

EHDS

HealthData@EU Pilot

Milestone M6.2

Technical working group on the transition from existing metadata templates to HealthDCAT-AP

« Working group minutes »

Pascal Derycke, Charles Andrew Vande Catsyne, Truls Korsgaard, Hans Aage Huru

31 January 2024



**Co-funded by
the European Union**

Disclaimer: Co-funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or HaDEA. Neither the European Union nor the granting authority can be held responsible for them.

Changes history

Date	Description of changes	Editor	Organisation
31/01/2024	First version with all content	Pascal Derycke	Sciensano

Document approval

Date	Description of changes	Editor	Organisation
30/03/2023	Approval of the milestone by the Steering Committee	Steering Committee	/

Introduction	3
Workshop on DCAT-AP provided by DIGIT Unit B2	5
TWG session held on the 14-07-2023 - Survey n°2	6
TWG session held on the 08-09-2023 - Survey n°2b	11
TWG session held on the 29-09-2023 - Survey n°3	17
TWG session held on the 13-10-2023 - Survey n°4 & 4b	48
TWG session held on the 27-10-2023 - Survey n°5	88
TWG session held on the 17-11-2023 - Survey n°6	98
TWG session held on the 24-11-2023 - Survey n°7	111
TWG session held on the 08-12-2023 - Survey n°8	121
TWG session held on the 22-12-2023 - Survey n°9 & mockups:	133

Introduction

'Milestone 6.2 of the EHDS2 pilot project, titled 'Technical Working Group (TWG) on the Transition from Existing Metadata Templates to HealthDCAT-AP', falls under the Work Package dedicated to designing the HealthDCAT Application Profile. This milestone primarily comprises the 'Working Group Minutes'.

The TWG, consisting of approximately 80 participants including representatives from the EHDS2 pilot project consortium and external stakeholders in the health sector, was established to offer support. From June to December 2023, 12 TWG sessions were organized. These sessions focused on discussing and deciding the properties of the application profile that are essential for better describing health data and enabling interoperability of metadata catalogues within the healthData@EU infrastructure.

The Technical Working Group (TWG) initiated its activities with a foundational workshop comprising two sessions. These sessions were expertly led by the Unit B2 Interoperability and Digital Government of EC DIGIT. Subsequently, the Federal Belgian Institute of Public Health, Sciensano, took the helm for an additional series of 10 sessions. Each of these sessions was centered around a specific EU Survey form, adding structure and focus to the discussions. To foster extensive consultations and contributions, the EU Survey forms were kept open for a duration of 10 days. After this period, the gathered responses were meticulously reviewed. The insights gleaned from this review were then presented in the subsequent TWG session.

To ensure accessibility and continuity, all sessions were recorded and have been made available to members of the project consortium. Accompanying this document are the anonymised results of these surveys. For ease of reference and analysis, the results have been compiled into tables corresponding to each survey and have been included in this document.

The documentation of the 'Working Group Minutes' is meticulously organised and presented in chronological order, as follows:

Workshop: Webinar DCAT-AP Health Session 1 & 2

Session 1 16-06-2023:

Session 2 30-07-2023:

TWG session held on the 14-07-2023 - Survey n°2*

- Implementation of dct:identifier as a Mandatory property using PURI (Persistent Uniform Resource Identifier)
- Introduction and utilisation of dcat:healthCategory

TWG session held on the 08-09-2023 - Survey n°2b

- dct:identifier mandatory property as PURI
- dcat:healthCategory – Art. 33 of the Regulation proposal
- NSIP requirements – review of mandatory properties

TWG session held on the 29-09-2023 - Survey n°3

- Create rich metadata record – User perspective
- Use case: faceted search

TWG session held on the 13-10-2023 - Survey n°4 & survey n°4b

“What, why, how, ...”

- Create rich metadata record – Data holder perspective
- Analyses of the HDRUK health Gateway and HealthInformationPortal.eu
- Use case: create metadata without administrative burden

- dct:alternative
- schema:typicalAgeRange
- schema:codingSyst
- schema:inCodeSet

TWG session held on the 27-10-2023 - Survey n°5

“provide a subset – proxy data for sensitive health datasets”

- adms:sample

TWG session held on the 17-11-2023 - Survey n°6

“providing dataset metrics”

- dct:mediator
- dct:valid
- ex:datametrics

TWG session held on the 24-11-2023 - Survey n°7

“metadata fit for AI”

- SeTA presentation by the JRC-Ispra
- dct:provenance, ex:populationtype, dpv :purpose

TWG session held on the 08-12-2023 - Survey n°8

“Data classification”

- dpv:dataController, ...
- Resource type, data type, funding type, ...

TWG session held on the 22-12-2023 - Survey n°9 & survey n°9b

- Draft version of the healthDCAT-AP (LINK-VACC dataset example)
- Mockups of the healthdata.europa.eu catalogue

*After the session on 14-07-2023, the decision was made to pause the TWG activities during the summer and resume in September, aiming to maximize participation in future sessions.

Workshop on DCAT-AP provided by DIGIT Unit B2

Topics covered:

Session n°1 16-06-2023:

- Approach & governance
- Short tour
- Creating a profile
- Use cases
- Use case 1: harvesting

Session n°2 30-07-2023:

- Use case 1: Example case BioBank (bbmri-eric.eu)
- Use case 2: Data findability
- Use case 3: access to data for performing data analysis.

Pavlina Fragkou from the Informatics Directorate-General (DIGIT), Unit B2 Interoperability and Digital Government of the European Commission, led a workshop on metadata creation and management for health data. Assisted by semantic architects Bert Van Nuffelen and Makx Dekkers, the workshop offered an in-depth overview of the processes involved. The focus was on DCAT (<https://github.com/SEMICEu/DCAT-AP/>), semantic web technologies, persistent identifiers, and the collaborative development of data standards through SEMIC.

In its second session, the workshop delved into the challenges and solutions associated with managing and categorising health datasets and metadata. It highlighted practical examples and emphasised the importance of community-driven decision-making in this domain.

Minutes of the TWG workshop (session 1)

1. Introduction and Objectives: Bert Van Nuffelen introduces the workshop, emphasizing the goal of creating a health profile for metadata standards and the importance of consensus and common understanding.
2. Overview of Metadata and Data Standards: The workshop begins with a high-level overview of metadata, its role in cataloging datasets, and the importance of creating a unified approach for data standards.
3. Discussion on Metadata Properties and Catalogs: Participants discuss various aspects of metadata, including the properties and classes used in cataloging datasets. The importance of persistent identifiers and context in metadata is highlighted.
4. Interactive Q&A and Feedback: The workshop is interactive, with participants like Andrzej Strug from GUMed BBMRI raising questions and providing insights into the application of metadata in health data management.
5. Harvesting Metadata and Data Portals: The concept of harvesting metadata from various data portals is explored. The discussion includes the complexities of aggregating metadata across different data sources and the technical challenges involved.
6. Use Cases and Practical Applications: The workshop focuses on practical use cases, demonstrating how metadata standards can be applied in health data management. The goal is to align the metadata structure with the needs of the health data community.
7. Closing Remarks and Future Directions: The workshop concludes with plans for future collaboration and the importance of community input in shaping metadata standards for health data.

Minutes of the TWG workshop (session 2)

1. Session Introduction and Agenda: Bert Van Nuffelen opens the session, outlining the agenda which includes a deeper dive into use cases of data harvesting based on previous sessions.
2. Concrete Examples and Research Findings: The workshop involves discussing research and examples related to Biobank data. Van Nuffelen emphasizes making the session more concrete with practical examples for each harvesting case.
3. Discussion on Data Sets and Catalogues: The complexities of data sets and catalogs in the context of health data are explored. The session includes discussions on how different biobanks present their data and the challenges in categorizing them as either datasets or catalogs.
4. Interactive Q&A and Community Input: Throughout the workshop, there is an emphasis on interaction, with participants raising questions and contributing to the discussions. The goal is to find common consensus among the working group members on various issues.
5. Challenges in metadata and data access: The session discusses the challenges associated with accessing data sets and metadata, and how different organizations manage and present this information.
6. Insights into Data Modeling and Cataloging: Insights are provided into the semantic modeling of data, including the representation of biobanks and the cataloging of data sets. The distinction between organizations and catalogs in the context of data representation is a key focus.
7. Closing Remarks and Future Directions: The session concludes with a discussion on the need for further research and the importance of community input in shaping the approach to data management and harvesting in health data systems.

TWG session held on the 14-07-2023 - Survey n°2

Topics covered:

- Implementation of `dct:identifier` as a Mandatory property using PURI (Persistent Uniform Resource Identifier)
- Introduction and utilisation of `dcat:healthCategory`

The TWG survey no. 2 centers on the use of Persistent Dereferenceable Identifiers (PURI) in `dct:identifier` to uniquely identify and link datasets across different portals. The focus is on ensuring unique dataset identification and improving data findability. It also explores creating and maintaining a codelist for the `dct:source` property to specify the data source, enhancing the findability of health data. The survey solicits opinions on making `dct:identifier` and the use of PURI mandatory in DCAT-Health, and invites suggestions for the "Data source" code list and its categorisation (mandatory, recommended, or optional).

Minutes:

1. Introduction and Objectives: Pascal Derycke introduces the session, outlining its objectives. The focus is on designing an extension for health data management and implementing a proof of concept for a pilot project.
2. Methodology and Tools: Derycke recaps the methodology used in the technical working group, introduces new tools implemented for metadata management, and emphasizes the importance of collecting feedback through surveys.

3. Discussion on Metadata Properties: The session discusses specific metadata properties, including identifiers and data source use cases. The goal is to decide which properties should be mandatory, recommended, or optional.
4. Interactive Feedback and Surveys: Participants are encouraged to provide feedback on metadata properties using surveys. The session includes a demonstration of how to use the survey forms and the importance of basing decisions on facts and research.
5. Review of Metadata Records: The session reviews existing metadata records, highlighting the completeness and accuracy of the information provided, and the importance of persistent identifiers.
6. Technical Challenges and Solutions: The session explores technical challenges in metadata management, such as ensuring persistent identifiers and understanding the granularity of data sets.
7. Closing remarks and next Steps: The session concludes with plans for future discussions and the importance of collaborative decision-making in the ongoing development of health data management systems.

Results of the survey:

TGW survey n°2	yes	no	Total number of responses	Comments
dct:identifier as mandatory property?	5	0	5	
Make the use of PURI mandatory?	5	0	5	
Use of dcat:source to establish a relation to resource types		no		Attention point: dct:source provides a relation to another dataset (not a resource) EMA gives a list of organisations acting as health data providers

Survey n°2 is available at [this link](#)

Health DCAT extension - TWG n°2

Fields marked with * are mandatory.

