



2016-11-14
StatDCAT-AP Meeting 6
SEMIC Phase 8
Meeting Minutes

Date: 01/12/2016



Webinar: StatDCAT-AP virtual meeting

Venue	Adobe Connect	Meeting date	2016-11-14
Author	Florian Barthélemy	Meeting time	13:00 – 15:00
Reviewed by	Makx Dekkers	Issue date	2016-12-01
Status	Accepted	Version	0.04

1. ATTENDEES

Name	Abbreviation	Organisation
Florian Barthélemy	FB	PwC EU Services
Gregor Boyd	GB	Scottish Government
Edoardo Dalla Mutta	EDM	Unknown
Makx Dekkers	MD	Member SEMIC team, editor
Anastasia Dimou	AD	imec/iMinds Data Science Lab, Belgium
Guillaume Duffes	GD	INSEE (French Statistical Institute)
Denis Grofils	DG	European Commission, Eurostat
Dragan Ivanovic	DI	European Commission, Eurostat
Valentina Janev	VJ	Institute Mihajlo Pupin, Serbia
Nikolaos Loutas	NL	PwC EU Services
Martial Menard	MM	European Commission
Chris Nelson	CN	Metadata Technology Ltd., UK
Bert Van Nuffelen	BN	TenForce, Belgium
Marco Pellegrino(chair)	MP	European Commission, Eurostat
Daniele Rizzi	DR	DG CONNECT
Søren Roug	SR	European environment agency
Alan Vask	AV	Marketing and Dissemination Department, Statistics Estonia
Agnieszka Zajac	AZ	Publications Office of the EU

AGENDA

ID	Description
1.	Opening, agenda, tour de table
2.	Objectives of this meeting
3.	Discussion of open issues
4.	Next steps

2. 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](#).

3. OBJECTIVES OF THIS MEETING

MP presented the objectives of the meeting:

- Discuss and agree on the resolution of the issues created based on the comments received during the public review period.
- Define next steps and finalise the StatDCAT-AP specification.

4. DISCUSSION OF OPEN ISSUES

MD said that 20 issues were raised during the public review period, submitted by five commenters, from the National Statistics Institutes of France and Norway, Open Data in the Czech Republic, the Permanent Representation of Denmark to the European Union and the Ministry of Finance of Brazil. All issues have been published on Joinup. The issues are divided in four groups:

- Issues for which we propose changes (Accept/Reject)
- Issues for which we propose the inclusion of clarifications for implementers
- Issues that need further work (outside the scope at the time being)
- Additional issues/errors that will be corrected (editorial issues, typos)

4.1. Proposed changes

Issue 1: Definition & label of property stat:statMeasure

1. Link: <https://joinup.ec.europa.eu/node/156473/>
2. Issue: The property stat:statMeasure is defined as "unit of measurement" not as SDMX/Data Cube `measure`.
3. Proposal:
 - Rename statMeasure to statUnitMeasure.
 - Add a property that is equivalent with SDMX/Data Cube measure.
4. Proposed resolution:
 - Accept proposal, rename property to stat:statUnitMeasure.
 - Consider adding 'measure' property for future revisions of StatDCAT-AP.

-
5. Discussion:
MD mentioned that the second point of the proposed solution could be moved to future work.
CN said that we need to be careful when solving this because there are many measures existing.
 6. Decision: The proposed resolution was agreed.

Issue 2: Property `stat:statMeasure` on Distribution

1. Link: <https://joinup.ec.europa.eu/node/153937/>
2. Issue: The property `stat:statMeasure` is a characteristic of the Dataset and applies to all Distributions.
3. Proposal:
 - Define the property on Distribution to allow different Distributions to use different units of measurement.
4. Proposed resolution:
 - Reject proposal.
 - Respond with clarification that attributes, dimensions and measures are 'conceptual' entities that apply to all distribution formats, e.g. in case a dataset is provided both in SDMX and in Data Cube, and therefore fit better on the Dataset level.
5. Discussion: No comment was raised.
6. Decision: The proposed resolution was accepted.

