

20231004 Meeting Minutes: Webinar on DCAT-AP for data spaces

Project:	SEMIC	Meeting Date/Time:	04/10/2023 14:00 - 16:00
Meeting Coordinator:	Pavlina Fragkou Makx Dekkers Bert Van Nuffelen	Issue Date:	10/10/2023

Meeting Agenda

1. Welcome
2. The big picture: Challenges within and across data spaces
3. Solution
4. Testimonials & real-life use cases
5. DCAT-AP ecosystem & SEMIC Service Offering
6. Q&A

Meeting Slides

[LINK](#)

Participants

Name	Initials	Organisation
Ahmed Abid	AA	NTT DATA
Aitor Corchero	AC	NTT Data
Alberto Abella	AA	FIWARE
Alessio Dragoni	AD	SciamLab

Participants		
Name	Initials	Organisation
Ana Martin	AM	/
Anastasia Sofou	AS	SEMIC Team
Anders	/	/
Andrea Pasare	AP	SEMIC Team
Andrea Perego	AP	European Parliament
Andreas Cola	AC	Telicent Limited
Andreas Eisenrauch	AE	Amadeus
Andrzej Strug	AS	GUMed Poland
Anssi Ahlberg	AA	Finnish Digital Agency
Arthur Schiltz	AS	SEMIC Team
Bart De Lathouwer	ADL	Geonovum & Muni
Bart Hanssens	BH	DG Digital Transformation
Benjamin Witsch	BW	Austriatech
Bernard Claes	BC	DIGIT
Bernd Ahlborn	BA	/
Bert Van Nuffelen	BVN	SEMIC Team
Casper	/	/

Participants		
Name	Initials	Organisation
Cécile Guasch	CG	DIGIT
Charles-Andrew Vande Catsyne	CAVC	Sciensano
Charline Alexandre	CA	SEMIC Team
Christian Hassler	CH	MoH-AT
Cor Melse	CM	RIVM Netherlands
Costas Simatos	CS	DIGIT
Csongor Nyulas	CN	SEMIC Team
Dan Turley	DT	Taligent
Daniel Arosa Otero	DAO	NTT Data
Danielle Welter	DW	LNDS
Dominique Riefstahl	DR	Ministry of Digitalisation, Luxembourg
Ekatarina	/	/
Emiel Dhondt	EDH	SEMIC Team
Fabian Kirstein	FK	Fraunhofer
Francesco Cartella	FC	DIGIT
Gabriele Bozzi	GB	IMEC

Participants		
Name	Initials	Organisation
Geraldine Nolf	GN	Digital Flanders
Gilles Hermes	GH	Ministry of Digitalisation, Luxembourg
Giuditta Del Buono	GDL	Gaia-X
Hagar Lowenthal	HL	Publications Office
Hans Aage	HA	HURU
Hans Overbeek	HO	Publications Office, Netherlands
Harold Kroeze	HK	Central Bureau for Statistics, Netherlands
Helena	/	/
Helena Lodenius	HL	CSC
Hosni Akremi	HA	NTT Data
Hugo Lafuente	HL	NTT Data
Ilias	/	/
Inge Voldseth	IV	Norwegian Directorate of eHealth
Ioannis Chondropoulos	IC	Ministry of Environment, Greece
Irene Goni	IG	GRNET
Italo Mairo	IM	DIGIT

Participants		
Name	Initials	Organisation
Jaime Meléndez	JM	IACS
Jakub Klímek	JK	Digital Information Agency, Czech Republic
Javier Conde	JC	Polytechnic University of Madrid
Jésus Ibáñez	JB	IACS
Jesper Zelditz	JZ	State Chancellery, Germany
Jelte Bootsma	JB	TNO
Jil Saligoe-Simmel	JSS	Esri
Jim Yang	JY	Norwegian Digitisation Agency
Jitse De Cock	JDC	SEMIC Team
Joachim Fugleberg	JF	Norwegian Digitisation Agency
Joachim Nielandt	JN	Digital Flanders
Joeri Robbrecht	JR	DG ENV
Johan Groenen	JG	Ministry of Justice and Security, Netherlands
Johan Mastenbroek	JM	/
Julian Rojas	JR	IMEC
Karen Dijkstra	KD	Federal Chancellery, Switzerland