DCAT Health extension

Technical Working Group session 14-07-2023

* Your email:

pascal.derycke@sciensano.be

Nota bene: This form remains open until Monday 24th of July

Dcat:healthMetadataIdentifier

Use case: One of the most important data processing steps of data catalogues is harvesting. The result is that the same dataset description is present on multiple data portals. One key information element is the identifier of a dataset. Aggregating data portals have difficulties to detect duplicates.

Solution: use Persistent Dereferenceable Identifiers (PURI) in dct:identifier

Ex: <https://national-catalogue.org/dataset/68112f77-8f2c-496a-8398-77e52b60c883> Unique, with context and dereferenceable

Ref: [DCAT-AP](#) [dct:identifier rdfs:Literal 0..n]

[Data.europa.eu SPARQL query](#)

[See metadata records in TWG sandbox](#)

Resources:

- The FAIR principles explained ([Video lecture by Luiz Bonino at 7:30](#))
- Workshop on DCAT by DIGIT2 (ref: 20230630_DCAT-AP for Health_2.pptx) Slides 14 to 26
- WP5 POC (harvesting architecture based on eDelivery)

Do you agree making dct:identifier as a mandatory property in DCAT-Health?

- yes
 no

Do you agree making the use of Persistent Dereferenceable Identifiers as a mandatory value in DCAT-Health?

- yes
- no

Provide any comments or suggestions:

Dcat:healthSource

Use case: Findability (Faceted search). Indicate the source of the health data, whether it is from hospitals, research institutions, public health agencies, or other relevant sources.

Solution: create and maintain a codelist. Use the DCAT-AP property dct:source to specify the entity responsible for providing the dataset

Ref: [DCAT-AP](#) [dct:source dcat:Dataset 0..n]

[Data.europa SPARQL query](#)

[See metadata records in TWG sandbox](#)

Resources:

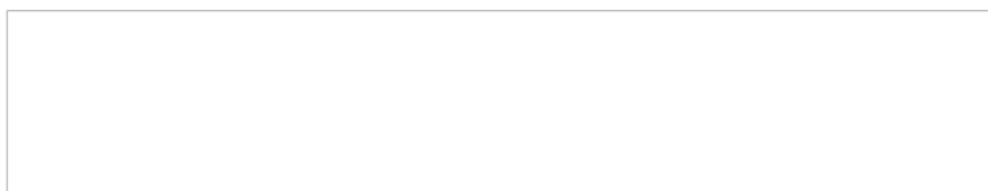
- Workshop on DCAT by DIGIT2 (ref: 20230630_DCAT-AP for Health_2.pptx) Slides 51 to 55
- Milestone M6.1: Report on the landscape analysis of available metadata catalogues.
- <https://op.europa.eu/en/web/eu-vocabularies/authority-tables> (ex: [corporate-body-classification](#))

* Provide your suggestions for the "Data source" code list:

Would you make the use of the "Data source" code list

- Mandatory
- Recommended
- or Optional?

Provide any comments or suggestions:



Contact

[Contact Form](#)

TWG session held on the 08-09-2023 - Survey n°2b

Topics covered:

- dct:identifier mandatory property as PURI
- dcat:healthCategory – Art. 33 of the Regulation proposal
- NSIP requirements – review of mandatory properties

The survey from the Technical Working Group (TWG) number 2b, updated on September 8, re-includes, as discussed key topic, the use of Persistent Dereferenceable Identifiers (PURI) as a mandatory value in dct:identifier for dataset uniqueness. Another key topic includes the creation of a codelist for EHDS2 datasets using the dct:subject property to categorise health data topics. Additionally, it discusses aligning DCAT-AP-Health with National Single Information Points (NSIPs) requirements concerning the reuse of public sector data. Participants are asked to provide feedback on these aspects, including the mandatory use of the EHDS2 data classification and adopting NSIP requirements in DCAT-Health.

Minutes:

1. Introduction and Session Overview: Pascal Derycke opens the session, reviewing the agenda, which includes discussions on dct:identifier which was already discussed at the previous session and the implications of data governance acts on health data.
2. Survey review and new survey: Pascal presents the results of previous survey and introduces the new one. These surveys aim to gather opinions on creating rich metadata records and understanding new property relevance for health data.
3. Discussion on metadata and Data Governance: The session delves into the technical aspects of metadata properties, emphasizing the need for comprehensive and accurate metadata records. There's a particular focus on the impact of data governance acts on metadata management.
4. Interactive participation and feedback: The session is interactive, with participants like Filip Pattyn providing insights and raising questions about metadata and data governance. Derycke encourages feedback and discussion throughout the session.
5. Technical Insights and Application Profiles: The session provides technical insights into various metadata properties, including subject categorisation and data quality. The importance of aligning with existing standards and application profiles is discussed.
6. Collaboration and decision-making: Participants are encouraged to collaborate and provide feedback on proposals. The session aims for a collective decision-making process on best practices and standards for health data metadata.
7. Closing remarks and future steps: The session concludes with plans for further analysis of feedback and the implementation of metadata records. Derycke emphasises the ongoing nature of the work and the importance of community input in shaping health data management strategies.

Results of the survey:

TGW survey n°2b	yes	no	Total number of responses	Comments
dct :identifier as mandatory property?	7	0	7	
Make the use of PURI mandatory?	6	1	7	
Art. 33	5	2	7	Comments are focusing on the

		Rationales _____ for answering "no" not correct		accuracy of the art. 33 list
NSIP requirements	5	2 Rationales _____ for answering "no" not correct	7	

Survey n°2b is available at [this link](#)

Health DCAT extension - TWG n°2 (updated) 08-09-2023

Fields marked with * are mandatory.

DCAT Health extension

Technical Working Group session 08-09-2023 (updated version of 17-07-2023)

* Your email:

pascal.derycke@sciensano.be

Nota bene: This form remains open until Monday September 18

Dcat:healthMetadataIdentifier

Use case: One of the most important data processing steps of data catalogues is harvesting. The result is that the same dataset description is present on multiple data portals. One key information element is the identifier of a dataset. Aggregating data portals have difficulties to detect duplicates.

Solution: use Persistent Dereferenceable Identifiers (PURI) in dct:identifier

Ex: <https://national-catalogue.org/dataset/68112f77-8f2c-496a-8398-77e52b60c883> Unique, with context and dereferenceable

Ref: [DCAT-AP](#) [dct:identifier rdfs:Literal 0..n]

[Data.europa.eu SPARQL query](#)

[See metadata records in TWG sandbox](#)

Resources:

- The FAIR principles explained ([Video lecture by Luiz Bonino at 7:30](#))
- Workshop on DCAT by DIGIT2 (ref: 20230630_DCAT-AP for Health_2.pptx) Slides 14 to 26
- WP5 POC (harvesting architecture based on eDelivery)

Do you agree making dct:identifier as a mandatory property in DCAT-Health?

- yes
 no

Do you agree making the use of Persistent Dereferenceable Identifiers as a mandatory value in DCAT-Health?

- yes
 no

Provide any comments or suggestions:

Dcat:healthSource

Use case: Findability (Faceted search). Indicate the topic category of the health data, whether it is from hospitals, research institutions, public health agencies, or other relevant sources.

Solution: create and maintain a codelist of EHDS2 datasets. Use the GeoDCAT-AP/Dublin Core property 'dct:subject' to specify the topic category of the dataset

Ref: [GeoDCAT-AP](#) [dct:subject dcat:Dataset 0..n]

[Data.europa SPARQL query](#)

[Article 33](#)

Minimum categories of electronic data for secondary use

Data holders shall make the following categories of electronic data available for secondary use in accordance with the provisions of this Chapter:

- (a) EHRs;*
- (b) data impacting on health, including social, environmental behavioural determinants of health;*
- (c) relevant pathogen genomic data, impacting on human health;*
- (d) health-related administrative data, including claims and reimbursement data;*
- (e) human genetic, genomic and proteomic data;*
- (f) person generated electronic health data, including medical devices, wellness applications or other digital health applications;*
- (g) identification data related to health professionals involved in the treatment of a natural person;*
- (h) population wide health data registries (public health registries);*
- (i) electronic health data from medical registries for specific diseases;*
- (j) electronic health data from clinical trials;*
- (k) electronic health data from medical devices and from registries for medicinal products and medical devices;*
- (l) research cohorts, questionnaires and surveys related to health;*
- (m) electronic health data from biobanks and dedicated databases;*
- (n) electronic data related to insurance status, professional status, education, lifestyle, wellness and behaviour data relevant to health;*
- (o) electronic health data containing various improvements such as correction, annotation, enrichment received by the data holder following a processing based on a data permit.*

Resources:

- Workshop on DCAT by DIGIT2 (ref: 20230630_DCAT-AP for Health_2.pptx) Slides 51 to 55
- Milestone M6.1: Report on the landscape analysis of available metadata catalogues.
- <https://semiceu.github.io/GeoDCAT-AP/releases/2.0.0/#dataset-topic-category> (Example of INSPIRE categories)
- Example of an INSPIRE geospatial metadata catalogue:
https://metadata.helcom.fi/geonetwork/srv/eng/catalog_search#/home

- Provide your comments about the use of "dct:subject":

Would you make the use of the EHDS2 data classification

- Mandatory
- Recommended
- or Optional?

Provide any comments or suggestions:

Align DCAT-AP-Health with National Single Information Points (NSIPs) requirements

"The Data Governance Act includes mechanisms to foster the reuse of public sector data that, for certain reasons, cannot be made available as open data. This could be health or mobility data that are, according to Article 3 of the Data Governance Act, protected on grounds of statistical or commercial confidentiality, intellectual property rights of third parties, or privacy." Source: [European Single Access Point on data.europa.eu](https://european-council.europa.eu/media/e0604606-1230-4b43-8059-116185402000/asset/document/20230606_01_DGA-NSIPs-Requirements)

The following metadata is mandatory for NSIP datasets:

Property	URI	Range	Usage note	Cardinality
Title (M)	dct:title *	rdfs:Literal	This property contains a name given to the Dataset. This property can be repeated for parallel language versions of the name.	1..n
Description (M)	dct:description *	rdfs:Literal	This property contains a free-text account of the Dataset. This property can be repeated for parallel language versions of the description.	1..n
Publisher (M)	dct:publisher	foaf:Agent	This property refers to an entity (organisation) responsible for making the Dataset available.	1..1
Access rights (M)	dct:accessRights	dct:RightsStatement	This property refers to information that indicates whether the dataset is open data, has access restrictions, or is not public. From the controlled vocabulary of the Publications Office of the EU ⁶ , the following codes should be used for NSIP data: "non-public" or "restricted". "open" is prohibited for NSIP data.	1..1
Distribution (M)	dct:distribution	dcat:Distribution	Distribution(s) available for a dataset.	1..n

The following metadata is mandatory for distributions:

Property	URI	Range	Usage note	Cardinality
Format (M)	dct:format	dct:MediaTypeOrExtent	This property refers to the file format of the Distribution. You can only specify one format per Distribution. If an NSIP offers the same data in different formats, each format must be specified as a separate distribution.	1..1
Size (M)	dcat:byteSize	rdfs:Literal	The size in bytes can be approximated (as a decimal) if the precise size is not known. ⁸	1..1
Access procedure (M)	dcat:accessURL *	rdfs:Resource	A URL of a Website that enables either access to the described data or that contains information on how to request the data.	1..n
Conditions for re-use (Rights) (M)	dct:rights	dct:RightsStatement	This property refers to a statement that specifies rights associated with the Distribution.	1..1

Do you agree adopting NSIP requirements in DCAT-Health?

