of W3C is publicly available:
http://www.w3.org/Consortium/Member/List. There are examples of public sector entities (e.g. ministries).
It is positive for MSs to participate in international standardization activities in the area of government data standards.
W3C has already taken ISA work to their standardization process (ADMS, RegOrg)

b/ L'apport du web sémantique pour l'administration en général
Examples of real use of semantic technologies in EU institutions: ESTAT, DG SANCO, European Environmental Agency, ISA pilots with Core Vocabularies, the Linked Organization of Local Authorities.
Benefits of semantic technologies are listed in the following publications by the ISA Programme:
Study on Business Models for Linked Open Government Data - BM4LOGD, October 2013.
https://joinup.ec.europa.eu/node/72473
10 Rules for persistent URIs, March 2013.
https://joinup.ec.europa.eu/community/semic/document/10-rules-persistent-uris
Case study on how Linked Data is transforming eGovernment, March 2013.
https://joinup.ec.europa.eu/node/61404

c/ Avez-vous identifiés des jeux de données de référence ("key register data") dont la conversion en format linked data est prioritaire?
 - Données de référence géographiques, entreprises, adresses...
- Existe-t-il un projet au niveau européen visant à prioriser la conversion de certains jeux de données dans l'ensemble des EM?

The study on high-value datasets from EU institutions provides an example of prioritization - at EU context - to define high-value datasets, which could give ideas for prioritising at the national level. Study to be available soon in Joinup.
The ISA Programme investigates this area too through the ISA Actions on semantics and on access to base registers.
The new EC Communication "Guidelines on recommended standard licenses, datasets and charging for the re-use of documents" (to be published before summer) sets priority areas as: geospatial, earth observation and environment, transport, statistics, companies.
The Danish initiative on Open up Basic data prioritize on geospatial data, personal data, companies, addresses, and real properties
http://uk.fm.dk/publications/2012/good-basic-data-for-everyone/~/media/Publikationer/Imported/2012/Gode%20grunddata%20til%20alle/BasicData_UK_web_2012.10.08.ashx

d/ L'apport du web sémantique pour les entreprises privées :
 - Impact sur l'efficacité, productivité et compétitivité?
 - Impact sur leur présence/rayonnement sur le web?

 We don't have an overview on this topic. Nevertheless, in relevant events (e.g. SemTech) large companies present important benefits. McKinsey: +3-5 Trillions US$ per year from open data.

e/ Benchmark européen :
 - Selon vous, y-a-t-il des EM plus avancés sur ces sujets? Si oui, lesquels?
 - Quelles sont les freins, difficultés généralement rencontrés?
 - Suite à d'éventuels retours d'expérience, avez-vous des recommandations à faire quant à la gouvernance de l'adoption du web sémantique pour le secteur publique ? Qui pilote? Quels partenaires? Quelle durée?
UK is quite advanced in the area of Linked Data as concluded from the number of identified cases in the BM4LOGD study. Also Estonia, Belgium, Germany have reported work with semantics and data standards in general.
The BM4LOGD study also lists barriers. Amongst others: licensing, privacy, data and metadata standards, no clear business models, lack of investments.
ISA has just completed the survey "Requirements and existing solutions for metadata management and governance in EU Institutions and Member States" to be available soon in Joinup.

f/ Formation initiale et continue :
 - Avez-vous connaissance de filière de formations universitaires aux métiers liés au web sémantique?
 - Avez-vous des projets de MOOC sur le web sémantique? Cela vous semble-t-il une bonne idée? L'ISA serait-elle prête à soutenir un tel projet?
There are universities that provide summer schools on linked data and semantic web topics, e.g. The Polytechnical University of Madrid (UPM) http://sssw.org/
The training material of the Open Data Support project provide a suitable introduction.
http://training.opendatasupport.eu/
ISA prepares introductory material on semantics, to be availble on Joinup before summer.

g/ Aspects juridiques :

- L'ISA mène-t-elle une réflexion sur les licences d'utilisation des jeux de données en linked data (algèbre de licences associant les données à une licence pour un usage donné)

- Avez-vous des recommandations en la matière?

The EC Communication "Guidelines on recommended standard licenses, datasets and charging for the re-use of documents" will provide guidelines on recommended standard licenses (to be published before summer)

h/ Evolutions attendues :

- Avez-vous un regard sur les évolutions à venir concernant la standardisation et la généralisation des formats "linked data"?
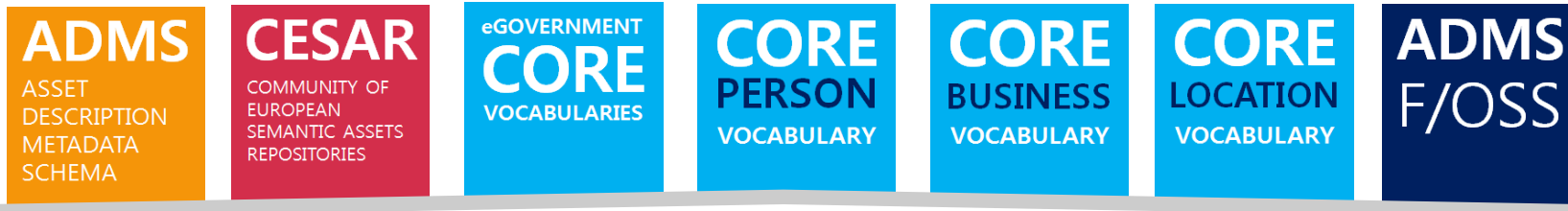
W3C promotes RDF as a well-established standard for publishing linked data.

Data standards are developed per domain of interest.

ISA contributes to this direction with the work on the DCAT-AP, Core Vocabularies and ADMS.

Questions & discussion...

**Join ISA initiatives at: http://ec.europa.eu/isa/
and @ http://joinup.ec.europa.eu**