Participants		
Name	Initials	Organisation
Katerina Moutogianni	KM	DG CNECT
Katharina Schneider	KS	BfArM
Katherina	/	/
Kerstin Engelhardt	KE	Directorate of e-health, Norway
Licínio Kustra Mano	LKM	DG SANTE
Lina Molinas Comet	LMO	Fraunhofer FIT
Luca Lianas	LL	CRS4, Italy
Lucie	/	/
Ludger Rinsche	LR]init[AG
Luis Alves de Sousa	LADS	ECDC
Luis Ibanez	LI	University of Southampton
Maco Comerio	MC	Cefriel
Makx Dekkers	MD	SEMIC Team
Marc Hanauer	MH	INSERM
Marco Combetto	MC	JRC
Marco Comerio	MC	Cefriel

Participants		
Name	Initials	Organisation
Mario Bravo	MB	NTT Data
Maryse Bücking	MB	NTM
Massimo Gaggero	MG	CRS4, Italy
Matthias Palmér	MP	Digg (Sweden) & MetaSolutions AB
Max De Wilde	MDW	SEMIC Team
Mélodie Bernaux	MB	DG SANTE
MHP	/	/
Michael Peolsson	MP	Swedish eHealth Agency
Michael van der Zel	MvdZ	HL7
Mihai Paunescu	MP	Publications Office, Luxembourg
Miriam Beusink	MB	Health-RI, Netherlands
Nuno Ferreira	NF	DG SANTE
Nuno Freire	NF	Europeana Foundation
Nuno Mesquita	NM	DG SANTE
Oystein Asnes	OA	Norwegian Digitisation Agency
Pascal Tanner	PT	Statistical Office Aargau

Participants		
Name	Initials	Organisation
Pavlina Fragkou	PV	DIGIT
Penny Labropoulou	PL	ILSP/RC "Athena"
Pero Ivanko	PI	CIPH
Peter Bruhn Andersen	BPA	Danish Agency for Digitisation
Peter Lubrich	PL	BASt, Germany
Pierre-Antoine Champin	PAC	W3C
Pieter van Everdingen	PVE	PLDN
Pim Volkert	PV	Health-RI, Netherlands
Pinja Perttunen	PP	Finnish Digital Agency
Rene Oostergo	PO	UMCG
Renée Thommes	RT	Ministry of Digitalisation, Lxembourg
Roberto	/	/
Ronald Kok	RK	City of Rotterdam
Sander Van Dooren	SVD	Digital Flanders
Sellitto Giovanni Paolo	SVP	ANAC, Italy
Severin Bäck	SB	MoH-AT

Participants		
Name	Initials	Organisation
Shrikant Ijmulwar	SI	Deloitte
Simou Stefania	SS	European Medicines Agency
Sorin Ostafiev	SO	ECDC
Sotiris Karampatakis	SK	Semantic Web Company
Stijn Tallir	ST	/
Tatjana Paveskovic	TP	CHIF, Croatia
Terje Sylvarnes	TS	Norwegian Digitisation Agency
Tessa	/	/
Tim Lieben	TL	OVAM
Trus Korsgaard	TK	Norwegian Directorate for eHealth
Vanessa Lima	VL	SPMS
Vanessa Mendes	VM	SPMS
Vincent	/	/
Vincent Sprengers	VS	Health-RI, Netherlands
Vittorio Meloni	VM	CRS4, Italy
Wai-Lam Wong	WLW	Nictiz

Participants		
Name	Initials	Organisation
William Verbeeck	WV	SEMIC Team
Zdenka Dudova	ZD	German Biobank Node, BBMRI.de
Zeljka Gluhak	ZG	/
Zia Alborzi	ZA	LNDS

Summary		
14:00 - 14:05	Welcome	Slides 1 - 7
14:05 - 14:10	The big picture: challenges within and across data spaces	Slides 8 - 12
14:10 - 14:20	Solution	Slides 13 - 19
14:20 - 15:15	Testimonials & real-life use cases	Slides 20 - 40
15:15 - 15:25	DCAT-AP ecosystem & SEMIC Service Offering	Slides 41 - 47
15:25 - 15:35	DSSC	Slides 48 - 59
15:35 - 15:55	Q&A & Closing	Slides 60 - 64