- yes
 no

Provide any comments or suggestions:

Contact

[Contact Form](#)

TWG session held on the 29-09-2023 - Survey n°3

Topics covered:

- Create rich metadata record – User perspective
- Use case: faceted search

The TWG survey no. 3 focuses on reviewing DCAT-AP properties of the dataset class, applying FAIR data principles. The session provides a comprehensive overview of the complexities involved in creating and managing metadata records for health data, emphasizing the need for rich, detailed, and accurate metadata to enhance data discovery and usability.

The survey asks participants to consider the importance of various dataset properties as filters in a metadata catalogue. It also invites feedback on optional properties and their usage. The survey is comprehensive, seeking to refine and enhance the metadata standards for health datasets.

Minutes:

1. Introduction and Agenda Overview: Pascal Derycke opens the session, outlining the agenda, which includes a review of previous survey results and a new survey on creating rich metadata records.
2. Discussion on Metadata Records: The session delves into a detailed analysis of metadata properties, including their significance and the rationale behind them. Participants are asked for feedback on each property, emphasizing the importance of creating comprehensive metadata records.
3. Survey Analysis and Feedback: Derycke presents survey results and encourages feedback, highlighting the significance of properties like the city identifier and discussing their impact on health data.
4. Technical Insights: The session provides technical insights into various metadata properties, including title, description, keywords, and publisher. The importance of tagging and language specification is also discussed.
5. Interactive Q&A and Participation Encouragement: Throughout the session, there is an emphasis on interactive participation, with Derycke encouraging responses in the chat and addressing questions from participants.
6. Discussion on Metadata Properties: The session covers a range of metadata properties, discussing their importance from both user and data holder perspectives. This includes properties like geographical coverage, temporal coverage, and access rights.
7. Closing Remarks and Future Steps: The session concludes with plans for further analysis of feedback and the implementation of metadata records, highlighting the ongoing nature of the work and the importance of community input.

Results of the survey:

TGW survey n°3	Free text	Code list	URL	Mandatory	Recommended	Optional	total	Comments
dcat:keyword				1	6	-	7	"garbage bag"

This property contains a keyword or tag describing the Dataset.								Some participants did not understand it is a free list of keywords?
dcat:contactPoint This property contains contact information that can be used for sending comments about the Dataset.			contact	5	2	-	7	HDAB should be the official contact point
dct:spatial This property refers to a geographic region that is covered by the Dataset.				3	4	-	7	
dct:publisher This property refers to an entity (organisation) responsible for making the Dataset available.			contact	6	1	-	7	
dct:temporal This property refers to a temporal period that the Dataset covers.				4	3	-	7	
dcat:theme This property refers to a category of the Dataset. A Dataset may be associated with multiple themes.				6	1	-	7	Important for filtering
dct:accessRights This property refers to information that indicates whether the Dataset is open data, has access restrictions or is not public.				4	2	1	7	Important to inform that the dataset is not open access – which label to choose? No-public, sensitive, ...
dct:conformsTo This property refers to an implementing rule or other specification.				0	2	5	7	Not all datasets conform to a certain standard
dct:creator This property refers to the entity responsible for producing the dataset.			contact	2	2	3	7	
foaf:page This property refers to a page or document about this Dataset			resource	0	3	4	7	What should the doc contain?
dct:accrualPeriodicity This property refers to the frequency at which the Dataset is updated.				1	3	3		Should use a code list
dct:hasVersion This property refers to a related Dataset that is a version, edition, or adaptation of the described Dataset.			Dataset	0	0	7		
dct:isReferencedBy This property is about a related resource, such as a publication, that references, cites, or otherwise points to the dataset.			Resource	0	1	6		Not maintainable
dct:isVersionOf This property refers to a related Dataset of which the described Dataset is a			Dataset	0	1	6		

version, edition, or adaptation.								
dcat:landingPage This property refers to a web page that provides access to the Dataset, its Distributions and/or additional information. It is intended to point to a landing page at the original data provider, not to a page on a site of a third party, such as an aggregator.			Resource	2	2	3		According to definition the landing page provides access to the dataset
dct:language This property refers to a language of the Dataset. This property can be repeated if there are multiple languages in the Dataset.				2	3	1	-1	
dct:modified This property contains the most recent date on which the Dataset was changed or modified.				1	3	3		What's about datasets modified daily?
adms:identifier This property refers to a secondary identifier of the Dataset, such as MAST/ADS , DataCite , DOI , EZID or W3ID .				0	1	6		
dct:provenance This property contains a statement about the lineage of a Dataset.				0	2	4	-1	Need for a template to structure info
prov:qualifiedAttribution This property refers to a link to an Agent having some form of responsibility for the resource.			Contact	0	1	6		
dcat:qualifiedRelation This property provides a link to a description of a relationship with another resource.			Resource	0	0	7		
dct:relation This property refers to a related resource.			Resource	0	0	6	-1	Useful links should not be more than optional?
dct:issued This property contains the date of formal issuance (e.g., publication) of the Dataset.				1	3	3		What does it mean for sensitive datasets?
adms:sample This property refers to a sample distribution of the dataset.			Dataset					
dct:source This property refers to a related Dataset from which the described Dataset is derived.			Dataset	0	1	6		<i>All relations to external resources should be made R to help understanding the dataset</i>
dcat:spatialResolutionInMeters This property refers to the minimum spatial separation resolvable in a dataset, measured in meters.				0	0	6	-1	Not always relevant
dcat:temporalResolution				1	0	6		Not always relevant

This property refers to the minimum time period resolvable in the dataset.								
dct:type This property refers to the type of the Dataset. A recommended controlled vocabulary data-type is foreseen.			4	2	1			
owl:versionInfo This property contains a version number or other version designation of the Dataset.			0	0	7			
adms:versionNotes This property contains a description of the differences between this version and a previous version of the Dataset. This property can be repeated for parallel language versions of the version notes.			0	2	5			
prov:wasGeneratedBy This property refers to an activity that generated, or provides the business context for, the creation of the dataset.			1	2	4			

Survey n°3 is available at [this link](#)

Health DCAT extension - TWG n°3 29-09-2023

Fields marked with * are mandatory.

DCAT Health extension

Technical Working Group session 29-09-2023

* Your email:

pascal.derycke@sciensano.be

Nota bene: This form remains open until Monday October 9

Review of the DCAT-AP properties of dataset class presented in the TWG sandbox.

Requirement derived from applying FAIR data principles:

Findable

F2. Data are described with rich metadata (defined by R1 below)	Define a short list of important metadata properties as mandatory properties to achieve a satisfying level of discoverability/reusability.
---	--

Reusable

R1. (Meta)data are richly described with a plurality of accurate and relevant attributes	See F1
--	--------

Use Case:

- [DCAT-AP usage analysis: use of dct:publisher and dcat:contactPoint](#)
- Bert Van Nuffelen: "The quality and completeness of the metadata in data.europa.eu is a EU wide challenge"

Among the following properties of the dataset class which of the properties would you consider as an important filter in a metadata catalogue?

- | | | |
|---|--|---|
| <input type="checkbox"/> geographical coverage
dct:spatial | <input type="checkbox"/> access rights
dct:accessRights | <input type="checkbox"/> release date
dct:issued |
| <input type="checkbox"/> keyword
dcat:keyword | <input type="checkbox"/> conforms to
dct:conformsTo | <input type="checkbox"/> spatial resolution
dcat:spatialResolutionInMeters |
| <input type="checkbox"/> publisher
dct:publisher | <input type="checkbox"/> frequency
dct:accrualPeriodicity | <input type="checkbox"/> temporal resolution
dcat:temporalResolution |
| <input type="checkbox"/> temporal coverage
dct:temporal | <input type="checkbox"/> language
dct:language | <input type="checkbox"/> data type
dct:type |
| <input type="checkbox"/> theme
dcat:theme | <input type="checkbox"/> modification date
dct:modified | <input type="checkbox"/> was generated by
prov:wasGeneratedBy |

Provide any comments, suggestions or questions:

DCAT-AP has 2 mandatory properties: "Title" and "Description".

TITLE

This property contains a name given to the Dataset. This property can be repeated for parallel language versions of the name.

Example:

	[dct:title rdfs:Literal 1..n]
Mandatory	dct:title "Austrian Corona Panel Project"@en dct:title "Project panel Corona autrichien"@fr .

DESCRIPTION

This property contains a free-text account of the Dataset. This property can be repeated for parallel language versions of the description.

Example:

	[dct:description rdfs:Literal 1..n]
Ma n d a t o r y	dct:description "The City of Hamburg (Environmental and Energy Authority) accompanies, supports and promotes energetic neighbourhood concepts that illuminate measures to increase..."@en .

Resources:

- [TWG Sandbox](#)
- [Data.europa.eu SPARQL query](#)

Provide any comments, suggestions or questions:

KEYWORD

This property contains a keyword or tag describing the Dataset.

Example:

	[dcat:keyword rdfs:Literal 0..n]
Recommen ded	dcat:keyword "COVID-19"@en, "SARS-CoV-2"@en, "Vaccination"@en, "Test"@en .

Resources:

- [TWG Sandbox](#)
- [Data.europa.eu SPARQL query](#)
- [Example of an open Health geospatial Dataset with many keywords](#)

Would you make the use of **dcat:keyword** as

- mandatory
- or keep it as is (recommended)?

Provide any comments, suggestions or questions:

CONTACT POINT

This property contains contact information that can be used for sending comments about the Dataset.

Example:

	[dcat:contactPoint vcard:Kind 0..n]
Recommended	<pre>dcat:contactPoint [a vcard:Organization; vcard:fn "Behörde für Umwelt, Klima, Energie und Agrarwirtschaft (BUKEA) "; vcard:hasAddress [a vcard:Address; vcard:country-name "DEU"; vcard:locality "Hamburg"; vcard:postal-code "D-21109"; vcard:street-address "Neuenfelder Strasse 19"]; vcard:hasEmail <mailto:waermekataster@bukea.hamburg.de>];</pre>

Resources:

- [TWG Sandbox](#)
- [Data.europa.eu SPARQL query](#)

Would you make the use of **dcat:contactPoint** as

- mandatory
- or keep it as is (recommended)?

Provide any comments, suggestions or questions:



GEOGRAPHICAL COVERAGE

This property refers to a geographic region that is covered by the Dataset.

