**2016-02-05**

**StatDCAT-AP Webinar**

**SEMIC Phase 7**

Meeting Minutes

|  |  |  |  |
| --- | --- | --- | --- |
| Webinar: StatDCAT-AP | | | |
| Venue | Adobe Connect | Meeting date | 2016-02-05 |
| Author | Stefanos Kotoglou | Meeting time | 10:30 – 12:30 |
| Reviewed by | Makx Dekkers | Issue date | 2016-02-09 |
| Status | Accepted | Version | 1.00 |

# Attendees

|  |  |  |
| --- | --- | --- |
| Name | Abbreviation | Organisation |
| Amandine Masuy | AM | IWEPS (Belgium) |
| Athanasios Karalopoulos | AK | ISA Programme, European Commission |
| Bert Van Nuffelen | BN | Tenforce (Belgium) |
| Chris Nelson | CN | Metadata Technology Ltd (UK) |
| Daniele Rizzi | DR | European Commission |
| Danny Delcambre | DD | Eurostat, European Commission |
| Denis Grofils | DG | Eurostat, European Commission |
| Gregor Boyd | GB | Scottish Government |
| Jan Dvorak | JD | euroCRIS (Czech Republic) |
| Makx Dekkers | MD | Independent Consultant |
| Marco Pellegrino | MP | European Commission |
| Martial Menard | MM | Sogeti (Luxembourg) |
| Nikolaos Loutas | NL | PwC EU Services |
| Norbert Hohn | NH | Publications Office |
| Rob Davidson | RD | Office for National Statistics (UK) |
| Soren Roug | SR | European Environmental Agency |
| Stefano Abruzzini | SA | European Commission |
| Stefanos Kotoglou | SK | PwC EU Services |
| Uros Milosevic | UM | Tenforce (Belgium) |
| Valentina Janev | VJ | Institute Mihajlo Pupin (Serbia) |
| Vassilios Peristeras | VP | ISA Programme, European Commission |
| Vuk Mijović | VM | Institute Mihajlo Pupin (Serbia) |
| Willem Van Gemert | WG | Publications Office |

# Agenda

|  |  |
| --- | --- |
| ID | Description |
|  | Practical arrangements |
|  | Welcome and roll call |
|  | Introduction of the participants - Round table |
|  | Scope and approach |
|  | Rationale/scenarios |
|  | Earlier work and existing solutions:   * Work done in 2015 * DCAT-AP/Data Portals ecosystem * SDMX/ESMS |
|  | Q&A |
|  | Next meetings |

# Practical arrangements

SK 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](https://joinup.ec.europa.eu/sites/default/files/isa_field_path/2016-02-05_statdcat-ap_webinar.pdf).

* All the material (i.e. presentation of the webinar and meeting minutes) is available at the [webpage of the meeting](https://joinup.ec.europa.eu/asset/stat_dcat_application_profile/event/webinar-statdcat-application-profile).
* StatDCAT-AP webpage: <https://joinup.ec.europa.eu/asset/stat_dcat_application_profile/description>

# Welcome to roll up call

VP welcomed everyone to the Webinar for the StatDCAT-AP.

VP mentioned that the objective of this work is to create an extension of DCAT-AP for describing statistical datasets for increasing their discoverability.

# Introduction of the participants - Round table

MD 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 |
| 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. |
| AK | Programme Officer of ISA and responsible for the SEMIC project. |
| BN | Working for Tenforce. He has been involved in the development of StatDCAT-AP since the beginning. |
| CN | Working for Metadata Technology Ltd. He does not have practical experience of using DCAT-AP. |
| DR | Working for SciamLab, Italy SciamLab has developed Open Data Hub Italy, the richest open data portal available in Italy, which is DCAT compliant. |
| DD | Working for Eurostat. |
| DG | Working for Eurostat. |
| 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/> |
| JD | Woking 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 the European Commission. |
| SK | Working for PwC. He provides assistance to the semantic interoperability team of the EC. |
| UM | Working for Tenforce, together with BN.3 |
| 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. |

# Scope and approach

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 provided the roadmap for this work:

***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.

MD mentioned that for developing the StatDCAT-AP, the focus is on the involvement of stakeholders from both the statistical domain and the open data domain.

MD 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

MD informed the participants that the work that will be delivered will be openly licensed. A [contributor agreement](http://joinup.ec.europa.eu/mailman/archives/stat_dcat_application_profile/2016-February/000025.html) has been shared via the mailing list.

# Rationale and Scenarios

NH mentioned that there is a lack of a common language for describing data between the statistics community and the data portals community.

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.

DR inquired whether the scope of this work should become more broad and general.

BN mentioned that the long term goal is to develop a generic specification that could be also used for implementations covering more general and bigger range.