Full meeting minutes

Welcome	Goal of today: discuss interoperability of data spaces.
The big picture: Challenges within and across data spaces	Dataset sharing: interoperability across data spaces European strategy for data in which we create data spaces to foster more exchange, products and data services through more

	<p>accessible data. Today we focus on how we can create more accessible data that is FAIR.</p> <p>Challenges Questions exist around data exchange protocols, data solutions and data models to be used. These are the challenges data spaces are attempting to tackle today. However, if we create unique separate dataspace we don't solve any issues. Hence, we have to make dataspace shareable, reusable, and interoperable.</p> <p>Example: A road agency with road network data, an environmental agency with air quality data and a health agency with illness data each have their own solution. However, analysing the impact of road related air pollution on health (work absence) can only be answered by a combination of these datasets. Therefore these datasets need to be interoperable, understandable and findable.</p>
<p>Solution</p>	<p>DCAT-AP This is where DCAT-AP comes into play. DCAT-AP harmonises the metadata while allowing variations and facilitates description, etc. It can be used across different dataspace resulting in stable and long term management of the metadata and the datasets itself. DCAT-AP is based on DCAT, a W3C standard. It has mature implementations and a strong community.</p>
<p>Testimonials & real-life use cases</p>	<p>Three testimonials on extending DCAT-AP were given during this segment of the webinar. Namely one on mobility, one on health and one on environment.</p> <p>Mobility presented by Peter Lubrich (BAST) He presented the NAPCORE project to coordinate the National Access Points (NAPs) that have data on national mobility (EU, Norway, UK, ...). These NAPs usually use their own access points, but they face the same challenge, namely standardised metadata. The goal is to harmonise NAPs.</p> <p>The idea is to create a DCAT-AP extension for mobility, namely MobilityDCAT-AP which will be published soon, aligned with DCAT-AP, using suitable RDF vocabularies that relate to mobility. Data quality is very important hence the use of Data Quality vocabularies.</p>

Collaboration with SEMIC slowly grew over time and SEMIC played a supportive role in the development of MobilityDCAT-AP, such as expert advice, a review of the specification and ongoing alignment.

Makx Dekkers addresses the flexibility of DCAT-AP which is especially highlighted in the context of this testimonial.

Health presented by Licinio Kustra Mano (DG SANTE)

LKM describes a dream where health data from all around Europe would be available so researchers can access it, product builders can improve their products and policymakers can easily make data driven decisions. In other words, make health data available to those that can make the biggest impact. That is the goal of DG SANTE: to make health data discoverable.

The challenge is that health data categories are very diverse: electronic, administrative, claims and reimbursement, the list goes on. Furthermore, there is a balance between meaningful description against the related administrative burden. Lastly, we need to describe the quality and utility of the data.

The solution is not to reinvent the wheel but keep it turning. Therefore DCAT-AP is the base and the model can be extended without breaking it. For example by addressing the highly specialised requirements of the European Health Data Spaces (EHDS). Specificity for health data is a key necessity that addresses aforementioned challenges through specific descriptors, quality & utility, and balancing burden & obligation.

DG SANTE has closely collaborated with SEMIC in its development of HealthDCAT-AP through the HealthData@EU Pilot and of course the DCAT-AP foundation that the application profile is built upon.

Environment presented by Joeri Robrecht (DG ENV)

JR describes the core tasks of DG ENV which are development and implementation of the Green Deal Data Space, better use of Geospatial Intelligence data, coordination of DIGITAL Strategy & Space related files, and supporting initiatives to make the Commission and DG ENV more data driven.

The major challenges when setting up their data space were identifying the stakeholders, identifying the essential relevant data and their interoperability. An example is the development of a smart grid for EV charging poles where data from several datasets is necessary for the successful deployment. Here metadata also plays a big part in the interoperability issue.

The foundations are currently being laid for the GREAT project data space. Once the building blocks are created they will be run by the DSSC. Reusability is a key point. It is not the intention to reinvent the wheel.