Example:

	[dct:spatial dct:Location 0..n]
R e c o m m e n d ed	<pre>dct:spatial <https://sws.geonames.org/3337388/>; dct:spatial [a dct:Location; locn:geometry "{\"type\": \"Polygon\", \" coordinates\": [[[10.326304, 53.394985], [10.326304, 53.964153], [8.420551, 53.964153], [8.420551, 53.394985], [10.326304, 53.394985]]]}\"^^<https://replacement.io/assignments/media-types /application/vnd.geo+json> , \"POLYGON ((10.3263 53.3950, 10.3263 53.9642, 8.4206 53.9642, 8.4206 53.3950, 10.3263 53.3950))\" ^^<http://replacement.io/ont/geosparql#wktLiteral>]; dct:spatial [a dct:Location; locn:adminUnitL2 <http://dcat-ap.de/def /politicalGeocoding/regionalKey/020000000000>];</pre>

Resources:

- [TWG Sandbox](#)
- [Data.europa.eu SPARQL query](#)

Machine actionable metadata:

```

1 |URL="https://sws.geonames.org/3337388"
2
3 import rdflib
4 from rdflib import Graph
5
6 # Create a Graph, parse in Internet data
7 cat = Graph().parse(URL)
8
9 print(cat.serialize())

```

```

@prefix cc: <http://creativecommons.org/ns#> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix gn: <http://www.geonames.org/ontology#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix wgs84_pos: <http://www.w3.org/2003/01/geo/wgs84_pos#> .

```

```

<https://sws.geonames.org/3337388/> a gn:Feature ;
  gn:alternateName "Flandre",
    "Vlaanderen",
    "Flandes"@ca,
    "Flämische Region"@de,
    "Flemish Region"@en,
    "Flandra Regiono"@eo,
    "Flandes"@es,
    "Région Flamande"@fr,
    "Flóndras"@ga,
    "Flandres"@pt ;
  gn:childrenFeatures <https://sws.geonames.org/3337388/contains.rdf> ;
  gn:countryCode "BE" ;
  gn:featureClass <https://www.geonames.org/ontology#A> ;
  gn:featureCode <https://www.geonames.org/ontology#A.ADM1> ;
  gn:locationMap <https://www.geonames.org/3337388/flanders.html> ;
  gn:name "Flanders" ;

```

Would you make the use of **dct:spatial** as

- mandatory
- or keep it as is (recommended)?

Provide any comments, suggestions or questions:

PUBLISHER

This property refers to an entity (organisation) responsible for making the Dataset available.

Example:

	[dct:publisher foaf:Agent 0..1]
R e c o m m e n d e d	<pre>dct:publisher [a foaf:Organization; locn:address [a locn:Address; locn:adminUnitL1 "DEU"; locn:postCode "D-21109"; locn:postName "Hamburg"; locn:thoroughfare "Neuenfelder Strasse 19"]; foaf:mbox <mailto:anselm.sprandel@bukea.hamburg.de>; foaf:name "Behörde für Umwelt, Klima, Energie und Agrarwirtschaft (BUKEA)"; foaf:phone <tel:+% 2B49%2040%204%2028%2040%20-%202085>]; dct:publisher <https://opendata.schleswig-holstein.de/organization /94498aaa-9bd0-4ac0-ae70-db319a3ca1d8>;</pre>

Resources:

- [TWG Sandbox](#)
- [Data.europa.eu SPARQL query](#)
- [Requesting Linked Open Data](#)
- Ex: [Opendata.schleswig-holstein.de \(Metadata\)](#)

Would you make the use of **dct:publisher** as

- mandatory
 or keep it as is (recommended)?

Provide any comments, suggestions or questions:

TEMPORAL COVERAGE

This property refers to a temporal period that the Dataset covers.

Example:

	[dct:temporal dct:PeriodOfTime 0..n]
--	--------------------------------------

Recommended	<pre>dct:temporal [a dct:PeriodOfTime; dcat:endDate "2019-11-26"^^<http://www.w3.org/2001 /XMLSchema#date>; dcat:startDate "2013-01-01"^^<http://www.w3.org/2001 /XMLSchema#date>];</pre>
-------------	---

Resources:

- [TWG Sandbox](#)
- [Data.europa.eu SPARQL query](#)

Would you make the use of **dct:temporal** as

- mandatory
 or keep it as is (recommended)?

Provide any comments, suggestions or questions:

THEME

This property refers to a category of the Dataset. A Dataset may be associated with multiple themes.

Example:

	[dcat:theme, subproperty of dct:subject skos:Concept 0..n]
Recommended	<pre>dcat:theme <http://publications.europa.eu/resource/authority/data-theme/ENVI> , <http://dcat-ap.ch/vocabulary/themes/territory> .</pre>

Resources:

- [TWG Sandbox](#)
- [Data.europa.eu SPARQL query](#)
- [Wikidata: A large-scale collaborative ontological medical database](#)

Ontology search tools:

- <https://ontobee.org/>
- <https://www.wikidata.org/>
- [Ontologie search](#)

Would you make the use of **dcat:theme** as

- mandatory
- or keep it as is (recommended)?

Provide any comments, suggestions or questions:

ACCESS RIGHTS

This property refers to information that indicates whether the Dataset is open data, has access restrictions or is not public.

Example:

	[dct:accessRights dct:RightsStatement 0..1]
Optional	dct:accessRights <http://publications.europa.eu/resource/authority/access-right/RESTRICTED>;

Resources:

- [TWG Sandbox](#)
- [Data.europa.eu SPARQL query](#)
- [EU vocabularies](#)

Would you make the use of **dct:accessRights** as

- mandatory
- recommended
- or keep it as is (optional)?

Provide any comments, suggestions or questions:

CONFORMS TO

This property refers to an implementing rule or other specification.

Example:

	<code>[dct:conformsTo dct:Standard 0..n]</code>
Optional	<code>dct:conformsTo <http://data.europa.eu/eli/reg/2010/1089>; dct:conformsTo [a dct:Standard; dct:issued "2013-04-04"^^<http://www.w3.org/2001/XMLSchema#date>; dct:title "Technical Guidance for the implementation of INSPIRE View Services"@en];</code>

Resources:

- [TWG Sandbox](#)
- [Data.europa.eu SPARQL query](#)
- [OMOP Common Data Model](#)

Would you make the use of **dct:conformsTo** as

- mandatory
- recommended
- or keep it as is (optional)?

Provide any comments, suggestions or questions:

CREATOR

This property refers to the entity responsible for producing the dataset.

Example:

	[dct:creator foaf:Agent 0..n]
Optional	<pre>dct:creator [a foaf:Organization; locn:address [a locn:Address; locn:adminUnitL1 "DEU"; locn:postCode "D-21109"; locn:postName "Hamburg"; locn:thoroughfare "Neuenfelder Strasse 19"]; foaf:mbox <mailto:anselm.sprandel@bukea.hamburg.de>; foaf:name "Behörde für Umwelt, Klima, Energie und Agrarwirtschaft (BUKEA)"; foaf:phone <tel:+%2B49%2040%204%2028%2040%20-%202085>];</pre> <pre>dct:creator <https://opendata.schleswig-holstein.de/organization/94498aaa-9bd0-4ac0-ae70-db319a3ca1d8>;</pre>

Would you make the use of **dct:creator** as

- mandatory
- recommended
- or keep it as is (optional)?

Provide any comments, suggestions or questions:

DOCUMENTATION

This property refers to a page or document about this Dataset.

Example:

	[foaf:page foaf:Document 0..n]
Optional	<pre>foaf:page <http://dati.comune.milano.it/dataset/ds1069_ricoveri-ordinari-tumori> .</pre>

Resources:

- [TWG Sandbox \(Completeness +/-80%\)](#)
- [Data.europa.eu SPARQL query](#)

Would you make the use of **foaf:page** as

- mandatory
- recommended
- or keep it as is (optional)?

Provide any comments, suggestions or questions:

FREQUENCY

This property refers to the frequency at which the Dataset is updated.

Example:

	[dct:accrualPeriodicity foaf:Document 0..1]
Optional	dct:accrualPeriodicity <http://publications.europa.eu/resource/authority/frequency/ANNUAL>;

Resources:

- [TWG Sandbox \(Completeness +/-75%\)](#)
- [Data.europa.eu SPARQL query](#)
- [EU vocabularies](#)

Would you make the use of **dct:accrualPeriodicity** as

- mandatory
- recommended
- or keep it as is (optional)?

Provide any comments, suggestions or questions:

HAS VERSION

This property refers to a related Dataset that is a version, edition, or adaptation of the described Dataset.

Example:

	[dct:hasVersion dcat:Dataset 0..n]
O p t i o n a l	dct:hasVersion <https://opendata.schleswig-holstein.de/dataset/91ca0886-0090-497a-8fc0-bd3055d72191> , <https://opendata.schleswig-holstein.de/dataset/627f1756-d865-43f9-b933-73b82c9826c7> , <https://opendata.schleswig-holstein.de/dataset/33fdd874-c6e1-4ae8-96cb-8860c7e3163e> , <https://opendata.schleswig-holstein.de/dataset/1c38ac2c-3fdd-40bd-af44-ba09526d7586> , ...

Resources:

- [TWG Sandbox \(Completeness +/-16%\)](#)
- [Data.europa.eu SPARQL query](#)

Would you make the use of **dct:hasVersion** as

- mandatory
- recommended
- or keep it as is (optional)?

Provide any comments, suggestions or questions:

IS REFERENCED BY

This property is about a related resource, such as a publication, that references, cites, or otherwise points to the dataset.

Example:

	[dct:isReferencedBy rdfs:Resource 0..n]
Optional	dct:isReferencedBy <https://www.ine.es/dyngs/IOE/es/operacion.htm?numinv%3D30418>;

Resources:

- [TWG Sandbox \(Completeness +/-40%\)](#)
- [Data.europa.eu SPARQL query](#)

Would you make the use of **dct:isReferencedBy** as

- mandatory
- recommended
- or keep it as is (optional)?

Provide any comments, suggestions or questions:

IS VERSION OF

This property refers to a related Dataset of which the described Dataset is a version, edition, or adaptation.

Example:

	[dct:isVersionOf dcat:Dataset 0..n]
Optional	dct:isVersionOf <https://opendata.schleswig-holstein.de/dataset/a4c09d4b-9922-40f2-8615-4f4d89ff339f>;

Resources:

- [TWG Sandbox \(Completeness +/-9%\)](#)
- [Data.europa.eu SPARQL query](#)

Would you make the use of **dct:isVersionOf** as

- mandatory
- recommended
- or keep it as is (optional)?

Provide any comments, suggestions or questions:

LANDING PAGE

This property refers to a web page that provides access to the Dataset, its Distributions and/or additional information. It is intended to point to a landing page at the original data provider, not to a page on a site of a third party, such as an aggregator.

