



European
Commission

StatDCAT-AP

Virtual Meeting 1

11 March 2016

ISA Programme Action 1.1

Agenda

1. Opening
2. Practical arrangements
3. Welcome and introductions
4. Scope and objectives
5. Use cases
6. Issues related to mapping and extraction
7. Introduction to the deliverable
8. Time plan and next meetings
9. Any other business
10. Closing



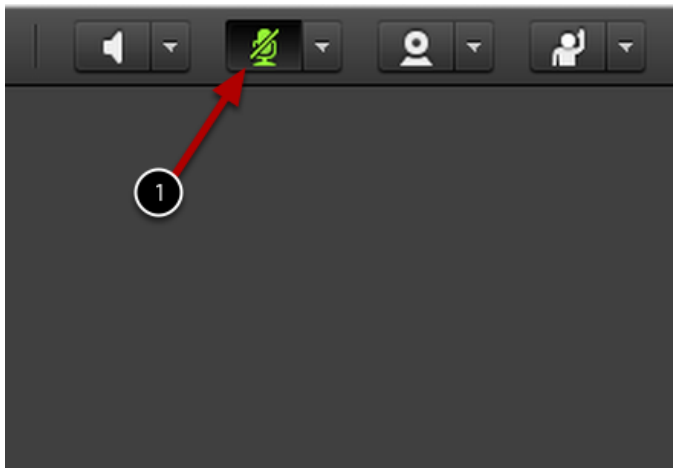
Practical arrangements

Practical arrangements

Join the meeting room:

- Go to http://ec-wacs.adobeconnect.com/stat_dcat/
- Click '**Enter as a guest**', fill in your name and click 'Enter meeting'
- Click '**ok**' to agree with the Web Conferencing Disclaimer
- Answer '**Start**' to the question: Would you like audio conferencing with this meeting?
- Enter your telephone number and have the system call you,
OR dial your [Local or International Access Number](#)
- Dial the Conference Room Number: *1323004#

Practical arrangements



Please:

- **Mute your mic** when not speaking
 - To mute your microphone click the Microphone Icon. When muted, the icon will remain green with a slash.
- **Raise your hand** to request the floor (and help making the speaker aware of who is raising hands)
- **Use the chat box** for sharing links and making comments
- **Voting:** use the chat box (+1 / -1 / abstain)

Collaboration Infrastructure

StatDCAT-AP release page

- Joinup page where the implementation guidelines are published:
https://joinup.ec.europa.eu/asset/stat_dcat_application_profile/description

Mailing list

- Publicly archived mailing list on Joinup used for submitting change requests on the draft specification:
stat_dcat_application_profile@joinup.ec.europa.eu

Issue tracker

- WG Members can log issues on the issue tracker or via the mailing list
- WG Members can comment on the issues that are already logged:
https://joinup.ec.europa.eu/asset/stat_dcat_application_profile/issue/all



Welcome and introductions



The work is funded under ISA Action 1.1 of the ISA Programme on improving semantic interoperability in European e-Government systems.



Find out more on <http://semic.eu>

Approach

- Involvement of stakeholders from both the statistical domain and the open data domain
- Building mutual understanding in open discussion
- Striving for consensus among stakeholders towards commonly agreed specification

- Working group led by Eurostat (Marco Pellegrino) and Publications Office (Norbert Hohn) with ISA and DG CONNECT
- Editor: Makx Dekkers, SEMIC team



Scope and objectives

Scope and objectives

- Enhance interoperability between descriptions of statistical data sets and general data portals
- Develop and reach consensus on an Application Profile of the Data Catalog Vocabulary (DCAT) for description of statistical data sets with an initial focus on discovery
- Elaborate initial guidelines on the extraction of relevant metadata from the existing implementation at Eurostat and others
- Propose extensions to DCAT-AP with descriptive elements particularly useful for discovery of statistical data sets

Longer term roadmap

First stage, until July 2016

1. Connecting descriptions of statistical datasets with general open data portals through a common basic exchange format StatDCAT-AP
2. Developing guidelines for the extraction of metadata from specific implementations of statistical standards towards StatDCAT-AP

Later stages

3. Harmonising implementations of statistical standards, possibly as an extension of the basic StatDCAT profile (for the metadata level) and through the use of W3C RDF Data Cube Vocabulary (for the data level)
4. Creating a set of tools to facilitate automatic extraction and validation of metadata from data described by statistical standards into StatDCAT-AP
5. Conducting practical pilots in various stages of the above activities to test and verify approaches and solutions.



