

# D05.01 20230202 Meeting Minutes: Webinar on the review of the Data Catalogue Application Profile

<b>Project:</b>	SEMIC: DCAT-AP	<b>Meeting Date/Time:</b>	02/02/2023 14:00 - 16:00
<b>Meeting Coordinator:</b>	Bert Van Nuffelen, Makx Dekkers	<b>Issue Date:</b>	17/03/2023

## Meeting Agenda

1. Welcome
2. Context DCAT-AP
3. Usage of Dataset Series & alignment I
4. Short break
5. Usage of Dataset Series & alignment II
6. Wrap-up & next steps

## Meeting Slides

[LINK](#)

## Participants

<b>Name</b>	<b>Initials</b>	<b>Organisation</b>
William Verbeeck	WV	SEMIC Team
Bert Van Nuffelen	BVN	SEMIC Team
Jitse De Cock	JDC	SEMIC Team
Anastasia Sofou	AS	SEMIC Team
Pavlina Fragkou	PF	SEMIC Team

Emidio Stani	ES	SEMIC Team
Makx Dekkers	MD	SEMIC Team
Florian Barthelemy	FB	SEMIC Team
Andrea Pörsch	AP	Helmholtz Metadata Collaboration
Anja Litka	AL	Division H II 5 Geoinformation   Federal Ministry of the Interior and Community
Anja Loddenkemper	AL	Landesamt für Geoinformation und Landesvermessung Niedersachsen, Koordinierungsstelle GDI-NI
Bart Hanssens	BH	OpenGov
Bernd Ahlborn	BA	Unknown
Camille Cloitre	CC	Council of European Municipalities and Regions
Carlos Carvalho	CaC	ECDC
Cassandra Fournot	CF	Health Data Hub
Ana Cochino	AC	European Medicines Agency
Danielle Welter	DW	Luxembourg National Data Service (LNDS)
David Robberechts	DR	Unknown
Katerina-Christina Deli	KD	European Medicines Agency
Licinio Kustra Mano	LKM	DG SANTE
Dr. Jesper Zedlitz Schleswig-Holstein	JZS	State Chancellery of Schleswig-Holstein
Eero Vegmann	EV	QuickBI
Eirini Christodoulou	EC	Post Doctoral Researcher
Persephone Doupi	PD	Finnish Institute for Health and Welfare
Geraldine Nolf	GN	Digital Flanders
Gudrun Kr. Gudfinnsdottir	GG	Icelandic Association of Local Authorities
Hagar Lowenthal	HL	Joint Research Centre
Hans Aage Huru	HAH	Norsk Helsenet SF

Irène Kesisoglou	IK	Sciensano
Jakov Vukovic	JV	Public Health Resident Croatia
Jakub Klimek	JK	Atlalogy
Jasper Beernaerts	JB	NGI Belgium
Javier Orozco-Messana	JOM	DG CNECT
Jeroen Belien	JEB	Health-RI
Pim Volkert	PV	Health-RI
Jill Saligoe-Simmel	JS	Esri
Joachim Nielandt	JN	Digital Flanders
JP Lovato	JL	Liip
Konstantins Bogucarskis	KB	Joint Research Centre
Ludger Rinsche	LR	]init[
Luis Alves de Sousa	LADS	European Centre for Disease Prevention and Control and other places
Manos Papageorgiou	MP	KIOS Research and Innovation Center of Excellence
Marc Hanauer	MH	Unknown
Marc Vanderperren	MV	Publications Office
Mari Mäkinen	MM	Finnish Institute for Health and Welfare (THL)
Marjan Meurisse	MaM	Sciensano
Matthias Palmer	MP	MetaSolutions
Michele Spichtig	MS	University of Applied Sciences of the Grisons FHGR
Mihai Paunescu	MiP	Publications Office
Norman Calleja	NC	Malta
Natan Cox	NaC	Cognizone
Nicolai Fog Hansen	NFH	Danish Agency for Digital Government
Nina Georgieva	NG	Unknown

Oystein Asnes	OA	Norwegian Digitalisation Agency
Pascal Derycke	PD	Sciensano
Pero Ivanko	PI	CIPH
Peter Lubrich	PL	Federal Highway Research Institute, Department for Connected Mobility, Germany
Philip Ashlock	PA	TTS Solutions
Pierre Smith	PS	Sciensano
Pooria	P	Unknown
Zsolt Peter Puskas	ZPP	HC eXpert Kft.
Simon Dutkowski	SD	FOKUS
Simone Cocchi	SC	Modula France
Simou Stefania	SS	European Medicines Agency
Sjaak Kempe	SK	Eucaris
Sonia Jimenez	SJ	International Data Spaces Association
Stavros Tsouderos	StS	Unknown
Steven Fletcher	SF	Unknown
Tala Haddad	TH	Orphanet
Thomas Tursics	TT	GovData
Truls Korsgaard	TK	Finnish Institute for Health and Welfare (THL)
Ulrika Domellöf Mattson	UDM	Digg
Willem van Gemert	WVG	Publications Office
Mantas Sakalas	MS	Baltic Institute of Advanced Technology (BPTI)