Example:

	[dcat:landingPage foaf:Document 0..n]
Optional	dcat:landingPage < ">https://www.pxweb.bfs.admin.ch/Default.aspx?> >;

Resources:

- [TWG Sandbox \(Completeness +/-77%\)](#)
- [Data.europa.eu SPARQL query](#)

Would you make the use of **dcat:landingPage** as

- mandatory
- recommended
- or keep it as is (optional)?

Provide any comments, suggestions or questions:

LANGUAGE

This property refers to a language of the Dataset. This property can be repeated if there are multiple languages in the Dataset.

Example:

	[dct:language dct:LinguisticSystem 0..n]
Optional	dct:language <http://publications.europa.eu/resource/authority/language/NLD>;

Resources:

- [TWG Sandbox \(Completeness +/-82%\)](#)
- [Data.europa.eu SPARQL query](#)
- [EU vocabularies](#)

Would you make the use of **dct:language** as

- mandatory
- recommended
- or keep it as is (optional)?

Provide any comments, suggestions or questions:

MODIFICATION DATE

This property contains the most recent date on which the Dataset was changed or modified.

Example:

	[dct:modified rdfs:Literal typed as xsd:date, xsd:dateTime, xsd:gYear or xsd:gYearMonth 0..1]

Optional	<code>dct:modified "2023-08-19T18:47:54Z"^^<http://www.w3.org/2001/XMLSchema#dateTime>;</code>
----------	--

Resources:

- [TWG Sandbox \(Completeness +/-23%\)](#)
- [Data.europa.eu SPARQL query](#)

Would you make the use of **dct:modified** as

- mandatory
- recommended
- or keep it as is (optional)?

Provide any comments, suggestions or questions:

OTHER IDENTIFIER

This property refers to a secondary identifier of the Dataset, such as MAST/ADS , DataCite , DOI , EZID or W3ID .

Example:

	<code>[adms:identifier adms:Identifier 0..n]</code>
Optional	<code>adms:identifier <https://doi.org/10.2906/101099100099/1>;</code> <code>adms:identifier "A IV 3 - j/12 HH";</code>

Resources:

- [TWG Sandbox \(Completeness +/-7%\)](#)
- [Data.europa.eu SPARQL query](#)

Would you make the use of **adms:identifier** as

- mandatory
- recommended

or keep it as is (optional)?

Provide any comments, suggestions or questions:

PROVENANCE

This property contains a statement about the lineage of a Dataset.

Example:

	[dct:provenance dct:ProvenanceStatement 0..n]
O p t i o n a l	dct:provenance [a dct:ProvenanceStatement; rdfs:label "Die Verkehrswegeachsen werden von den Gebietskoerperschaften bzw. Infrastrukturbetreibern fuer den Graphen anhand von Orthophotos digitalisiert. Sie sind mindestens fuer einen ModellmaBstab von 1:10000 erfasst."@de];

Resources:

- [TWG Sandbox \(Completeness +/-9%\)](#)
- [Data.europa.eu SPARQL query](#)

Would you make the use of **dct:provenance** as

- mandatory
 recommended
 or keep it as is (optional)?

Provide any comments, suggestions or questions:

QUALIFIED ATTRIBUTION

This property refers to a link to an Agent having some form of responsibility for the resource.

Example:

	[prov:qualifiedAttribution prov:Attribution 0..n]
Optional	<pre>prov:qualifiedAttribution [a prov:Attribution; dcat:hadRole <http://inspire.ec.europa.eu/metadata-codelist/ResponsiblePartyRole/pointOfContact>; prov:agent [a foaf:Organization; locn:address [a locn:Address; locn:adminUnitL1 "Deutschland"; locn:adminUnitL2 "Nordrhein-Westfalen"; locn:postCode "53175"; locn:postName "Bonn"; locn:thoroughfare "Heinemannstr. 6"]; foaf:homepage <file:///usr/verticles/www.eisenbahn.bundesamt.de>; foaf:mbox <mailto:Ref53@eba.bund.de>; foaf:name "Eisenbahn-Bundesamt"; foaf:phone <tel:+%2B49%280%29228%209826-0>]];</pre>

Resources:

- [TWG Sandbox \(Completeness +/-36%\)](#)
- [Data.europa.eu SPARQL query](#)

Would you make the use of **prov:qualifiedAttribution** as

- mandatory
 recommended
 or keep it as is (optional)?

Provide any comments, suggestions or questions:

QUALIFIED RELATION

This property provides a link to a description of a relationship with another resource.

Example:

	[dcat:qualifiedRelation dcat:Relationship 0..n]
Optional	dcat:qualifiedRelation [a dcat:Relationship ; dct:relation <http://example.org/Original987> ; dcat:hadRole <http://www.iana.org/assignments/relation/original>] .

Resources:

- [TWG Sandbox \(Completeness +/-9%\)](#)
- [Data.europa.eu SPARQL query](#)

Would you make the use of **dcat:qualifiedRelation** as

- mandatory
 recommended
 or keep it as is (optional)?

Provide any comments, suggestions or questions:

RELATED RESOURCE

This property refers to a related resource.

Example:

	[dct:relation rdfs:Resource 0..n]
Optional	dct:relation < https://www.fedlex.admin.ch/eli/cc/1995/1328_1328_1328/fr > , < https://www.fedlex.admin.ch/eli/cc/1993/2080_2080_2080/fr > , ...

Resources:

- [TWG Sandbox \(Completeness +/-18%\)](#)
- [Data.europa.eu SPARQL query](#)

Would you make the use of **dct:relation** as

- mandatory
- recommended
- or keep it as is (optional)?

Provide any comments, suggestions or questions:

RELEASE DATE

This property contains the date of formal issuance (e.g., publication) of the Dataset.

Example:

	[dct:issued rdfs:Literal typed as xsd:date, xsd:dateTime, xsd:gYear or xsd:gYearMonth 0..1]
Optional	dct:issued "2023-08-19T18:47:54Z"^^<http://www.w3.org/2001/XMLSchema#dateTime>;

Resources:

- [TWG Sandbox \(Completeness +/-36%\)](#)
- [Data.europa.eu SPARQL query](#)

Would you make the use of **dct:issued** as

- mandatory
- recommended
- or keep it as is (optional)?

Provide any comments, suggestions or questions:

SAMPLE

This property refers to a sample distribution of the dataset.

**This metadata property is of range `dc:Distribution` and will be discussed in a further TWG session.
[adms:sample `dc:Distribution 0..n`]**

SOURCE

This property refers to a related Dataset from which the described Dataset is derived.

Example:

	[<code>dct:source dc:Dataset 0..n</code>]
Optional	<code>dct:source <https://opendata.schleswig-holstein.de/dataset/a4c09d4b-9922-40f2-8615-4f4d89ff339f>;</code>

Resources:

- [TWG Sandbox \(Completeness +/-16%\)](#)
- [Data.europa.eu SPARQL query](#)

Would you make the use of **dct:source** as

- mandatory
- recommended
- or keep it as is (optional)?

Provide any comments, suggestions or questions:

SPATIAL RESOLUTION

This property refers to the minimum spatial separation resolvable in a dataset, measured in meters.

Example:

	[dcat:spatialResolutionInMeters rdfs:Literal typed as xsd:decimal 0..1]
Optional	dcat:spatialResolutionInMeters "10"^^<http://www.w3.org/2001/XMLSchema#decimal>;

Resources:

- [TWG Sandbox \(Completeness +/-25%\)](#)
- [Data.europa.eu SPARQL query](#)

Would you make the use of **dcat:spatialResolutionInMeters** as

- mandatory
- recommended
- or keep it as is (optional)?

Provide any comments, suggestions or questions:

TEMPORAL RESOLUTION

This property refers to the minimum time period resolvable in the dataset.

Example:

	[dcat:temporalResolution rdfs:Literal typed as xsd:duration 0..1]
Optional	dcat:temporalResolution "P1D"^^<http://www.w3.org/2001/XMLSchema#duration>; dcat:temporalResolution "PT24H"^^<http://www.w3.org/2001/XMLSchema#duration>; dcat:temporalResolution "P0.14285W"^^<http://www.w3.org/2001/XMLSchema#duration>;

Resources:

- [TWG Sandbox \(Completeness +/-14%\)](#)
- [Data.europa.eu SPARQL query](#)

Would you make the use of **dcat:temporalResolution** as

- mandatory
- recommended
- or keep it as is (optional)?

Provide any comments, suggestions or questions:

TYPE

This property refers to the type of the Dataset. A recommended controlled vocabulary data-type is foreseen.

Example:

	[dct:type skos:Concept 0..n]
Optional	dct:type <http://publications.europa.eu/resource/authority/dataset-type/SYNTHETIC_DATA>;

Resources:

- [TWG Sandbox \(Completeness +/-14%\)](#)
- [Data.europa.eu SPARQL query](#)
- [EU vocabularies](#)

Would you make the use of **dct:type** as

- mandatory
- recommended
- or keep it as is (optional)?

Provide any comments, suggestions or questions:

VERSION

This property contains a version number or other version designation of the Dataset.

Example:

	<code>[owl:versionInfo rdfs:Literal 0..1]</code>
Optional	<code>owl:versionInfo "Statistical Bulletin 2023";</code>

Resources:

- [TWG Sandbox \(Completeness +/-20%\)](#)
- [Data.europa.eu SPARQL query](#)

Would you make the use of **owl:versionInfo** as

- mandatory
- recommended
- or keep it as is (optional)?

Provide any comments, suggestions or questions:

VERSION NOTES

This property contains a description of the differences between this version and a previous version of the Dataset. This property can be repeated for parallel language versions of the version notes.

Example:

	<code>[adms:versionNotes rdfs:Literal 0..n]</code>
O p t i	<code>adms:versionNotes "XML file v2.0 has been available since 2020-12-09. All information regarding the shortage is displayed for each medicinal package separately in a structured manner, for example dates relevant</code>

o n al	for the shortage, contact information and possible alternative treatment options."@en;
--------------	--

Resources:

- [TWG Sandbox \(Completeness +/-18%\)](#)
- [Data.europa.eu SPARQL query](#)

Would you make the use of **adsm:versionNotes** as

- mandatory
- recommended
- or keep it as is (optional)?

Provide any comments, suggestions or questions:

WAS GENERATED BY

This property refers to an activity that generated, or provides the business context for, the creation of the dataset.

Example:

	[prov:wasGeneratedBy prov:Activity 0..n]
Optional	<pre> prov:wasGeneratedBy [a prov:Activity; dcterms:type <http://dbpedia.org/resource/Observation> ; prov:startedAtTime "2000-11-01"^^xsd:date ; prov:used dap:Parkes-radio-telescope ; prov:wasInformedBy dap:ATNF ; rdfs:label "P366 - Parkes multibeam high-latitude pulsar survey" @en ; rdfs:seeAlso <https://doi.org/10.1111/j.1365-2966.2006.10100.x>]; </pre>

Resources:

- [TWG Sandbox \(Completeness +/-23%\)](#)
- [Data.europa.eu SPARQL query](#)

Would you make the use of **prov:wasGeneratedBy** as

- mandatory
- recommended
- or keep it as is (optional)?