They have looked at several ways to standardise their metadata and DCAT-AP is certainly a candidate. However, the minimum interoperability mechanism is necessary for their interface in their data services.

DG Environment has always had a good collaboration with SEMIC, for example through the creation of GeoDCAT-AP. JR remarks that DCAT-AP 3.0 is almost perfectly aligned with GeoDCAT-AP 2.0.

DG ENV collaborated on a pilot on open data spaces with the SOLID concept to explore the feasibility of sharing personal EV data within the boundaries of the European legal framework. This pilot included two use cases namely analysing the smart charging grid and the battery lifespan.

Makx mentions the active community and the necessity for interoperable data spaces.

Q&A testimonials:

Q: How do the speakers see the support DCAT-AP can give to data interoperability, rather than the metadata? How do you link DCAT-AP and the underlying data models?

JR: It is difficult to have a common model that meets all requirements, therefore extensions can be built and checked by validators. They have shown to be very difficult to maintain. JR hopes AI will help facilitate this process and relieve the concept of complex taxonomies. A direct link with an online data model, such as DCAT-AP, would be sufficient.

PL: Data providers would handpick and choose what works for them. Some models are close to DCAT-AP and could be integrated relatively easily. PR found that a lot of data providers would use keywords to describe mobility data, however, ontologies turned out to work better.

LKM: The starting point should be finding the right balance between meaningful descriptors that allows people to find the data. The data model for the metadata is one thing, which is defined through regulation. This is however not necessarily linked with the data model that is inside the data. Finally, we have the benefit of the internationally-used ontologies. These are some of the things we would like to capture in the properties of the metadata of health data. There is a requirement for ontologies that are open but are not applicable to all data. We may need to work with the idea of value sets/codelists.

BVN: With DCAT-AP we can make the existence of these data models visible and transparent. Based on that knowledge we can discuss if there are any expectations on how these data models are represented to increase interoperability. That is the connection where we go from the metadata perspective to the data perspective. Ideally, the community adopts common approaches where at least we know what kind of model we can expect without them being fully interoperable.

MD: Different data spaces are busy getting their own act together, but we would like a common way to approach unique issues within data spaces to end up with standardised results. The more data spaces work together through DCAT-AP the more we can serve everyone. There may also be application profiles that allow some level of interoperability. Everyone can work together in building their own application profile that meets their own requirements and then through SEMIC we can align all these data spaces. Later we can look at how we can facilitate services on data interoperability itself. Currently discoverability is the first priority which will lead into later ideas on wider interoperability.

BVN: This kind of phasing was conceptualised from a partner within the health data space, hence there is a necessity created from within the various data providers.

	<p>JZ: In France, the Frictionless Specification is used to describe the data. I haven't checked yet to see if they have linked the descriptions in DCAT-AP.</p> <p>JB: From my perspective, DCAT can be used for the self-descriptions in a data space, it is a vocabulary to define metadata of datasets and resources. Furthermore, the dataset described in a self-description (via DCAT) can refer via the conformsTo attribute to a standard/vocabulary the dataset is compliant to. These standards/vocabularies refer to ontologies, data models, and schema specifications, etc. and are published in a Vocabulary Hub. Yet remark that these standards/vocabularies are also just datasets, meaning that we can also describe metadata about a data space specific vocabulary with DCAT. TNO has its own implementation of a Vocabulary Hub, and currently we are researching how to achieve semantic interoperability within federated data space. Therefore, we currently work on federating Vocabulary Hubs by providing metadata of the vocabularies within a Vocabulary Hub via DCAT.</p>
<p>DCAT-AP ecosystem & SEMIC Service Offering</p>	<p>The benefits of DCAT-AP are findability of data sets and making catalogues interconnected and interoperable. SEMIC offers support in adopting this specification. An example of this support is the recent. DCAT-AP for HVD annex.</p> <p>While creating this annex an outward look was taken toward our peers. We look at the additional properties and whether they are useful outside the scope of the project at hand. If so, they are lifted to DCAT-AP at large to facilitate reusability. A key aspect is also to balance description and burden which relates to what LKM said.</p> <p>Different levels of the SEMIC service offering address different needs of data providers. The end result should always be the same, regardless of the level of SEMIC involvement, namely a data space specification that meets the data providers' needs.</p> <p>SEMIC aims to organise events to facilitate adoption and collaboration.</p>
<p>DSSC</p>	<p>The Data Spaces Support Centre (DSSC) was created to facilitate the sharing of data in a trustworthy and sovereign manner. A lack of common guidelines and practices might hamper these goals. The DSSC supports initiatives to create data spaces that are in line with the European Data Strategy. Exploration of needs and defining</p>

