



2016-03-11
StatDCAT-AP Virtual Meeting 1
SEMIC Phase 7
Meeting Minutes

Date: 12/04/2016



Workshop: StatDCAT-AP Virtual Meeting 1

Venue	Adobe Connect	Meeting date	2015-03-11
Author	Stefanos Kotoglou	Meeting time	09:30 – 12:15
Reviewed by	Makx Dekkers	Issue date	2015-03-16
Status	For review	Version	0.01

ATTENDEES

Name	Abbreviation	Organisation
Ana Barreda	AB	Spanish National Catalogue www.datos.gob.es (Spain)
Anastasia Dimou	AD	iMinds, rml.io (Belgium)
Agnieszka Zajac	AZ	Publications Office
Amandine Masuy	AM	IWEPS (Belgium)
Andrei Melis	AMe	DG CONNECT
Alan Vask	AV	Marketing and Dissemination Department (Estonia)
Bert Van Nuffelen	BN	TenForce (Belgium)
Cristina Miranda	CM	Spanish National Catalogue www.datos.gob.es (Spain)
Chris Nelson	CN	Publications Office - Metadata Technology Ltd.
Danny Delcambre	DD	European Commission
Denis Groflis	DG	European Commission
Daniele Rizzi	DR	European Commission
Gregor Boyd	GB	Scottish Government
Giacomo Gamba	GG	ISPAT (Italy)
Makx Dekkers	MD	Freelancer
Martial Menard	MM	European Commission – Sogeti

Marco Pellegrino	MP	European Commission
Norbert Hohn	NH	Publications Office
Nikolaos Loutas	NL	PwC EU Services
Rob Davidson	RD	Office for National Statistics (UK)
Stefano Abruzzini	SA	DG CONNECT
Stefanos Kotoglou	SK	PwC EU Services
Soren Roug	SR	European Environmental Agency
Uros Milosevic	UM	TenForce
Valentina Janev	VJ	Institute Mihajlo Pupin
Vassilios Peristeras	VP	ISA Programme, European Commission
Willem Van Gemert	WG	Publications Office

AGENDA

ID	Description
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 and next steps

1. OPENING

MP opened the session, and informed the participants that the meeting minutes will be made available on [Joinup](#).

2. PRACTICAL ARRANGEMENTS

MD provided all the practical information regarding the communication and the collaboration between the members of the Working Group. The members of the Working Group will use a mailing list for discussion. All the information is available in the [slide-deck of the presentation](#).

- All the material (i.e. presentation of the webinar and meeting minutes) is available at the [webpage of the meeting](#).
- StatDCAT-AP webpage: <https://joinup.ec.europa.eu/node/147940>

3. WELCOME AND INTRODUCTIONS

MP mentioned that the concrete objective of the work is to develop and reach consensus on an Application Profile of the Data Catalog Vocabulary ([DCAT](#)) to be used for the description of statistical data sets with an initial focus on discovery of those data sets in a wider context.

MP invited everyone to make a self-presentation, providing more information about their organisation and their interest in StatDCAT-AP.

The following table provides an overview of the received input:

Table 1 Round table

Abbreviation	Description
AB	Working for the Spanish National Catalogue (www.datos.gob.es).
AD	Working for the Data Science Lab (former MMLab), Ghent University – iMinds.
AZ	Working for the Publications Office.
AM	Working for the IWEPS (the Walloon institute of statistics and evaluation research). They do not have any experience with DCAT-AP so far, but they use a standardised metadata form for all their indicators. They have developed a portal named 'WalStat' (http://www.wastat.iweps.be) based on open sources and open data. Their goal is to update their level of 'openness' through the use of the StatDCAT.
AMe	Working for DG CONNECT.
AV	Working for Marketing and Dissemination Department in Estonia.
BN	Working for Tenforce. He has been involved in the development of StatDCAT-AP since the beginning.
CM	Working for the Spanish National Catalogue (www.datos.gob.es).