Provide any examples of "activities" generating health datasets (e.g.: [EHR](#) or [Survey](#), ...)

Provide any comments, suggestions or questions:

Contact

[Contact Form](#)

TWG session held on the 13-10-2023 - Survey n°4 & 4b

Survey 4 - Topics covered: "What, why, how, ..."

- Create rich metadata record – Data holder perspective
- Analyses of the HDRUK health Gateway and HealthInformationPortal.eu
- Use case: create metadata without administrative burden

The TWG survey no. 4 addresses several aspects of the Health DCAT extension. It reviews metadata records of two popular health data catalogues (i.e.: HDRUK Gateway and the Health Information Portal) and the European data portal, focusing on defining mandatory and recommended metadata properties from a data holder perspective. Licinio Kustra-Mano@EC, in a previous high-level meeting, highlighted the need to balance meaningful descriptions with the administrative workload. Key topics include the application of FAIR data principles, the completeness of metadata. The survey aims to assess the quality and completeness of metadata and solicit feedback on potential improvements.

Survey 4b - Topics covered:

- dct:alternative
- schema:typicalAgeRange
- schema:codingSyst
- schema:inCodeSet

The TWG survey no. 4bis focuses on introducing new properties not currently mapped in DCAT-AP for the HealthInformationPortal.eu and HDRUK Gateway. These include properties like 'associated national node', 'acronym', 'datasource url status', and others. The survey seeks feedback on the necessity and implementation of these new properties, specifically asking about the introduction and categorisation (mandatory, recommended, or optional) of properties like dct:alternative for alternative names, schema:typicalAgeRange for age groups, and schema:codingSystem for coding systems. Participants are encouraged to provide comments or suggestions on these proposals.

Minutes:

1. Opening Remarks: Pascal Derycke starts the session by addressing a previous technical issue and sharing the agenda, which includes reviewing survey results and discussing new surveys on metadata records.
2. Survey Results and New Surveys: Derycke reviews the results of a previous survey and introduces new surveys. One survey focuses on creating rich metadata records from a data holder perspective, and another on new property relevance for the health domain.
3. Metadata Records Analysis: The session moves into a detailed analysis of metadata records, including discussions on the completeness of metadata and the use of controlled vocabularies.
4. Challenges in Metadata Management: Various challenges in managing metadata are discussed, such as ensuring meaningful descriptions and balancing administrative burden. The importance of understanding properties in metadata and ensuring accuracy is emphasized.
5. Interactive Discussion: The session includes interactive discussions with participants like Truls Korsgaard and Bert Van Nuffelen, addressing technical and conceptual questions about metadata properties and their implementation.
6. Technical Insights and Best Practices: Technical insights into the use of metadata, including the use of global identifiers and control vocabularies, are shared. The

discussion also covers best practices for categorizing and tagging datasets in health data portals.

7. Closing Remarks and Future Directions: The session concludes with plans for future discussions on the application of metadata in health data management, emphasizing the importance of aligning technical requirements with use cases.

Results of the survey n°4:

TGW survey n°4		Fre e tex t	Cod e list	URL	Mandator y	Recommen ded	Option al	tota l	Comments
dct:title	what	1							
dct:description	what	2							
dcat:keyword This property contains a keyword or tag describing the Dataset.	what	3			6	2	-	8	
dcat:contactPoint This property contains contact information that can be used for sending comments about the Dataset.	who			1 contact	6	2	-	8	
dct:spatial This property refers to a geographic region that is covered by the Dataset.	where		1		4	4	-	8	
dct:publisher This property refers to an entity (organisation) responsible for making the Dataset available.	who			2 contact	7	1	-	8	Various definitions across domains ex: data holder vs creator, ...
dct:temporal This property refers to a temporal period that the Dataset covers.	when	4			3	5	-	8	Ongoing data collection
dcat:theme This property refers to a category of the Dataset. A Dataset may be associated with multiple themes.	what		2		4	4	-	8	
dct:accessRights This property refers to information that indicates whether the Dataset is open data, has access restrictions or is not public.	How to access		3		5	3	0	8	
dct:conformsTo This property refers to an implementing rule or other specification.	what		4		0	5	3	8	Medical-data-models.org 25192 medical data models What is the difference between M and R? in case it can't always apply to the dataset : R
dct:creator This property refers to the entity responsible for producing the dataset.	who			3 contact	1	5	2	8	

foaf:page This property refers to a page or document about this Dataset	what			4 resource	1	4	2	-1	
dct:accrualPeriodicity This property refers to the frequency at which the Dataset is updated.	when		5		2	4	2		EMA provides extra info to indicate which month it is updated. Dct:modified says when it has been updated last time
dct:hasVersion This property refers to a related Dataset that is a version, edition, or adaptation of the described Dataset.	Related			5 Dataset	2	3	3		
dct:isReferencedBy This property is about a related resource, such as a publication, that references, cites, or otherwise points to the dataset.	Related			6 Resource	0	3	5		
dct:isVersionOf This property refers to a related Dataset of which the described Dataset is a version, edition, or adaptation.	Related			7 Dataset	0	3	5		
dcat:landingPage This property refers to a web page that provides access to the Dataset, its Distributions and/or additional information. It is intended to point to a landing page at the original data provider, not to a page on a site of a third party, such as an aggregator.	How to access			8 Resource	3	3	2		According to definition the landing page provides access to the dataset Dcat:distribution > accessURL > HDAB
dct:language This property refers to a language of the Dataset. This property can be repeated if there are multiple languages in the Dataset.	what		6		3	4	0	-1	
dct:modified This property contains the most recent date on which the Dataset was changed or modified.	when	5			2	4	2		
adms:identifier This property refers to a secondary identifier of the Dataset, such as MAST/ADS , DataCite , DOI , EZID or W3ID .	what				0	3	4	-1	
dct:provenance This property contains a statement about the lineage of a Dataset.	how	6			2	4	2		
prov:qualifiedAttribution	who			9 Contact	1	3	2	-2	Is the data holder in

This property refers to a link to an Agent having some form of responsibility for the resource.									case of data linkage such an agent having a form of responsibility? Or CSI contact point?
dcat:qualifiedRelation This property provides a link to a description of a relationship with another resource.	Related			10 Resource	0	2	5	-1	Need examples! hadRole = ex:registry
dct:relation This property refers to a related resource.	Related			11 Resource	0	0	7	-1	Useful links should not be more than optional?
dct:issued This property contains the date of formal issuance (e.g., publication) of the Dataset.	when	7			2	4	1	-1	
adms:sample This property refers to a sample distribution of the dataset.	Access			12 Dataset	0	5	3		Possibility to show the variables
dct:source This property refers to a related Dataset from which the described Dataset is derived.	Related			13 Dataset	1	2	3	1	Ex: of a linkage situation
dcat:spatialResolutionInMeters This property refers to the minimum spatial separation resolvable in a dataset, measured in meters.	where	8			0	2	6		
dcat:temporalResolution This property refers to the minimum time period resolvable in the dataset.	when		7		0	3	3	-2	
dct:type This property refers to the type of the Dataset. A recommended controlled vocabulary data-type is foreseen.	what		8		1	5	1	-1	A code list helps to mitigate uncertainty how to fill the property
owl:versionInfo This property contains a version number or other version designation of the Dataset.	what	9			3	1	4		
adms:versionNotes This property contains a description of the differences between this version and a previous version of the Dataset. This property can be repeated for parallel language versions of the version notes.	what	10			0	4	4		
prov:wasGeneratedBy This property refers to an activity that generated, or provides the business context for, the creation of the dataset.	how	11			1	2	4	-1	

		12	8	13				
--	--	----	---	----	--	--	--	--

Survey n°4 is available at [this link](#)

Results of the survey n°4b:

TGW survey n°4b	yes	no	total	Comments
dct :alternative	9	0	9	
Schema :typicalAgeRange	5	3	8	
Schema :codingSyst	6	2	8	
Schema :inCodeSet	4	2	6	

Survey n°4b is available at [this link](#)

Health DCAT extension - TWG n° 4 13-10-2023

Fields marked with * are mandatory.

DCAT Health extension

Technical Working Group session 13-10-2023

* Your email:

pascal.derycke@sciensano.be

Nota bene: This form remains open until Monday October 23

Review of the metadata records of two catalogues having a metadata governance:

- <https://www.healthdatagateway.org> (650 records)
- <https://www.healthinformationportal.eu> (265 records)
-

Review of the health metadata records of the European data portal:

- <https://data.europa.eu> (23 176 records)

Objective of the session: define mandatory/recommended metadata properties from a data holder perspective

- Licinio Kustra-Mano@EC: "Balance meaningful descriptions with administrative burden"

Requirement derived from applying FAIR data principles:

Findable

F2. Data are described with rich metadata (defined by R1 below)	Define a short list of important metadata properties as mandatory properties to achieve a satisfying level of discoverability/reusability.
---	--

Reusable

R1. (Meta)data are richly described with a plurality of accurate and relevant attributes	See F1
--	--------

Challenges:

- [Report on DCAT-AP use \(2018\)](#)
- Bert Van Nuffelen: "The quality and completeness of the metadata in data.europa.eu is a EU wide challenge"

Level of completeness	HDRUK gateway	Healthinformationportal	data.europa.eu *
[dct:title rdfs:Literal 1..n]	100%	100%	100%
[dct:description rdfs:Literal 1..n]	99%	100%	100%
[dcat:keyword rdfs:Literal 0..n]	93%	100%	82%
[dcat:contactPoint vcard:Kind 0..n]	100%	100%	43%
[dct:spatial dct:Location 0..n]	95%	100%	52%
[dct:publisher foaf:Agent 0..1]	100%	80%	86%
[dct:temporal dct:PeriodOfTime 0..n]	97%	100%	24%
[dcat:theme, subproperty of dct:subject skos:Concept 0..n]	93%	99%	76%
[dct:accessRights dct:RightsStatement 0..1]	100%	–	29%
[dct:conformsTo dct:Standard 0..n]	90%	–	26%
[dct:creator foaf:Agent 0..n]	81%	95%	2%
[foaf:page foaf:Document 0..n]	31%	–	8%

[dct:accrualPeriodicity foaf:Document 0..1]	96%	54%	33%
[dct:hasVersion dcat:Dataset 0..n]	-	-	0%
[dct:isReferencedBy rdfs:Resource 0..n]	27%	-	14%
[dct:isVersionOf dcat:Dataset 0..n]	43%	-	9%
[dcat:landingPage foaf:Document 0..n]	60%	100%	22%
[dct:language dct:LinguisticSystem 0..n]	90%	100%	73%
[dct:modified rdfs:Literal typed as xsd:date, xsd:dateTime, xsd:gYear or xsd:gYearMonth 0..1]	-	-	79%
[adms:identifier adms:Identifier 0..n]	12%	9%	2%
[dct:provenance dct:ProvenanceStatement 0..n]	79%	100%	20%
[prov:qualifiedAttribution prov:Attribution 0..n]	-	-	10%
[dcat:qualifiedRelation dcat:Relationship 0..n]	53%	%	0%
[dct:relation rdfs:Resource 0..n]	-	-	1%
[dct:issued rdfs:Literal typed as xsd:date, xsd:dateTime, xsd:gYear or xsd:gYearMonth 0..1]	46%	-	59%
[adms:sample dcat:Distribution 0..n]/td>	-	-	0.4%
[dct:source dcat:Dataset 0..n]	-	-	0.2%
[dcat:spatialResolutionInMeters rdfs:Literal typed as xsd:decimal 0..1]	-	88%	0.1%
[dcat:temporalResolution rdfs:Literal typed as xsd:duration 0..1]	-	-	0%

[dct:type skos:Concept 0..n]	-	-	14%
[owl:versionInfo rdfs:Literal 0..1]	-	-	3%
[adms:versionNotes rdfs:Literal 0..n]	-	-	0.06%
[prov:wasGeneratedBy prov:Activity 0..n]	77%	-	0%

* Percentages provided for the 23176 DCAT-AP, GeoDCAT, StatDCAT metadata records of the data.europa.eu portal on the 11/10/2023

Provide any comments, suggestions or questions:

DCAT-AP has 2 mandatory properties: "Title" and "Description".