	<p>common requirements is a key aspect in this mission to help stakeholders make better decisions.</p> <p>The DSSC is based on an asset approach. The blueprint asset is a set of guidelines that help the different stakeholders to support data spaces, it contains a glossary, a conceptual model and (technical) building blocks that work together to achieve the full functionality of a data space. Semantic interoperability is essential and DCAT-AP helps facilitate this in the blueprint in the data models, data, services, offerings, catalogues, publication and discovery.</p> <p>The DSSC supports the community through support activities at different maturity levels and collaborates in a user-centric manner.</p>
<p>Q&A</p>	<p>Codelists:</p> <p>MP: Vocabularies or codelists should not be maintained manually and it would be valuable to consider them harvestable. A mechanism to consider is using natural repositories that could be harvested, unpackaged and provided.</p> <p>MD: There is a need to have a common codelist on vocabularies, but also the need for domain specific ones. It is beneficial to have both.</p> <p>JR: We need to take into account a formal and non-formal approach, for example in legal context they will be highly formalised to ensure correct use. For example, we need the same keyword across data spaces to identify HVD.</p> <p>MD: We have to think about governance. We can't have codelists disappear and lose meaning for certain domains. Common codelists need prolonged maintenance, because it creates codependency.</p> <p>FB: There are codelists together with OP that are captured and published. However, new input for relevant codelists is always welcome. Discoverability of codelists is another relevant topic.</p> <p>JR: The INSPIRE registry provides a core codelist set around Member States which can be extended, but is centred around a legal base and proper governance. Here SEMIC can provide additional value to bring this to the attention of adopters.</p>

GN: You should not only check using the same codelist, but there is the issue of data providers introducing their own keywords which leads to the necessity to constant updates and permanent ongoing maintenance.

MD: This is a governance issue. Someone needs to be responsible for what is going on in the data spaces and needs to be aware of what is going on in the data spaces.

TK: I see challenges on how to map data sources to DCAT-AP because it is unclear what exactly defines a health data source. Within the data source, the question is whether they are semantically interoperable and whether these codelists are the same. For example a cancer registry can have sub registries for types of cancers, but this cannot be implemented in DCAT-AP, because there is a lack of classes that belong to each other.

BVN states the topic raised by TK is indeed a challenge. DCAT-AP offers good support for the sharing of dataset descriptions, described in crude terms; but when going to estimate the concrete usability (or effort to become usable) then this becomes "domain specific" and "very detailed". Balancing this is a big challenge.

AA: New regulation is not implemented at the same pace as it is implemented in the market and there is a need for that.

BVN says you need to distinguish adoption in a technical and semantic sense. If you make a piece of software and hardcode a property, when you need a new property there comes complexity to implement it. SEMIC is always evolving but only endorse for longer periods. That means you can always extend your own version. It is semantically correct, however inclusion in DCAT-AP might take longer.

There may be a clash between an extension and what the community agrees on. On one hand DCAT-AP is very broad, on the other hand it can be very precisely used in a certain data space. For example in DCAT-AP 1, how you distribute your dataset is identified in a single class distribution. Debate within the community led to a difference between file distribution and service distribution.

MD: Incremental, frequent updates have been discussed a lot. However, for people building systems, this is not desirable. The structure that is adopted now is to update annually.

	<p>AA: Agrees with MD's comment. He would like more frequent releases of beta versions for testing to have practical advice to be implemented in the alpha version.</p> <p>MP: Developers can implement extensions, test their models against DCAT-AP validation and provide feedback. This is an alternative to what AA is asking.</p>
Closing	<p>DCAT-AP can help with interoperability of metadata within data spaces.</p>