Use cases

Rationale/scenarios

The statistics community and the data (portals) community are apart, lacking a common language for describing data

Scenarios:

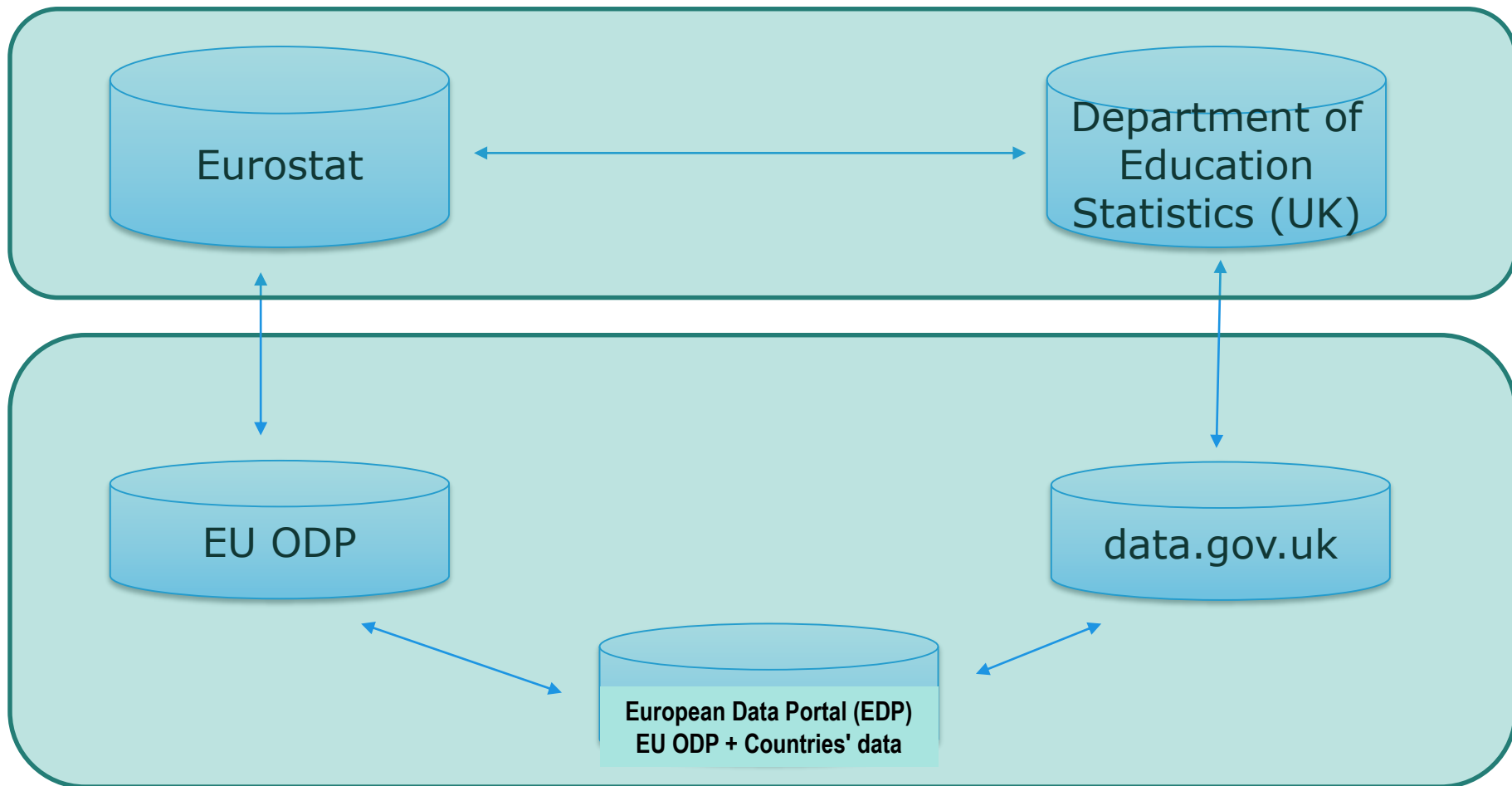
Simple exchange for retrievability and basic discovery