CN	Working for Metadata Technology Ltd. He does not have practical experience of using DCAT-AP.
DR	Working for the European Data Portal. He is also involved in geospatial datasets, and he worked for the extension of DCAT to GeoDCAT-AP.
DD	Working for Eurostat on dissemination of metadata.
DG	Working for Eurostat as an enterprise architect.
GB	Working as statistician at the Scottish Government. They are currently developing an open data platform for Scottish statistical data. It uses DCAT for the dataset metadata although they don't currently implement all aspects of it. The Beta version of the system is available at: http://statisticsbeta.com/ .
GG	Working for the regional statistics institute of Trento, Italy, Data dissemination unit.
JD	Working at euroCRIS (http://www.eurocris.org) where they are trying to advance interoperability in the domain of research information. They are interested in StatDCAT-AP from the perspective of research dataset re-use and research indicators. He is also affiliated with the Charles University in Prague.
MD	Working for the SEMIC team. He is the editor of this work.
MP	Working for Eurostat. He is co-chairing the Working Group with NH.
MM	Working as an IT consultant at Eurostat. She is involved in the project of the renovation of the Eurostat Dissemination Chain, in charge of semantic aspects of the project, among other tasks. She will attend the meetings of the working group as an observer but she will not participate actively. Their goal is to anticipate and to plan the necessary actions in order to integrate the outcome of StatDCAT-AP in this renovation.
NL	Working for PwC. He is the project manager, assisting the EC.
NH	Working for the Publications Office. He is co-chairing the Working Group with MP.
RD	Working at the Office for National Statistics (UK). The Office for National Statistics publishes spreadsheets in support of their statistical bulletins but these are not linked to standardised metadata. Each publication does have a descriptive page of text that helps it to be indexed by the website search engine and/or external search engines. There is keen desire to link data to common data specifications in order to improve findability and increase reuse etc.
SR	Working at the European Environmental Agency. They have converted the SDMX and the ESMS formats to RDF in the past. He is interested in contributing.
SA	Working for DG CONNECT.
SK	Working for PwC. He provides assistance to the semantic interoperability team of the EC.
UM	Working for Tenforce, together with BN.

VJ	Working for PUPIN, Serbia. They were involved together with Tenforce and other research centres in designing and implementing the LOD2 Statistical Workbench. One of the outcomes was to automatic publish from the Workbench to the CKAN.
VP	Programme Officer of ISA and responsible for the SEMIC project.
VM	Working for PUPIN, Serbia. They were involved together with Tenforce and other research centres in designing and implementing the LOD2 Statistical Workbench. One of the outcomes was to automatic publish from the Workbench to the CKAN.
WG	Working for the Publications Office.

MP presented the approach that will be followed at this work:

- 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

MP provided more details regarding the key roles of this work:

- Working group led by Eurostat (MP) and Publications Office (NH).
- ISA and DG CONNECT are observers.
- Editor: MD and the SEMIC team

4. SCOPE AND OBJECTIVES

MD presented the scope and the objectives of StatDCAT-AP:

- 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.

MD presented the roadmap for this work which is divided in two stages:

First stage, until July 2016

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

Later stages

- 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).
- Creating a set of tools to facilitate automatic extraction and validation of metadata from data described by statistical standards into StatDCAT-AP.
- Conducting practical pilots in various stages of the above activities to test and verify approaches and solutions.

5. USE CASES

NH explained by presenting couple of scenarios that the main goal of this work is to improve:

- the findability of statistical data; and
- the exchange of statistical data between data portals.

NH invited the participants to provide their use cases and their feedback before the next virtual meeting of the Working Group.

MD mentioned that the participants could provide use cases based on their experiences.

MD mentioned that the specification (DCAT-AP) is flexible, and there is no limitation in the format that data can be published and/or harvested. The data can be in RDF, in excel etc.

6. ISSUES RELATED TO MAPPING AND EXTRACTION

DG mentioned that the goal of the mappings is to map different standards and enhance the catalogue with elements that re describing statistical datasets.

DG mentioned that during the preliminary work on mapping DCAT-AP with SDMX, there were some issues/questions that arose:

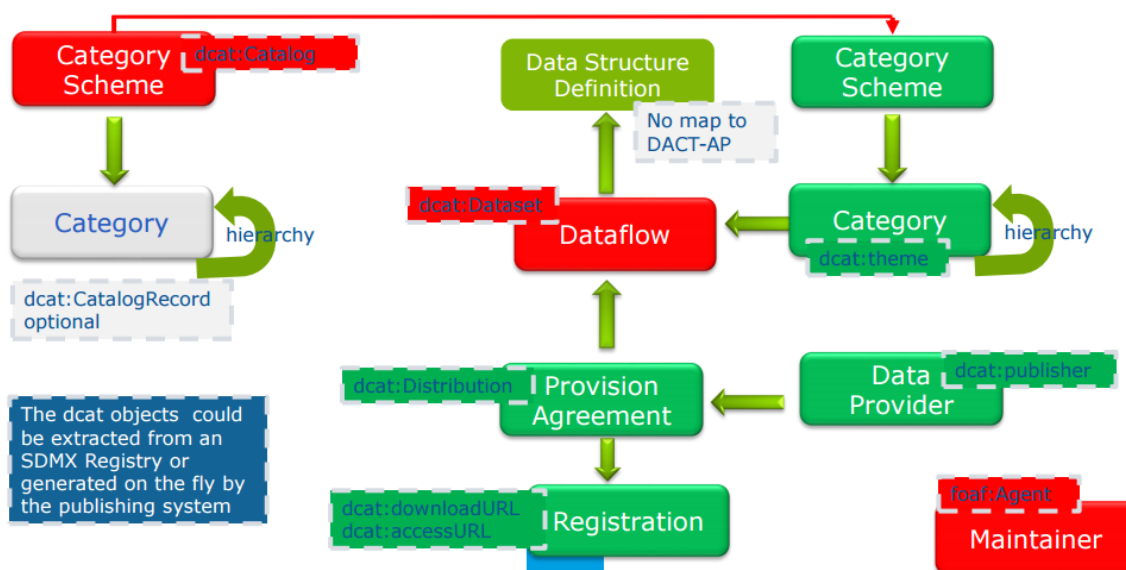
- Is the mapping to be used at operational level today, i.e. to produce DCAT-AP files immediately based on existing information?
- Is it intended to guide longer-term joint implementations of SDMX, DCAT and possibly other statistical standards?
- Who is supposed to use the mapping? Eurostat? National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

DG mentioned that 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;
- Particular statistical dissemination systems;
- Generic Statistical Information model (GSIM).

CN presented the structural metadata mapping of SDMX to DCAT-AP (the dashed boxes are elements of DCAT-AP):

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



BN mentioned that we need to collaborate with Working Groups from SDMX for producing generic mappings.

CN mentioned that the goal of StatDCAT-AP is to create a harmonised scheme that can be used by everyone in the domain of statistics. Organisations are not using a harmonised scheme, but they are using their own category schemes.

NL mentioned that DCAT-AP allows local category schemes, and recommends the use of the Data Themes NAL. However, once one needs to federate the datasets in one point, e.g. the European Data Portal, the local category schemes need to be aligned with the [common one](#), to allow a unique view over the datasets.

NH agreed with NL, and mentioned that we should have a common scheme with a list of themes.

Bert mentioned that each organisation has a list of themes and domains. Every time we agree in data exchanging, there is the need for identifying the data we share.

MP suggested documenting and sharing with the group all the issues.

BN mentioned that regarding the issue and the discussion that was created via the mailing list on “discovering statistical datasets”, the key item for searching for statistical datasets is the measures.

BN presented two suggestions for “discovering statistical datasets”:

1) Add a property “statistical breakdown”

This property collects used dimensions, measure and attributes in the dataset. The main challenge is the range of the property.

BN mentioned that there is the need for creating an open ended controlled vocabulary (RAMON could be a reference range list).

2) Refine /new property dcat:theme for the statistical domain

BN explained that we should reuse the “theme” in a particular category.

BN invited the participants to discuss the suggestions via the mailing list.

MD proposed having open discussion via the mailing list regarding the range of the property.

7. INTRODUCTION TO THE DELIVERABLE

MD informed the participants that the overall goal is the creation of a deliverable. A [draft version of the specification](#) is available on Joinup.

MD drew attention on the fact that the Working Group should build consensus on the data model. The guidelines will be developed based on the agreed data model.

MD mentioned that one of the objectives of this meeting is to collect requirements, and create issues for discussion on Joinup.

MP informed the participants that currently we are developing [mapping between SDMX and DCAT-AP](#) for extracting implementation guidelines.

8. TIME PLAN AND NEXT MEETINGS

MP mentioned that there will be a public review period starting in June.

MP mentioned that doodle polls will be distributed amongst the members of the Working Group for defining the date of the next Working Group meeting.

9. ANY OTHER BUSINESS

VP invited the participants to the [SEMIC 2016](#) and the back-to-back [workshop on DCAT-AP](#). The two meetings will take place on 12 and 13 May 2016 in Rome.

VP invited the participants to provide their ideas about the setting of the workshop

CLOSING – NEXT STEPS

ID	Description	Owner	Due date
1.	To send out doodle polls for the next meeting (mid-April)	PwC	20/03/2016
2.	To share a draft of the specification via Joinup	PwC	20/03/2016
3.	To provide use cases and feedback before the next virtual meeting of the Working Group	WG	08/04/2016
4.	To document and share the issues with the WG	PwC	08/04/2016
5.	To provide ideas about the setting of the workshop on DCAT-AP	WG	11/04/2016