TITLE

This property contains a name given to the Dataset. This property can be repeated for parallel language versions of the name.

Example:

	[dct:title rdfs:Literal 1..n]
Mandatory	dct:title "Austrian Corona Panel Project"@en dct:title "Project panel Corona autrichien"@fr .

DESCRIPTION

This property contains a free-text account of the Dataset. This property can be repeated for parallel language versions of the description.

Example:

	[dct:description rdfs:Literal 1..n]
--	-------------------------------------

M a n d a t a b a s e	dct:description "The City of Hamburg (Environmental and Energy Authority) accompanies, supports and promotes energetic neighbourhood concepts that illuminate measures to increase..."@en .
---	---

KEYWORD

This property contains a keyword or tag describing the Dataset.

Example:

	[dcat:keyword rdfs:Literal 0..n]
Recommended	dcat:keyword "COVID-19"@en, "SARS-CoV-2"@en, "Vaccination"@en, "Test"@en .

Resources:

- [TWG Sandbox \(completeness +/- 80% *\)](#)
- [HDRUK Gateway \(93%\)](#)
- [HealthInformationPortal.eu \(100%\)](#)
- [Data.europa.eu SPARQL query \(82%\)](#)

* The information provided is not always accurate

Would you make the use of **dcat:keyword** as

- mandatory
 or keep it as is (recommended)?

Provide any comments, suggestions or questions:

CONTACT POINT

This property contains contact information that can be used for sending comments about the Dataset.

Example:

	[dcat:contactPoint vcard:Kind 0..n]
Recommended	<pre>dcat:contactPoint [a vcard:Organization; vcard:fn "Behörde für Umwelt, Klima, Energie und Agrarwirtschaft (BUKEA) "; vcard:hasAddress [a vcard:Address; vcard:country-name "DEU"; vcard:locality "Hamburg"; vcard:postal-code "D-21109"; vcard:street-address "Neuenfelder Strasse 19"]; vcard:hasEmail <mailto:waermekataster@bukea.hamburg.de>];</pre>

Resources:

- [TWG Sandbox \(Completeness +/- 84%\)](#)
- [HDRUK Gateway \(100%\)](#)
- [HealthInformationPortal.eu \(100%\)](#)
- [Data.europa.eu SPARQL query \(43%\)](#)

Would you make the use of **dcat:contactPoint** as

- mandatory
 or keep it as is (recommended)?

Provide any comments, suggestions or questions:

GEOGRAPHICAL COVERAGE

This property refers to a geographic region that is covered by the Dataset.

Example:

	[dct:spatial dct:Location 0..n]
--	---------------------------------

R e c o m m e n d ed	<pre>dct:spatial <https://sws.geonames.org/3337388/>; dct:spatial [a dct:Location; locn:geometry "{\"type\": \"Polygon\", \" coordinates\": [[[10.326304, 53.394985], [10.326304, 53.964153], [8.420551, 53.964153], [8.420551, 53.394985], [10.326304, 53.394985]]]}\"^<https://replacement.io/assignments/media-types /application/vnd.geo+json> , \"POLYGON ((10.3263 53.3950, 10.3263 53.9642, 8.4206 53.9642, 8.4206 53.3950, 10.3263 53.3950))\" ^^<http://replacement.io/ont/geosparql#wktLiteral>]; dct:spatial [a dct:Location; locn:adminUnitL2 <http://dcat-ap.de/def /politicalGeocoding/regionalKey/020000000000>];</pre>
---	--

Resources:

- [TWG Sandbox \(Completeness +/- 82%\)](#)
- [HDRUK Gateway \(95%\)](#)
- [HealthInformationPortal.eu \(100%\)](#)
- [Data.europa.eu SPARQL query \(52%\)](#)

The EU controlled vocabulary "[Countries and territories](#)" MUST be used. (or <http://publications.europa.eu/resource/authority/place/>, <http://sws.geonames.org/>)

Would you make the use of **dct:spatial** as

- mandatory
- or keep it as is (recommended)?

Provide any comments, suggestions or questions:

PUBLISHER

This property refers to an entity (organisation) responsible for making the Dataset available.

Example:

	[dct:publisher foaf:Agent 0..1]
--	---------------------------------

R e c o m m e n d e d	<pre>dct:publisher [a foaf:Organization; locn:address [a locn:Address; locn:adminUnitL1 "DEU"; locn:postCode "D-21109"; locn:postName "Hamburg"; locn:thoroughfare "Neuenfelder Strasse 19"]; foaf:mbox <mailto:anselm.sprandel@bukea.hamburg.de>; foaf:name "Behörde für Umwelt, Klima, Energie und Agrarwirtschaft (BUKEA)"; foaf:phone <tel:+% 2B49%2040%204%2028%2040%20-%202085>]; dct:publisher <https://opendata.schleswig-holstein.de/organization /94498aaa-9bd0-4ac0-ae70-db319a3ca1d8>;</pre>
---	---

Resources:

- [TWG Sandbox \(Completeness +/- 86%\)](#)
- [HDRUK Gateway \(100%\)](#)
- [HealthInformationPortal.eu \(80%\)](#)
- [Data.europa.eu SPARQL query \(86%\)](#)

Would you make the use of **dct:publisher** as

- mandatory
 or keep it as is (recommended)?

Provide any comments, suggestions or questions:

TEMPORAL COVERAGE

This property refers to a temporal period that the Dataset covers.

Example:

	[dct:temporal dct:PeriodOfTime 0..n]
--	--------------------------------------

Recommended	<pre>dct:temporal [a dct:PeriodOfTime; dcat:endDate "2019-11-26"^^<http://www.w3.org/2001 /XMLSchema#date>; dcat:startDate "2013-01-01"^^<http://www.w3.org/2001 /XMLSchema#date>];</pre>
-------------	---

Resources:

- [TWG Sandbox \(Completeness +/-86%\)](#)
- [HDRUK Gateway \(97%\)](#)
- [HealthInformationPortal.eu \(100%\)](#)
- [Data.europa.eu SPARQL query \(24%\)](#)

Would you make the use of **dct:temporal** as

- mandatory
 or keep it as is (recommended)?

Provide any comments, suggestions or questions:

THEME

This property refers to a category of the Dataset. A Dataset may be associated with multiple themes.

Example:

	[dcat:theme, subproperty of dct:subject skos:Concept 0..n]
Recommended	<pre>dcat:theme <http://publications.europa.eu/resource/authority/data- theme/ENVI> , <http://dcat-ap.ch/vocabulary/themes/territory> .</pre>

Resources:

- [TWG Sandbox \(Completeness +/-43%\)](#)
- [HDRUK Gateway \(93%\)](#)
- [HealthInformationPortal.eu \(99%\)](#)

- [Data.europa.eu SPARQL query \(76%*\)](#)

* for the 1 574 598 metadata records of the catalogue

- A controlled vocabulary will be established based on the Article 33 of the regulation
- Health datasets MUST be annotated with the following code: <http://publications.europa.eu/resource/authority/data-theme/HEAL> (This concept identifies datasets covering the domain of health. Health is a state of physical, mental and social well-being in which disease and infirmity are absent. Dataset examples: COVID-19 Coronavirus data; European Cancer Information System.)

Ontology search tools:

- <https://ontobee.org/>
- <https://www.wikidata.org/>
- [Ontologie search](#)

Would you make the use of **dcat:theme** as

- mandatory
- or keep it as is (recommended)?

Provide any comments, suggestions or questions:

ACCESS RIGHTS

This property refers to information that indicates whether the Dataset is open data, has access restrictions or is not public.

Example:

	[dct:accessRights dct:RightsStatement 0..1]
Optional	dct:accessRights < http://publications.europa.eu/resource/authority/access-right/RESTRICTED >;

Resources:

- [TWG Sandbox \(Completeness +/-80%\)](#)
- [HDRUK Gateway \(100%\)](#)
- [HealthInformationPortal.eu \(-\)](#)
- [Data.europa.eu SPARQL query \(29%\)](#)

The EU controlled vocabulary "[Access right](#)" MUST be used.

Would you make the use of **dct:accessRights** as

- mandatory
- recommended
- or keep it as is (optional)?

Provide any comments, suggestions or questions:

CONFORMS TO

This property refers to an implementing rule or other specification.

Example:

	[dct:conformsTo dct:Standard 0..n]
Optional	<pre>dct:conformsTo <http://data.europa.eu/eli/reg/2010/1089>; dct:conformsTo [a dct:Standard; dct:issued "2013-04-04"^^<http://www.w3.org/2001/XMLSchema#date>; dct:title "Technical Guidance for the implementation of INSPIRE View Services"@en];</pre>

Resources:

- [TWG Sandbox \(Completeness +/-39%\)](#)
- [HDRUK Gateway \(90%\)](#)
- [HealthInformationPortal.eu \(-\)](#)
- [Data.europa.eu SPARQL query \(26%\)](#)

Data models:

- OMOP
- FHIR
- OpenEHR
- ex: national data model

Provide any other health data models:

Would you make the use of **dct:conformsTo** as

- mandatory
- recommended
- or keep it as is (optional)?

Provide any comments, suggestions or questions:

CREATOR

This property refers to the entity responsible for producing the dataset.