### Summary

14:00–14:05 Opening, slides 1-4  
14:05–14:10 Context, slide 5-9  
14:10–14:15 Status, slides 10-13  
14:15–14:45 Usage, slides 15-28  
14:45–15:00 Properties issues 1-3, slides 29-36

15:00–15:05 Break  
 15:05–15:50 Properties issues 4-9, slides 39-55  
 15:50–16:00 Wrap-up, closing, slides 56-60

## Points discussed and decisions taken

<b>Topic discussed</b>	<b>Outcome</b>
<b>Special Case 1</b>	Not approved <ul style="list-style-type: none"> <li>• Discussion moved to Github</li> </ul>
<b>Special Case 2</b>	Approved
<b>Special Case 3</b>	Not approved <ul style="list-style-type: none"> <li>• Discussion moved to Github</li> </ul>
<b>Topic 1 Descriptive information</b>	Not approved <ul style="list-style-type: none"> <li>• Discussion moved to Github</li> </ul>
<b>Topic 2 Identification</b>	Approved
<b>Topic 3 Life cycle</b>	Not approved <ul style="list-style-type: none"> <li>• Discussion moved to Github</li> </ul>
<b>Topic 4 Agent roles</b>	Approved
<b>Topic 5 Legal information</b>	Not approved <ul style="list-style-type: none"> <li>• Discussion moved to Github (Create github issue: relationship between series and datasets)</li> </ul>
<b>Topic 6 Dimensions</b>	Not approved <ul style="list-style-type: none"> <li>• Discussion moved to Github</li> </ul>
<b>Topic 7 Update evolution</b>	Approved
<b>Topic 8 Preferred navigation pattern</b>	Discussion moved to Github for the sake of time
<b>Topic 9 Membership of catalogue</b>	Discussion moved to Github for the sake of time

## Full meeting minutes

<b>Welcome</b>	PF welcomed the participants and presented the agenda.
<b>Context DCAT-AP</b>	BVN gave a status update on DCAT 3 and the context of DCAT-AP.
<b>Usage of Dataset Series &amp; alignment I</b>	<p>BVN explained the motivation for the usage of the Dataset Series.</p> <p>TK asked for an example of what is “into” the different “boxes”</p> <p>JZS shared an example on how we use dataset series in their portal (with the limited possibilities of DCAT-AP 1):  <a href="https://opendata.schleswig-holstein.de/dataset/windkraftanlagen-2022-07-01">https://opendata.schleswig-holstein.de/dataset/windkraftanlagen-2022-07-01</a></p> <p>MP disagreed with the formulation in the last slide "most recent version of the data". It indicates that the different datasets in a series are variants of each other. He thinks that is wrong, they complement each other, e.g. as a time series or a geographic division.</p> <p>JK agreed, stating that DCAT3 has previousVersion and nextVersion for that.</p> <p>To which MD replied that the most recent “edition” would be better than “version”.</p> <p><b>Special case 1 (slide 25)</b></p> <p>JZS stated that if only data is added to the dataset/distribution (and not all data is deleted or altered) we do not need a data series. It could be a dataset with one distribution and the temporal extent is changed.</p> <p>JK mentioned that maybe we could come up with a list of various Series types that could be used to further characterise the DatasetSeries? E.g. Time-based, Geo-based, Snapshots, etc.</p> <p>GN mentioned that ISO uses the following definition:  <i>Dataset series = collection of datasets sharing the same product specification.</i>  <i>Dataset = identifiable collection of data.</i>  <i>NOTE A dataset may be a smaller grouping of data which, though limited by some constraint such as spatial extent or feature type, is located physically within a larger dataset.</i>  <i>Theoretically, a dataset may be as small as a single feature or</i></p>

*feature attribute contained within a larger dataset. A hardcopy map or chart may be considered a dataset.*

LR agreed with JK about the list, as noted on GitHub and suggested creating a separate ticket and start collecting possible entries.

→ **list to be created on GitHub**

IK asked if data series could describe a data registry?  
BVN suggested taking this conversation offline.

LD asked if dataset series is more often used in the case of different “versions”.

### **Special case 2 (slide 26)**

General agreement.

### **Special case 3 (slide 27)**

GN stated that in ISO there are distributions in dataset series; but the semantics in ISO and DCAT of a distribution are different >> in ISO you can link to a higher level than the downloadable dataset as a whole.

NFH wondered what happens if the service has datasets that are not in the series?

To which BVN replied that their relationship never states you are only serving that information.

MD suggested to continue the discussion on GitHub.

MP mentioned to take into account the portal perspective.

### **Dataset Series property alignment (slide 28)**

JN asked if it is assumed that any property of the dataset series is applicable for the contained datasets?

To which MD replied that is not always the case, the publisher can be different for example.

LR wondered if in case of conflict we should recommend that (the more specific) value of the dataset “overwrites” the series?

MD replied that there is no formal way to express overwriting.

BVN suggested to continue the discussion on Github.

MP stated that the part "that is published separately" seems forced.