There is a need for mutual exposition of metadata descriptions in both statistical and data repositories in order to facilitate:

- Visibility (Retrievability), e.g. expose metadata of ECB statistics in EU Open Data Portal, and metadata of a data portal dataset in a statistical repository
- Basic discovery
- Exchange: Push-button export of StatDCAT-AP metadata from existing statistical (meta)data into data portals and vice versa, instead of ad-hoc mappings

Scenarios: Enhanced discovery thanks to an extension of DCAT AP by key statistical properties

- Enrich DCAT AP with key specific properties of statistical datasets (extension)
- Enhance discovery and filtering: Query DCAT AP based portals via key concepts of statistical metadata e.g. quality, dimensions





European Commission

National activities

Year	Spain	Italy	Other
2010	1.0000	1.0000	1.0000
2011	1.0000	1.0000	1.0000
2012	1.0000	1.0000	1.0000
2013	1.0000	1.0000	1.0000
2014	1.0000	1.0000	1.0000
2015	1.0000	1.0000	1.0000
2016	1.0000	1.0000	1.0000
2017	1.0000	1.0000	1.0000
2018	1.0000	1.0000	1.0000
2019	1.0000	1.0000	1.0000
2020	1.0000	1.0000	1.0000

eurostat
Your key to European statistics

18-year-olds in education

Additional information

Code: EDU18

Unit: %

Year: 2020

Country: EU

Number of values: 1

GOV.UK

Education and training statistics for the UK: 2013

Documents

Download

European Union Open Data Portal

18-year-olds in education

Publisher: Eurostat

Description: This indicator gives the percentage of all 18-year-olds who are still in any kind of school (all OECD months). It gives an indication of the number of young people who have not abandoned their efforts to improve their skills. Through initial education and it includes both those who had a regular education career without any delay, as well as those who are continuing even if they had to repeat some steps in the past.

Keywords: education, student

Basic Vocabulary: education, vocational training

GOV.UK

Education and training statistics for the United Kingdom

18-year-olds in education

Download

European Data Portal (EDP)
EU ODP + Countries' data



European Commission



Statistics (UK)

Education and training statistics for the UK: 2013

Education and training statistics for the UK: 2013

Documents

EDP

Education and training statistics for the United Kingdom

2013

2014

2015

2016

2017

European Data Portal (EDP)
EU ODP + Countries' data

National activities

National activities

Year	Spain	Italy	France
2010	100000	100000	100000
2011	100000	100000	100000
2012	100000	100000	100000
2013	100000	100000	100000
2014	100000	100000	100000
2015	100000	100000	100000