Issue 3: Properties `stat:dimension` and `stat:attribute` on RDF Distribution

1. Link: <https://joinup.ec.europa.eu/node/153934/>
2. Issue: On RDF Distribution `stat:dimension` and `stat:attribute` could be replaced by link to `qb:DataStructureDefinition`.
3. Proposal:
 - Use `dct:conformsTo` to link to `qb:DataStructureDefinition` instead of `stat:dimension` and `stat:attribute` for RDF Distributions.
4. Proposed resolution:
 - Reject proposal.
 - Respond with clarification that using the `stat` properties for all types of distributions preserves coherence across formats; a RDF distribution may use `dct:conformsTo` in addition.
5. Discussion:
MP thought about providing clarifications on this issue, previous and following ones.
CN agreed with the proposed resolution.
AD mentioned that `dct:conformsTo` has `dct:Standard` as range, if we want to be consistent, `qb:DataStructureDefinition` should be subclass of `dct:Standard`.
6. Decision: The proposed resolution was accepted. Clarifications will be provided to support the resolution of issues 2, 3, 4.

Issue 4: Definition of `numSeries` property

1. Link: <https://joinup.ec.europa.eu/node/156471/>
2. Issue: Definition is not formally correct.

3. Proposal:

- Define as: "Cartesian Product of the number of modalities of each dimension, excluding what Data Cube calls the measure dimension (that denotes which particular measure is being conveyed by the observation)". The numSeries is the actual number of series in the data set as referenced in the Distribution. This is usually less than the theoretical number calculated as the Cartesian Product (and sometimes significantly less). The actual number of series is, when combined with the dimension list, a useful indication of the detail of the data in the data set.
- Proposed resolution:
- Accept proposal.
- Change definition in specification.

4. Discussion:

MD proposed to take the first sentence as definition and the second as a description.

CN said that we could just keep one adapted definition.

GD who had raised this issue agreed that the second sentence could be used as definition.

CN pointed out that the biggest difference was about the 'usually less than the theoretical number'.

MD proposed to change the 'usually less' to 'maybe less'.

MP agreed and said that last part could be used as context. Others agreed.

MP added that than the support from the WG to create awareness around this was required.

MD added that a review will be necessary after last comments will have been addressed.

5. Decision:

- The definition will be: "Cartesian Product of the number of modalities of each dimension, excluding what Data Cube calls the measure dimension (that denotes which particular measure is being conveyed by the observation). The numSeries is the actual number of series in the data set as referenced in the Distribution." And the context will consist of: "The actual number might be less than the theoretical number calculated as the Cartesian Product. The actual number of series is, when combined with the dimension list, a useful indication of the detail of the data in the data set."
- A review of the definition will be proposed.

Issue 5: Choice between xsd:date and xsd:dateTime

1. Link: <https://joinup.ec.europa.eu/node/156467/>

2. Issue: Specification allows both xsd:date and xsd:dateTime

3. Proposal:

- Make a clear choice between the two based on statistical use cases.

4. Proposed resolution, one of:

- Leave both options in specification; preserves maximum interoperability with DCAT-AP-compliant systems as those already need to support both options.

-
- Choose `xsd:dateTime`; requires indication of time always (possibly default e.g. 00:00:00 or 23:59:59 if not known).
→ Note `xsd:date` cannot be chosen as this does not support publication more than once per day.
5. Discussion:
- MD proposed to keep the current situation.
For CN, it could be `xsd:date` or `xsd:dateTime`, but the important thing is to be clear on what we are using in the XML.
From a mutual agreement, the WG decided to let things as they are.
MP added that there is a need to provide some comments on the specification.
MD will provide this.
6. Decision: The proposed resolution was accepted as to know to keep the current specification. Comments will be provided to support this resolution.

4.2. Clarifications

Issue 6: stat:attribute/dimension versus qb:attribute/dimension

1. Link: <https://joinup.ec.europa.eu/node/156469/>
2. Issue: why are these properties defined in the `stat:` namespace and does StatDCAT-AP not use the Data Cube properties directly?
3. Proposed resolution:
 - Add clarification in the specification that the `qb:` properties are defined in the context of a `DataStructureDefinition` and that using them in a different context received negative advice from the Data Cube editors.
4. Discussion:

MP said that some more clarifications about why we are not using the Data Cube properties are needed.
MD agreed and clarified that the context is different between StatDCAT-AP and Data Cube.
5. Decision: Clarifications on the contexts between StatDCAT-AP and Data Cube will be provided to justify the reasons why StatDCAT-AP is not using the Data Cube properties.

Issue 7: Properties stat:dimension/attribute for non-RDF distributions?

1. Link: <https://joinup.ec.europa.eu/node/153933/>
2. Issue: can these properties, and StatDCAT-AP in general, only be used on RDF distributions?
3. Proposed resolution:
 - Add clarification in section 6.1 that StatDCAT-AP can be used to describe datasets in any format, including SDMX, Data Cube, CSV and other formats.
 - Add clarification in section 6.2.1 that these properties can be used irrespective of the format of the Distribution; add examples of how the values would be extracted from SDMX.
4. Discussion: No comment was raised.
5. Decision: The proposed resolution was accepted.