JZS said that a dataset series is not always ordered, e.g. a geographic collection.

GN agrees with JZS, for example: temporal, spatial,...

### **Principles for alignment**

LR stated that we should not assume that is a summary of shared characteristics.

MP mentioned that from the dcat3 specification: "Typically, this means that, for each of the relevant properties, the dataset series takes as value the union of those specified in child datasets."

LD asked if we can consider whether the datasets within a dataset series must have the same schema?

MD replied that there is no reason to require this.

BVN answered that this is an open topic.

### **T1 Descriptive information**

General agreement.

Some people have second thoughts about the description being mandatory → MD suggested moving the discussion to Github.

### **T2 Identification**

General agreement.

### **T3 Life cycle**

JK added that 'not maintained' implies deprecated. He wondered what happens if all the datasets in the series are alive, but their number simply does not change? e.g. series of regions.

MP asked why not to rely on accrualPeriodicity of the series itself?

N suggested again using the ISO 8601 duration for dynamic modifications.

JB and PV wondered how they should handle a dataset that is being archived, it is an action while you do not know the status if it is not maintained.



	<p>OA asked if changing the order of a dataset (by changing PREV/NEXT) will affect modification date for both the dataset and the Dataset Series.</p> <p>JN commented that the 'freshest' date would be duplicate information, deductible from the dataset information. He wondered if this statement is correct?</p> <p>MP agreed with JK and stated that the creation date + accrualPeriodicity should be enough for those that prefer it.</p> <p>LR suggested using the following link  <a href="http://purl.org/adms/status/UnderDevelopment">http://purl.org/adms/status/UnderDevelopment</a>.</p> <p>MD commented that if you add a dataset, you want to reflect it in the modification of the dataset series - Ideally provide the date of last modification.</p> <p>MP stated that accrualPeriodicity should be monthly/yearly.</p> <p>BVN asked if we consider that other organisations can add without notification/communication to the maintainer of the dataset series? (Ludger Rinsche indicated this)  MD suggested making a Github issue to further discuss this topic.</p> <p>MP added that the handling of the open world assumption in a harvesting environment is a challenging problem. His belief is that in general this kind of "injection" should be hindered as it would affect the trustworthiness of what you are seeing on portals.</p>
<p><b>Short break</b></p>	
<p><b>Usage of Dataset Series &amp; alignment II</b></p>	<p><b>T4 Agent roles</b>  JB and PV asked if dataset series can be part of another dataset series.  General agreement</p> <p><b>T5 Legal information</b>  JB and PV asked if legal information includes "terms of use"?  MD answered that a datasets series is not usable, it is the data in the datasets that is usable.</p> <p>MP's perception is that dataset series is not the sum of the datasets.</p>

BVN mentioned that it is not specifically stated in the specification, but some sentences are inclined to this.

→ Create github issue(relationship between series and datasets)

JK stated that in DCAT3, rights/licences SHOULD be provided for Distributions, even though it is allowed for Resources (Datasets). Since DatasetSeries has no distributions, there should be no legal information on it.

MH assumed it really could be depending on the way a resource organises "data series", for instance if I manage a catalogue with many dataset which could be "grouped" by countries, I may have specific legal information at the dataserie level (due to this grouping).

### **T6 Dimensions**

JZS wondered if "topic" is also a dimension or if there is another way to combine datasets that are needed to understand a particular topic like water quality for example.

BVN answered that in this case it is about evolution and that he will put the example as a new use case.

MP added that for instance, the series would not have any order. No next or prev. The datasets would be complementary, similar to geographical division. But the question is whether they could be said to "share some characteristics".

JN thought it is worrying that you can have a dataset series, defined according to a geographical attribute, and be able to sort it (because you need to be able to call 'prev' and 'next'). Based on time of course it makes sense.

MS asked why we need to duplicate temporal and spatial coverage if this can be calculated from the datasets in the dataset series.

MD replied that this is because otherwise there is no way to calculate it.

JB and PV asked how a temporal dataset series of spatial datasets fits in?

BVN replied that it can be both and MD said that the publisher needs to decide what view he/she wants the users to have.

Discussions moved to github

### **T7. Update evolution**

	<p>JK came up with a counterexample: Geo-based series - e.g. streets in town is a dataset, the series is for all towns. The series is not updated that frequently (towns not added frequently), but the streets in the individual datasets may change more often.</p> <p><b>General agreement</b></p> <p><b>T8. Navigation</b></p> <p>MD commented that in his understanding first is the oldest and last is the newest (regarding Dcat:first).</p> <p>LR added dcat:seriesMember.</p> <p>BVN asked if we agree that we should not use the prev and navigation properties as the only way to indicate membership of a dataset in a dataset series and we only rely on dcat:inSeries or dcat:inSeriesMember.</p> <p>Suggestion of MP: require dcat:inSeries</p> <p>MD moved the discussion to Github for the sake of time</p> <p><b>T9 Membership of Catalogue</b></p> <p>MD suggested to put all other discussions to GitHub</p>
<b>Wrap-up &amp; next steps</b>	The Session was wrapped up and everyone was thanked for their participation.