2016-05-13

StatDCAT-AP Meeting 4

SEMIC Phase 7

Meeting Minutes

Date: 20/05/2016

Workshop: StatDCAT-AP face to face and virtual meeting			
Venue	Sheraton Parco de' Medici Rome Hotel & Adobe Connect	Meeting date	2016-05-13
Author	Stefanos Kotoglou	Meeting time	13:00 – 15:00
Reviewed by	Makx Dekkers	Issue date	2016-05-20
Status	For review	Version	0.03

1. ATTENDEES

Name	Abbreviation	Organisation
Alan Vask	AV	Marketing and Dissemination Department
Andrea Perego	AP	Joint research Centre – European Commission
Athanasios Karalopoulos	AK	European Commission
Bert Van Nuffelen	BN	TenForce
Chris Nelson	CN	Metadata Technology Ltd.
Daniele Rizzi	DR	European Commission
Denis Groflis	DG	European Commission
Deirdre Lee	DL	Derilinx
Hubertus Cloodt	HC	Eurostat – European Commission
Jakub Klimek	JK	University of Economics in Prague and Ministry or Interior of the Czech Republic
Jim J. Yang	JJY	Agency for Public Management and eGovernment (Difi)
Makx Dekkers	MD	Member SEMIC team, editor
Marco Combetto	MC	Informatica Trentina S.p.A.
Marco Pellegrino (co-chair)	MP	Eurostat - European Commission

Martial Menard	MM	European Commission – Sogeti
Nikolaos Loutas	NL	PwC EU Services
Norbert Hohn (co-chair)	NH	Publication Office of the European Union
Paolo Starace	PS	Sciamlab
Peter Winstanley	PW	The Scottish Government
Pierre Dumas	PD	Swiss Federal Archives SFA
Stefanos Kotoglou	SK	PwC EU Services
Uros Milosevic	UM	TenForce
Valentina Janev	VJ	Institute Mihajlo Pupin
Willem Van Gemert	WG	Publications Office of the European Union

AGENDA

ID	Description
1.	Opening, agenda, tour de table
2.	Objective of the meeting
3.	StatDCAT-AP overall characteristics
4.	Proposed extensions
5.	SDMX-based transformation mechanism
6.	Next steps

1. OPENING, AGENDA, TOUR DE TABLE

MP presented the agenda for the meeting, and invited all the participants to present themselves. A full list of all the participants of the working groups' meetings is available on Joinup.

2. OBJECTIVES OF THE MEETING

MD presented the intended outcomes:

- Discuss and agree on potential extensions.
- Discuss and agree on SDMX-based transformation. SDMX is the reference standard for this work.
- Decide on the content of the specification.

3. StatDCAT-AP overall characteristics

MD listed the basic requirements that StatDCAT-AP should fulfil:

- StatDCAT-AP should be a fully conformant extension of DCAT-AP;
- The "core" of StatDCAT-AP should be understandable for general portals.

MD mentioned that the general data portals can enhance their services by processing the additional properties in StatDCAT-AP; after all, statistical datasets are an important collection for portals.

4. Proposed extensions

MD mentioned that in total, eight issues were raised via the mailing list, and documented on the issue tracker.

- ***Number of observations***

CN and DG mentioned that this number cannot be calculated, and it is not available in statistical systems.

MD proposed **not including** the "Number of Observations" in the extension of StatDCAT-AP. This was agreed by the participants.

- ***Number of data series***

CN mentioned that this data can easily be generated by data sources, and it is already available in some data portals.

CN mentioned that data series, unlike observations, do not include the time dimension.

NL and UM mentioned that for avoiding language issues, improving discoverability and enhancing reality check, this information should be machine readable.

MD proposed including a **new property for data series with type "text"**, and invited the participants to provide their feedback via the issue tracker.

- ***Link to visualisation***

NH mentioned that Eurostat develops widgets for visualising datasets.

NH and NL proposed using the related "resource" property in case the provider includes the source of the dataset.

MD proposed **keeping the issue open**, and invited the participants to provide their feedback.