eurostat

Your key to European statistics

18-year-olds in education

Additional information

2013

2014

2015

2016

2017

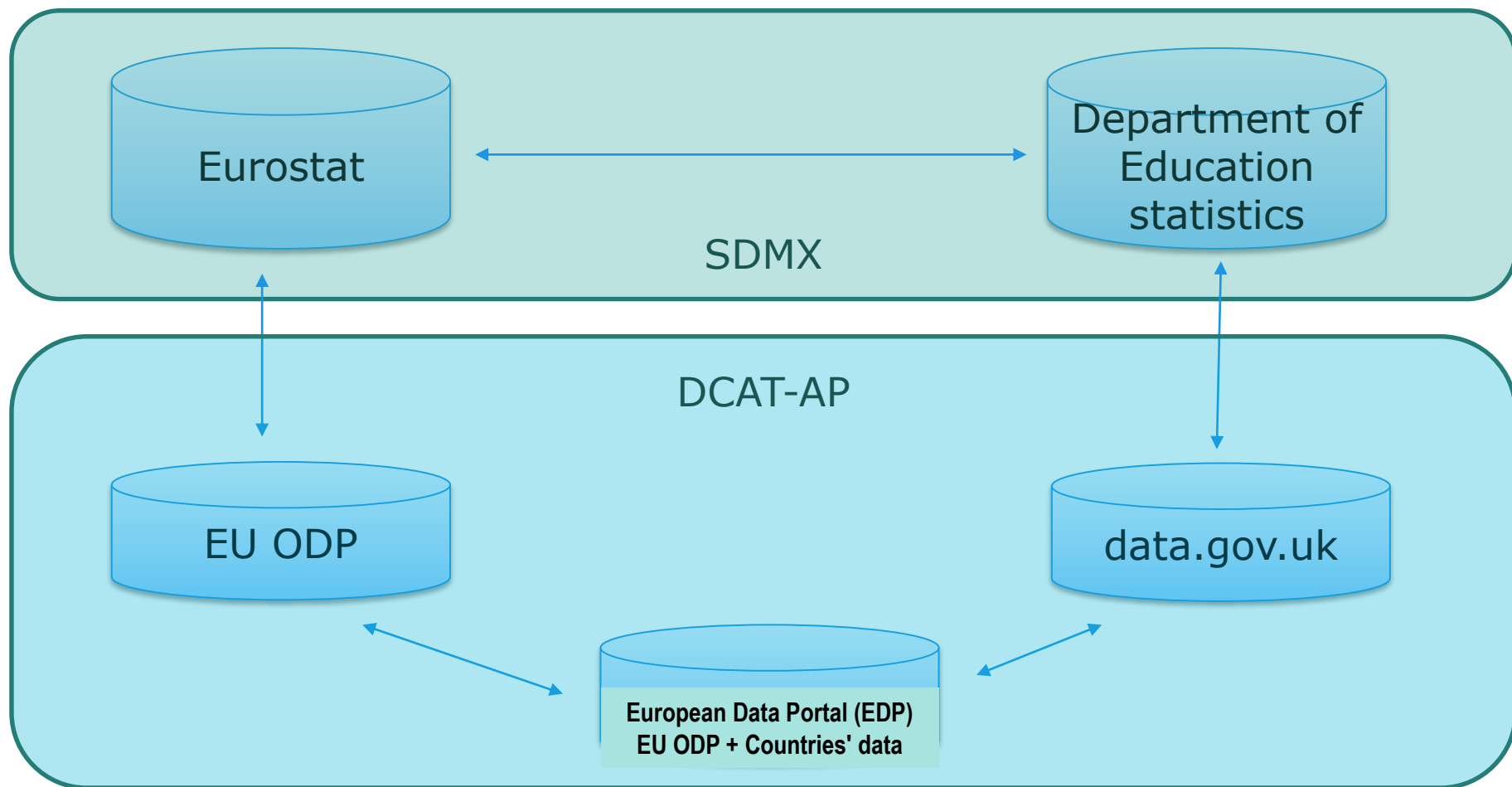
European Union Open Data Portal

18-year-olds in education

Keywords: education, student

Basic facts

Education and communications, Employment and working conditions



More use cases? Comments, questions?



Issues related to mapping and extraction

SDMX Concepts

Goal of the mapping

To produce a useful mapping of (Stat)DCAT-AP to typical statistical information.

Clarity is necessary about the precise goal of the mapping:

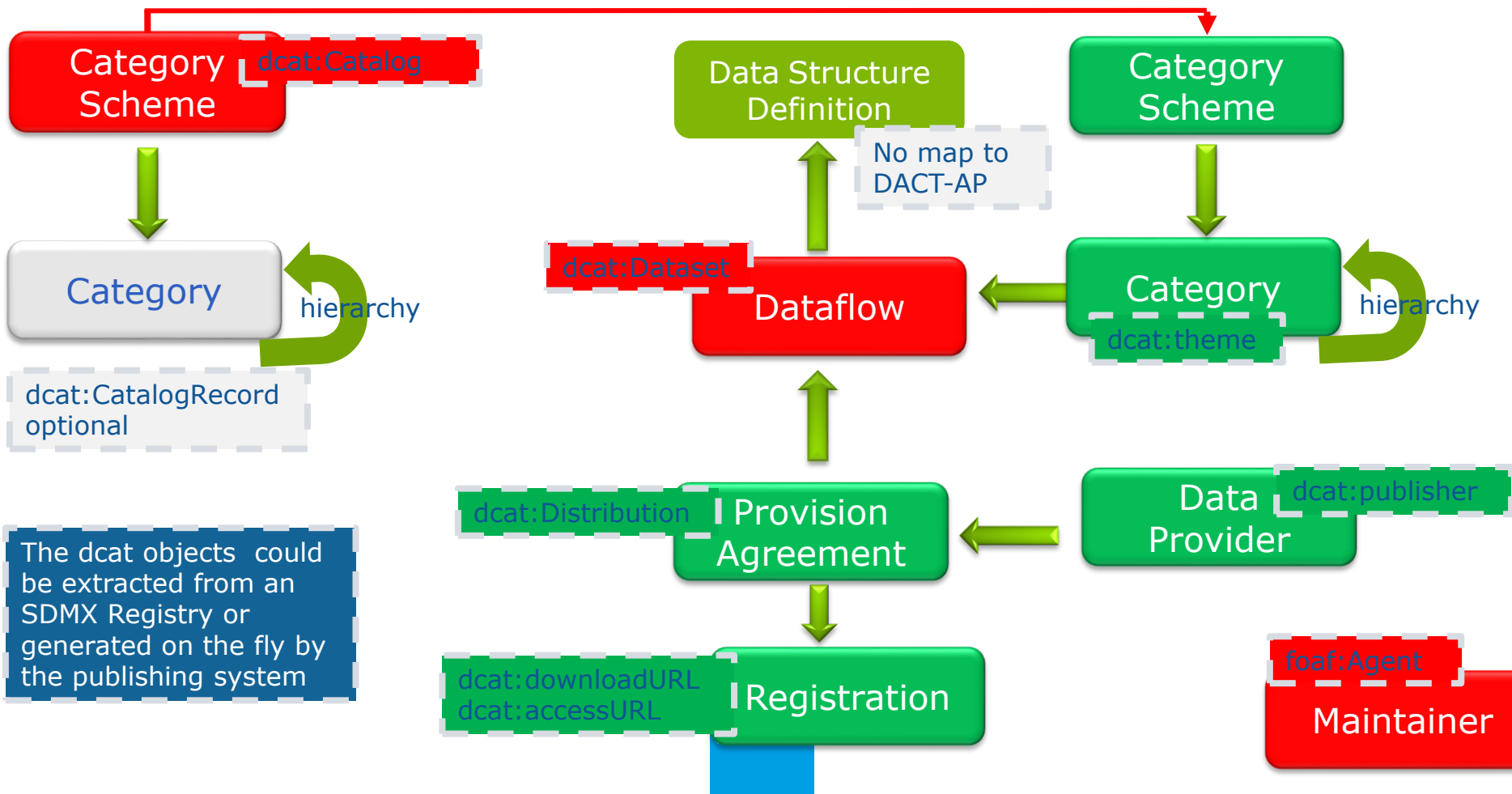
- Is the mapping to be **used at operational level today**, i.e. to produce DCAT-AP files immediately based on existing information?
- Is it intended to **guide longer-term joint implementations** of SDMX, DCAT and possibly other statistical standards?
- **Who is supposed to use the mapping?** Eurostat? National or regional statistical institutions? Other organisations?

Target of the mapping

Based on answers to the previous questions, one or several **target(s) of (Stat)DCAT-AP** mapping should be selected, such as:

- **SDMX information model:** Provides a standards-based mapping but does not necessarily correspond to effective implementations and thus not necessarily allow obtaining the required information.
- **Particular statistical dissemination systems:** Organisation-specific mapping(s) but would allow effectively obtaining required information from data providers in a shorter time horizon.
- Pivot model like the **Generic Statistical Information Model (GSIM)** mapping to DCAT, SDMX and particular statistical dissemination systems.

SDMX V 2.1 Structural Metadata Mapping to DCAT-AP – Mandatory and Recommended



Catalogs...