VP mentioned that avoiding duplication in the harvesting process for general purpose portals is out of the scope for the time being.

# Earlier work and existing solutions

## Work done in 2015

Answering to the main driving question: *“what information is being disseminated in the current data exchanges within the statistical ecosystem that can be related to DCAT-AP?”* BN provided a summary of the preparation activities that were performed in 2015:

* analysis of data exchange formats (e.g. [SDMX](http://ec.europa.eu/eurostat/data/sdmx-data-metadata-exchange)); and
* analysis of statistical data portals (e.g. [Eurostat](http://ec.europa.eu/eurostat), [OECD](http://www.oecd.org/belgium/))

BN mentioned that after the performed analysis, we realised that there is the need for terminological framework for bridging the statistical with the data portal communities.

BN explained that [ESMS](http://ec.europa.eu/eurostat/data/metadata/metadata-structure) is structured human readable information, whereas DCAT-AP is machine readable metadata. These factor create two issues:

* High overlap for trivial info: title, description, contact person, publisher; and
* Growing gap (from some possible mapping to no correspondence at all) for contextual info.

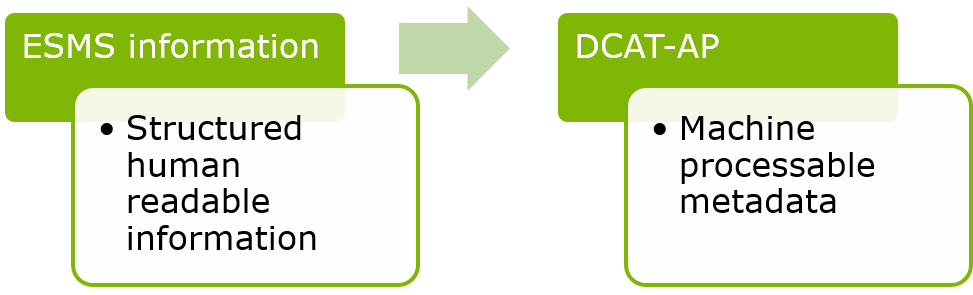


Figure 1 ESMS and DCAT-AP

BN presented the results after mapping SDMX with DCAT-AP (see Figure 2).

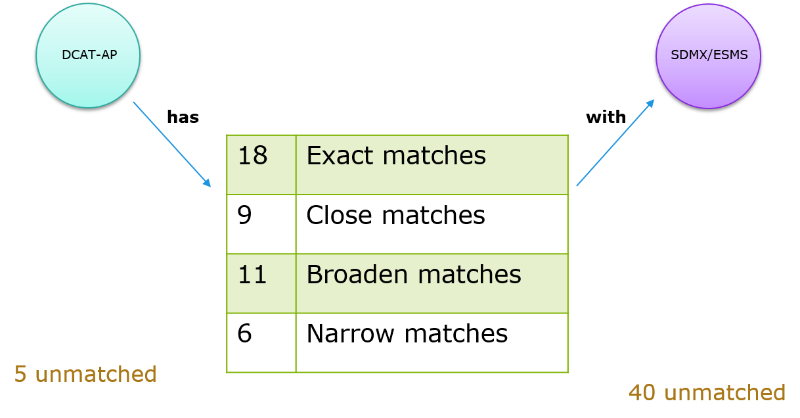


Figure 2 Mapping DCAT-AP to SDMX

BN presented the outcomes of the work that was done in 2015:

* Only one unique property has been spotted (the number of observations as measure of the size).
* Some properties might get a more generic definition in the statistical context (geographical & temporal coverage are specific cases of dimensions)
* There are common practices for grouping establishing relationships between datasets.
* [DataCube](https://www.w3.org/TR/vocab-data-cube/) and DDI RDF notation provide metadata structures for describing data sets.
* SDMX data cube structure definitions offer information that can be exploited by DCAT-AP.
* ESMS is not as-is transformable into DCAT-AP.
* Statistical data portals overlap highly with DCAT-AP information.

## DCAT-AP/Data Portals ecosystem

MD presented [DCAT-AP](https://joinup.ec.europa.eu/asset/dcat_application_profile/description):

* DCAT-AP is an Application Profile of the W3C DCAT Recommendation to be used for the exchange of descriptions of datasets between data portals in Europe.
* [DCAT](https://www.w3.org/TR/vocab-dcat/) was developed by the Government Linked Data Working Group at W3C in 2012-2013 as an RDF vocabulary designed to facilitate interoperability between data catalogues published on the Web.
* DCAT-AP was developed by the SEMIC activity under the ISA programme in 2013 and revised in 2015 for specific use in Europe, among others to support the European Data Portal.