- ***Dimension as keywords* and *Dimension as a property***

BN mentioned that the intension is to provide a loose semantics insight of the structure.

BN suggested keeping the issue open by investigating the possibility for generating a property that can be component-oriented.

The participants agreed not to express the "Dimension as keywors".

MD proposed **keeping the "Dimension as a property" issue open**, and invited the participants to provide their feedback.

- ***Quality aspects***

Quality is an important aspect, but given the time constraint, the working group should decide whether a quality aspects extension will be included in the first version of the StatDCAT-AP, or in the second version (i.e. version 1.1, or version 2.0). MD proposed **keeping the issue open**.

- ***Statistical population* and *Statistical unit***

MP mentioned that both the issues are of high priority.

MD proposed **keeping the issues open**, and invited the participants to provide their feedback on Joinup.

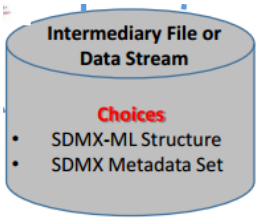
- ***New issue:*** "Length of time series"

MD proposed creating a new issue: "Length of time series".

NL inquired whether coverage and time series serve the same purpose. This may be discussed on Joinup.

5. SDMX-BASED TRANSFORMATION MECHANISM

CN mentioned that there are two different ways for describing metadata which lead to two different transformation mechanisms.



If an organisation prefers to use a tool to create DCAT-AP then the development of two tools are under consideration:

- SDMX structural metadata to DCAT-AP
- SDMX metadata set to DCAT-AP

CN listed down the advantages and the disadvantages of using the two different transformation mechanisms:

Table 1 Transformation mechanisms: advantages and disadvantages

Transformation mechanism	Advantages	Disadvantages
SDMX structural metadata (MSD)	• Familiar to organisations using SDMX • Can be generated easily from an SDMX Registry	The XML can be complex and verbose: • Annotations cannot be coded, hierarchical, validated or given mandatory and optional status. • Could create unnecessary "noise" when exchanging structural metadata with other organisations
SDMX metadata set	• Simple XML structure. • Attributes can be assigned any type of representation, validated, hierarchical, and the	• Not always well understood by SDMX users (may result in

	usage status can be mandatory or optional. • The Attribute Set can reference any object that can be identified. • It does not affect the structural metadata components. • If present, a Metadata Attribute can be "presentational", just giving structure to child attributes	some reluctance to use this mechanism). • Not widely used.

CN provided a summary about the transformation mechanisms:

- Transformation mechanism is **optional.**
- SDMX structural metadata (MSD) will suit organisations using SDMX, especially those using an SDMX Registry
- SDMX Metadata Set will suit organisations wishing to output STAT-DCAT metadata to a simple XML format, especially those not familiar with RDF and RDF vocabularies
- Metadata destined for DCAT-AP will need to be validated prior to generating DCAT-AP – ids, properties, references etc.
    - SDMX MSD can be used here irrespective of the intermediary format used.

CN mentioned that the validation is executed against the metadata structure definition (MSD).

CN mentioned that a registry is not able to perform validations on the annotations, unless the validator is built in the registry.

CN made clear that in SDMX a URL is a URI.

The participants decided to focus on MSD as the transformation mechanism. Also, it was decided that the transformation mechanism will be moved to an annex in the specification.

5. NEXT STEPS

MD listed the **next steps for the next months**:

- *MP to bring the proposal for using MSD as transformation mechanism to the SDMX specialists group.*
- *MD: to move the implementation sections to annex explaining that these solutions are for SDMX implementers that are migrating to StatDCAT-AP.*
- *WG and the Publications Office to map the MDR data themes to Eurostat themes (due by 3 June, to be presented in the next WG call).*
- *To include the possible extensions with proposal for RDF properties.*

FUTURE PLANNING

- End of May 2016: third draft, including full mapping proposal and usage of controlled vocabularies.
- 3 June 2016: fourth virtual WG meeting to agree schedule for public review
- July and August 2016: public review period.
- Mid-September 2016: fifth virtual WG to discuss and resolve public comments received.
- End of September 2016: approval of StatDCAT-AP version 1 for publication.