Issue 8: Specifying the length of time series

1. Link: <https://joinup.ec.europa.eu/node/153788/>
2. Issue: how can weekly and quarterly series be expressed?
3. Proposed resolution:
 - Add clarification in section 6.2.4 how the combination of temporal coverage and frequency can be used to express all types of time series.
4. Discussion: No comment was raised.
5. Decision: The proposed resolution was accepted.

Issue 9: Relationship StatDCAT-AP and Data Cube

1. Also issue **10** and **11**
2. Links: <https://joinup.ec.europa.eu/node/153935/>, <https://joinup.ec.europa.eu/node/156468/> and <https://joinup.ec.europa.eu/node/153932/>
3. Issue: How are dcat:Dataset, qb:DataSet, dcat:Distribution and qb:DataStructureDefinition related?
4. Proposed resolution:
 - Add full example of a StatDCAT-AP-compliant description of a Data Cube DataSet.
 - Respond that further advice may come out of the W3C SDSVoc workshop in Amsterdam 30 November-1 December 2016.
5. Discussion:

For MD, one issue is that StatDCAT-AP is not referring to Data Cube while the opposite is not true. This will be discussed in the coming workshop later this month.

MP asked whether the WG will be able to interact because the workshop is coming before the finalisation of the specification.

MD said that no decision will be taken during the workshop, it is only an opportunity to give some examples about those two datasets.

On the comment from DI 'qb:DataSet = an instance of dcat:Dataset?', MD mentioned that except from saying that the two are different and giving examples, we have to wait for W3C.
6. Decision: The proposed resolution was agreed. A full example of a StatDCAT-AP-compliant description of a Data Cube dataset will be provided. Further advices are required from W3C.

4.3. Further work

Issue 12: Relationship between StatDCAT-AP and DDI

1. Link: <https://joinup.ec.europa.eu/node/155606/>
2. Issue: How can DDI descriptions of datasets be expressed using StatDCAT-AP.
3. Proposed resolution:
 - Respond that contacts have been established with the DDI initiative, which may lead to development of an additional Annex in the future.
 - No change in the specification.
4. Discussion: No comment was raised.
5. Decision: The proposed resolution was accepted.

Issue 13: Full RDF expression of SIMS

1. Link: <https://joinup.ec.europa.eu/node/156476/>
2. Issue: Suggests development of a full RDF expression of SIMS.
3. Proposed resolution:
 - Respond that this is out of scope for the StatDCAT-AP activity but may be taken up by others.
 - No change in the specification.
4. Discussion:

MD said that a full expression of SIMS is a bit beyond the scope of the StatDCAT-AP working group.

GD mentioned that INSEE has some maturity on this and they could help further.
5. Decision: The proposed resolution was accepted.

Issue 14: SIMS as part of the StatDCAT vocabulary

1. Link: <https://joinup.ec.europa.eu/node/156474/>
2. Issue: Suggests that SIMS could be a part of StatDCAT-AP.
3. Proposed resolution:
 - Respond that this may be considered in future work in relation with more elaborate treatment of quality aspects.
 - No change in the specification.
4. Discussion: Future work could be how StatDCAT-AP could benefit from SIMS.
5. Decision: The proposed resolution was accepted.

Issue 15: Modelling SDMX Metadataflow in StatDCAT-AP

1. Link: <https://joinup.ec.europa.eu/node/156474/>
2. Issue: Suggests an extension to model the SDMX metadata flow in StatDCAT-AP.
3. Proposed resolution:
 - The subject of SIMS has been raised in Issues 13 and 14 and the relevance of support for the SDMX Metadataflow will be taken into account when assessing how best to expand the support for quality metadata in the next version of StatDCAT-AP.
 - No change in the specification.
4. Discussion: see discussion under issue 14.
5. Decision: The proposed resolution was accepted.

Issue 16: Visualisation for other types of datasets

1. Link: <https://joinup.ec.europa.eu/node/153936/>
2. Issue: Suggests that the approach to identify visualisations is also relevant for other types of data sets, not just statistical datasets.
3. Proposed resolution:
 - Respond that this suggestion will be considered in further development of DCAT-AP.
 - No change in the specification.
4. Discussion: No comment was raised.
5. Decision: The proposed resolution was accepted.