The latest version of the DCAT-AP specification can be found at: [DCAT-AP v1.1](https://joinup.ec.europa.eu/asset/dcat_application_profile/asset_release/dcat-ap-v11).

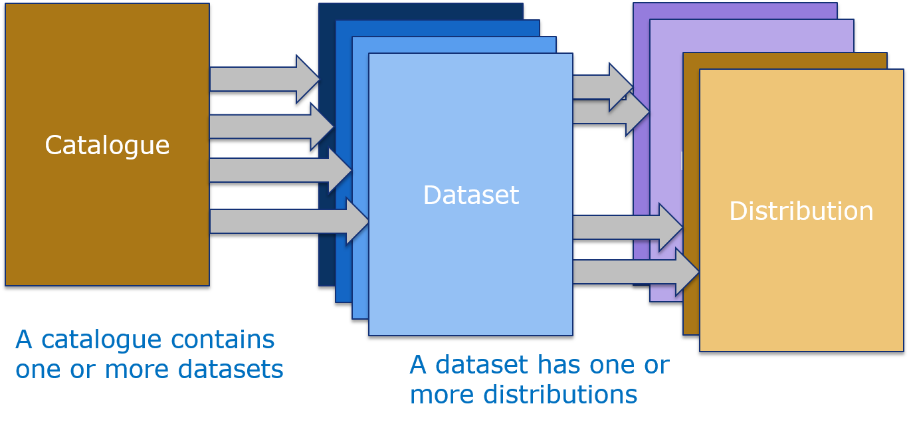


Figure 3 DCAT-AP model overview

MD listed the main aspects of DCAT-AP:

* DCAT-AP provides a common target for exchange of metadata
* It is applicable across domains as it does not limit the kinds of datasets that can be described
* Its objective is to support exchange of metadata for the main purpose of discoverability
* As such, it only describes the characteristics of datasets that are relevant for cross-domain discovery
* Additional characteristics for datasets in particular domains can be specified in extension profiles (e.g. [GeoDCAT](https://joinup.ec.europa.eu/asset/dcat_application_profile/asset_release/geodcat-ap-v10), StatDCAT)

MD presented [GeoDCAT-AP](https://joinup.ec.europa.eu/asset/dcat_application_profile/asset_release/geodcat-ap-v10):

* It is an extension of DCAT-AP for describing geospatial datasets, dataset series, and services.
* Its basic use case is to make spatial datasets, data series, and services searchable on general data portals, thereby making geospatial information better searchable across borders and sectors.

## SDMX/ESMS

MP presented SDMX:

* The Statistical Data and Metadata Exchange (SDMX) initiative is sponsored by seven institutions (the BIS, the ECB, Eurostat, the IMF, the OECD, the UN and the World Bank) in order to provide standards for the exchange of statistical information.
* SDMX describes and universalises the way to exchange statistical data, and provides standard formats for data and metadata, content guidelines as well as IT architecture for exchange of data and metadata.
* SDMX provides:
* A logical model to describe statistical data, together with guidelines on how to structure the content
* A standard for automated communication from machine to machine
* A technology supporting standardised IT tools that can be used by all parties involved in the data exchange and processing

CN provided a more technical presentation regarding the uses of SDMX. SDMX can be used for searching by:

* Data provider
* Concept
* Code
* Category

CN presented the mapping between SDMX and StatDCAT-AP (available [here](https://joinup.ec.europa.eu/sites/default/files/isa_field_path/2016-02-05_statdcat-ap_webinar.pdf)).

# Q&A

SA inquired whether we should specify the different types of metadata, and their use.

CN mentioned that in SDMX are two types of metadata:

* Code schemes; and
* Reference metadata

CN mentioned that dct:Standard could be included in SDMX by using reference data.

MP suggested creating a central registry for StatDCAT-AP (SDMX could cover this function).

WG mentioned that DCAT-AP authority tables are published in the metadata registry of the Publications Office.

MP and SA agreed on the fact that there is the need for agreement regarding the uses cases to be addressed.

# Next meeting

Next meeting will take place in the second week of March:

* <http://doodle.com/poll/yu759si5s4wsru75>

# Action Points

|  |  |  |  |
| --- | --- | --- | --- |
| ID | Description | Owner | Due date |
|  | To fill the doodle poll for the next meeting:  <http://doodle.com/poll/yu759si5s4wsru75> | WG members | 01/03/2016 |
|  | To share questions and experiences via the mailing list: stat\_dcat\_application\_profile@joinup.ec.europa.eu | WG members |  |
|  | To log new issues via the issue tracker: <https://joinup.ec.europa.eu/asset/stat_dcat_application_profile/issue/all> | WG members |  |

# Chat log