- Is it intended to have a single, harmonised, list of “themes” to support data discovery, or would each Catalog owner have its own scheme of “themes”
- Will there be a central catalog service?
 - Is it intended that there will be multiple catalogs linked to a central service that can search multiple catalogs in multiple locations?

Discovering statistical datasets

- What information is used to search for a statistical dataset?
 - via the dimensions
 - Via the measures
 - Via the attributes of the dimensions and measures
- Some arguments from the mailing list
 - Dimensions are valuable, provide semantical context
 - Measures seem to be the key items, however
 - they often are just value carriers
 - Semantics is provided by the dimensions

Discovering statistical datasets

Suggestion 1

- Add a property “statistical breakdown” that collects the used dimensions, measure, attributes in the dataset
 - It provides insight in the “breakdown”/“granularity” of the dataset
- Challenge: the range of the property
 - Examples:
 - sex, age range, commodity, financial instrument, ...
 - No complete, centralized list exists.
 - For Eurostat: could RAMON be such reference range list?

Discovering statistical datasets

Suggestion 2

- Refine /new property dcat:theme for the statistical domain
- The theming of the statistical domain (certainly in official statistics) is according to a different/more fine-grained than the MDR NAL data-theme
 - Question: is there a reference list available?
 - If yes, can we map the concepts of those to MDR NAL data-theme?
- Questions:
 - Shall we go for mapping (complete reuse), either for refinement (other range) or new property?

More issues? Comments, questions?



Introduction to the deliverable

Outline of the deliverable

1. Introduction
2. Terminology
3. Related work
- 4. Use cases and scenarios**
5. Methodology
6. Analysis and decision framework
- 7. Overview of data model**
- 8. Extraction guidelines**
- 9. Controlled vocabularies**
10. Conformance statement
11. Acknowledgements

Development of extraction guidelines

- Work on the extraction guidelines takes place at <https://docs.google.com/spreadsheets/d/1JPM0gM-MkM7o4FLItLczXWMR-R1-fRITvExfkE6oYIA/edit?pref=2&pli=1&hl=en#gid=0>



Time plan and next meeting

Tentative time plan

- **December 2015:** invitations to stakeholders, set up collaboration infrastructure
 - **January 2016:** collect requirements and suggestions
 - **5 February 2016:** Familiarisation Webinar
 - **February 2016:** first draft based on initial analysis and issues raised
 - **11 March 2016:** first virtual WG meeting to discuss first draft
-
- **March 2016:** second draft based on discussions and decision in the WG
 - **April 2016:** second virtual WG meeting to discuss second draft
 - **May 2016:** third draft, preparing for public review
 - **June 2016:** third virtual WG meeting to discuss schedule for public review
 - **July and August 2016:** public review period
 - **September 2016:** fourth virtual WG meeting, discuss and resolve public comments received, publication of StatDCAT-AP

Next meetings

Proposed periods for the next meetings:

- Monthly meetings
- Doodle polls will be shared to schedule meetings:
 1. Early April
 2. Early May
- Drafts of specifications to be shared ahead of meetings
- Discussions to take place through electronic means (mailing list, Joinup issue tracker)



Any other business

Save the date!



12 May 2016
Rome – Italy

Stay tuned at:

<https://joinup.ec.europa.eu/node/148436>

Any other business



Project Officers Vassilios.Peristeras@ec.europa.eu
Athanasios.Karalopoulos@ec.europa.eu

Visit our initiatives

ADMS ASSET DESCRIPTION METADATA SCHEMA	StatDCAT-AP FOR STATISTICAL DATASETS	GeoDCAT-AP FOR GEOSPATIAL DATASETS	DCAT-AP FOR DATA PORTALS IN EUROPE	CORE PUBLIC ORGANISATION VOCABULARY
CORE PERSON VOCABULARY	REGISTERED ORGANISATION VOCABULARY	CORE CRITERION & EVIDENCE VOCABULARY	CORE LOCATION VOCABULARY	CORE PUBLIC SERVICE VOCABULARY

Get involved



Follow [@SEMICEu](#) on Twitter



Join the [SEMIC](#) group on LinkedIn



Join the **SEMIC** community on Joinup