4.4. Editorial comments

Issue 17: Typo DVAT

1. Link: <https://joinup.ec.europa.eu/node/156475/>

Issue 18: Property schema:population in UML class diagram

1. Link: <https://joinup.ec.europa.eu/node/153938/>

Issue 19: Example of Visualisation

1. Link: <https://joinup.ec.europa.eu/node/153939/>

Issue 20: Missing namespaces

1. Link: <https://joinup.ec.europa.eu/node/153940/>
2. Proposed resolution for all the errors/editorial comments:
 - o All reported errors will be corrected.

5. NEXT STEPS

ID	Description	Owner	Due date
1.	Finalise the StatDCAT-AP specification by addressing the received comments from today's call.	ISA	18/11/2016
2.	Provide clarifications and comments for the following issues: 2, 3, 4, 5, 6, 7, 8, 9.	ISA	18/11/2016
3.	Provide a single file for comments to avoid raising comments on each issue on Joinup.	ISA	18/11/2016
4.	Participate to the W3C SDSVoc workshop in Amsterdam 30 November-1 December 2016 to collect new insights and examples on the relation between StatDCAT-AP and Data Cube.	ISA	01/12/2016
5.	Publish the specification on Joinup in December (14 th the latest).	ISA	14/12/2016

5.1. Next steps: share your ideas!

AZ: how should we follow up the work here in order to exchange some experience or comments? Could this work methodology be shared with e.g. Member States to spread awareness?

MP: we have started already: presentation during the WG sessions, conferences etc. and we will continue with this.

DR: The European Data Portal (EDP) has some room for work in their contracts.

NL: Some work has been done between geoDCAT-AP and EDP. Once the specifications for StatDCAT-AP will be finalised, the purpose is to do some pilots for StatDCAT-AP and EDP as well. See an example [here](#) with DCAT-AP implementation guidelines.

6. CHAT

The chat history has been cleared.

Martial Menard: hello

Chris Nelson: have connected to dial in to london but it tells me the host has not yet arrived! I used room number 1323004

Chris Nelson: I will try again

Marco Pellegrino: room 1323004 should be correct

Athanasios KARALOPOULOS: Chris, alternatively the system can call you in

Athanasios KARALOPOULOS: try this option too

Anastasia Dimou: Hi! Yes, I'm only on listening mode

Chris Nelson: it still says the host has not arrived

Chris Nelson: I prefer to use a headphone that's why I use Skype

Chris Nelson: OK will need to use the phone out option

Anastasia Dimou: Hi Anastasia from Ghent University, imec

Anastasia Dimou: (iMinds merged with iMinds so former iMinds)

Anastasia Dimou: with imec

Martial Menard: Hi, I am working on the renovation of the Eurostat Dissemination Chain. I gather requirements in this respect. Thanks

Martial Menard: Thanks Marco

Søren Roug, EEA: From European Environment Agency. I'm only on listening mode for the time being.

Anastasia Dimou: dct:conformsTo has dct:Standard as range

Anastasia Dimou: sure but if we want to be consistent then qb:DataStructureDefinition should be subclass of dct:Standard

Dragan Ivanovic: Are we speaking about the cardinality of the Cartesian Product $D_1 * \dots * D_n$, which is $|D_1| * \dots * |D_n|$?

Gregor Boyd: happy to leave as they are

OP - Agnieszka Zajac: leave it as it is

Anastasia Dimou: Implementing both of them it's not that difficult :)

Dragan Ivanovic: qb:DataSet = an instance of dcat:Dataset?

Chris Nelson: W3C can suggest that the DCAT dataset is renamed dataflow

OP - Agnieszka Zajac: @Nikos: what e-mail/contact point should we give for participation in the pilots you mentioned?

Nikos Loutas, PwC: contact@semic.eu or nikolaos.loutas@be.pwc.com

OP - Agnieszka Zajac: ok. thank you

Makx Dekkers: makx@makxdekkers.com

Valentin Janev (Institute Mihajlo Pupin): Valentina Janev, I am sorry for being late for 15min. Thanks to all participants for the discussion, Institute Mihajlo Pupin

Nikos Loutas, PwC: Indeed thank you very much fo your valuable contributions

OP - Agnieszka Zajac: Thank you for a great work

Makx Dekkers: thanks! we'll be in touch

Daniele Rizzi - DG CONNECT: Thank you Marco and all!

OP - Agnieszka Zajac: Bye

Gregor Boyd 2: thanks bye

Anastasia Dimou: bye

Martial Menard: Thanks you, bye