Example:

	[dct:creator foaf:Agent 0..n]
O p t i o n a l	<pre>dct:creator [a foaf:Organization; locn:address [a locn:Address; locn:adminUnitL1 "DEU"; locn:postCode "D-21109"; locn:postName "Hamburg"; locn:thoroughfare "Neuenfelder Strasse 19"]; foaf:mbox <mailto:anselm.sprandel@bukea.hamburg.de>; foaf:name "Behörde für Umwelt, Klima, Energie und Agrarwirtschaft (BUKEA)"; foaf:phone <tel:+%2B49%2040%204%2028%2040%20-%202085>]; dct:creator <https://opendata.schleswig-holstein.de/organization/94498aaa-9bd0-4ac0-ae70-db319a3ca1d8>;</pre>

Resources:

- [TWG Sandbox \(Completeness +/-80%\)](#)
- [HDRUK Gateway \(81%\)](#)
- [HealthInformationPortal.eu \(95%\)](#)
- [Data.europa.eu SPARQL query \(2%\)](#)

Would you make the use of **dct:creator** as

- mandatory
- recommended
- or keep it as is (optional)?

Provide any comments, suggestions or questions:

DOCUMENTATION

This property refers to a page or document about this Dataset.

Example:

	[foaf:page foaf:Document 0..n]
Optional	foaf:page <http://dati.comune.milano.it/dataset/ds1069_ricoveri-ordinari-tumori> .

Resources:

- [TWG Sandbox \(Completeness +/-80%\)](#)
- [HDRUK Gateway \(31%\)](#)
- [HealthInformationPortal.eu \(-\)](#)
- [Data.europa.eu SPARQL query \(8%\)](#)

Would you make the use of **foaf:page** as

- mandatory
- recommended
-

or keep it as is (optional)?

Provide any comments, suggestions or questions:

FREQUENCY

This property refers to the frequency at which the Dataset is updated.

Example:

	[dct:accrualPeriodicity foaf:Document 0..1]
Optional	dct:accrualPeriodicity <http://publications.europa.eu/resource/authority/frequency/ANNUAL>;

Resources:

- [TWG Sandbox \(Completeness +/-75%\)](#)
- [HDRUK Gateway \(96%\)](#)
- [HealthInformationPortal.eu \(54%\)](#)
- [Data.europa.eu SPARQL query \(33%\)](#)

The EU controlled vocabulary "[Frequency](#)" MUST be used.

Would you make the use of **dct:accrualPeriodicity** as

- mandatory
- recommended
- or keep it as is (optional)?

Provide any comments, suggestions or questions:

HAS VERSION

This property refers to a related Dataset that is a version, edition, or adaptation of the described Dataset.

Example:

	[dct:hasVersion dcat:Dataset 0..n]
O p t i o n a l	dct:hasVersion <https://opendata.schleswig-holstein.de/dataset/91ca0886-0090-497a-8fc0-bd3055d72191> , <https://opendata.schleswig-holstein.de/dataset/627f1756-d865-43f9-b933-73b82c9826c7> , <https://opendata.schleswig-holstein.de/dataset/33fdd874-c6e1-4ae8-96cb-8860c7e3163e> , <https://opendata.schleswig-holstein.de/dataset/1c38ac2c-3fdd-40bd-af44-ba09526d7586> , ...

Resources:

- [TWG Sandbox \(Completeness +/-16%\)](#)
- [HDRUK Gateway \(-\)](#)
- [HealthInformationPortal.eu \(-\)](#)
- [Data.europa.eu SPARQL query \(0%\)](#)

Metadata PURI identifiers MUST to be used with dct:hasVersion, dct:isVersionOf and dct:source.

Would you make the use of **dct:hasVersion** as

- mandatory
 recommended
 or keep it as is (optional)?

Provide any comments, suggestions or questions:

IS REFERENCED BY

This property is about a related resource, such as a publication, that references, cites, or otherwise points to the dataset.

Example:

	[dct:isReferencedBy rdfs:Resource 0..n]
Optional	dct:isReferencedBy <https://www.ine.es/dyngs/IOE/es/operacion.htm?numinv%3D30418>;

Resources:

- [TWG Sandbox \(Completeness +/-40%\)](#)
- [HDRUK Gateway \(27%\)](#)
- [HealthInformationPortal.eu \(-\)](#)
- [Data.europa.eu SPARQL query \(14%\)](#)

Would you make the use of **dct:isReferencedBy** as

- mandatory
- recommended
- or keep it as is (optional)?

Provide any comments, suggestions or questions:

IS VERSION OF

This property refers to a related Dataset of which the described Dataset is a version, edition, or adaptation.

Example:

	[dct:isVersionOf dcat:Dataset 0..n]
Optional	dct:isVersionOf <https://opendata.schleswig-holstein.de/dataset/a4c09d4b-9922-40f2-8615-4f4d89ff339f>;

Resources:

- [TWG Sandbox \(Completeness +/-9%\)](#)

- [HDRUK Gateway \(43%\)](#)
- [HealthInformationPortal.eu \(-\)](#)
- [Data.europa.eu SPARQL query \(9%\)](#)

Metadata PURI identifiers MUST to be used with dct:hasVersion, dct:isVersionOf and dct:source.

Would you make the use of **dct:isVersionOf** as

- mandatory
- recommended
- or keep it as is (optional)?

Provide any comments, suggestions or questions:

LANDING PAGE

This property refers to a web page that provides access to the Dataset, its Distributions and/or additional information. It is intended to point to a landing page at the original data provider, not to a page on a site of a third party, such as an aggregator.

Example:

	[dcat:landingPage foaf:Document 0..n]
Optional	dcat:landingPage <https://www.pxweb.bfs.admin.ch/Default.aspx?>;

Resources:

- [TWG Sandbox \(Completeness +/-77%\)](#)
- [HDRUK Gateway \(60%\)](#)
- [HealthInformationPortal.eu \(100%\)](#)
- [Data.europa.eu SPARQL query \(22%\)](#)

Would you make the use of **dcat:landingPage** as

- mandatory
- recommended
- or keep it as is (optional)?

Provide any comments, suggestions or questions:

LANGUAGE

This property refers to a language of the Dataset. This property can be repeated if there are multiple languages in the Dataset.

Example:

	[dct:language dct:LinguisticSystem 0..n]
Optional	dct:language <http://publications.europa.eu/resource/authority/language/NLD>;

Resources:

- [TWG Sandbox \(Completeness +/-82%\)](#)
- [HDRUK Gateway \(90%\)](#)
- [HealthInformationPortal.eu \(100%\)](#)
- [Data.europa.eu SPARQL query \(73%\)](#)

The EU controlled vocabulary "[Language](#)" MUST be used.

Would you make the use of **dct:language** as

- mandatory
- recommended
- or keep it as is (optional)?

Provide any comments, suggestions or questions:

MODIFICATION DATE

This property contains the most recent date on which the Dataset was changed or modified.

Example:

	[dct:modified rdfs:Literal typed as xsd:date, xsd:dateTime, xsd:gYear or xsd:gYearMonth 0..1]
Optional	dct:modified "2023-08-19T18:47:54Z"^^<http://www.w3.org/2001/XMLSchema#dateTime>;

Resources:

- [TWG Sandbox \(Completeness +/-23%\)](#)
- [HDRUK Gateway \(-\)](#)
- [HealthInformationPortal.eu \(-\)](#)
- [Data.europa.eu SPARQL query \(79%\)](#)

Would you make the use of **dct:modified** as

- mandatory
 recommended
 or keep it as is (optional)?

Provide any comments, suggestions or questions:

OTHER IDENTIFIER

This property refers to a secondary identifier of the Dataset, such as MAST/ADS , DataCite , DOI , EZID or W3ID .

Example:

	[adms:identifier adms:Identifier 0..n]
Optional	adms:identifier <https://doi.org/10.2906/101099100099/1>; adms:identifier "A IV 3 - j/12 HH";

Resources:

- [TWG Sandbox \(Completeness +/-7%\)](#)
- [HDRUK Gateway \(12%\)](#)
- [HealthInformationPortal.eu \(9%\)](#)
- [Data.europa.eu SPARQL query \(2%\)](#)

Would you make the use of **adms:identifier** as

- mandatory
- recommended
- or keep it as is (optional)?

Provide any comments, suggestions or questions:

PROVENANCE

This property contains a statement about the lineage of a Dataset.

Example:

	[dct:provenance dct:ProvenanceStatement 0..n]
O p t i o n a l	dct:provenance [a dct:ProvenanceStatement; rdfs:label "Die Verkehrswegeachsen werden von den Gebietskoerperschaften bzw. Infrastrukturbetreibern fuer den Graphen anhand von Orthophotos digitalisiert. Sie sind mindestens fuer einen Modellmaßstab von 1:10000 erfasst."@de];

Resources:

- [TWG Sandbox \(Completeness +/-9%\)](#)
- [HDRUK Gateway \(79%\)](#)
- [HealthInformationPortal.eu \(100%\)](#)
- [Data.europa.eu SPARQL query \(20%\)](#)

The HDRUK and HealthInformationPortal.eu use code lists for providing provenance information.
 HDRUK: CARE, IN-PATIENTS, ADMINISTRATIVE, TRIAL, DISEASE REGISTRY, STUDY, AUDIT, ...
 HealthInformationPortal.eu: SURVEY, ADMINISTRATIVE DATA, REGISTRY DATA, INTERVIEW DATA, HOSPITAL RESOURCES, ...

Would you make the use of **dct:provenance** as

- mandatory
- recommended
- or keep it as is (optional)?

Provide any comments, suggestions or questions:

QUALIFIED ATTRIBUTION

This property refers to a link to an Agent having some form of responsibility for the resource.

Example:

	[prov:qualifiedAttribution prov:Attribution 0..n]
Optional	<pre> prov:qualifiedAttribution [a prov:Attribution; dcat:hadRole <http://inspire.ec.europa.eu/metadata-codelist/ResponsiblePartyRole/pointOfContact>; prov:agent [a foaf:Organization; locn:address [a locn:Address; locn:adminUnitL1 "Deutschland"; locn:adminUnitL2 "Nordrhein-Westfalen"; locn:postCode "53175"; locn:postName "Bonn"; locn:thoroughfare "Heinemannstr. 6"]; foaf:homepage <file:///usr/verticles/www.eisenbahn.bundesamt.de>; foaf:mbox <mailto:Ref53@eba.bund.de>; foaf:name "Eisenbahn-Bundesamt"; foaf:phone <tel:+%2B49%280%29228%209826-0>]]; </pre>

Resources:

- [TWG Sandbox \(Completeness +/-36%\)](#)

- [HDRUK Gateway \(-\)](#)
- [HealthInformationPortal.eu \(-\)](#)
- [Data.europa.eu SPARQL query \(10%\)](#)

Would you make the use of **prov:qualifiedAttribution** as

- mandatory
- recommended
- or keep it as is (optional)?

Provide any comments, suggestions or questions:

QUALIFIED RELATION

This property provides a link to a description of a relationship with another resource.

Example:

	[dcat:qualifiedRelation dcat:Relationship 0..n]
Optional	<pre>dcat:qualifiedRelation [a dcat:Relationship ; dct:relation <http://example.org/Original987> ; dcat:hadRole <http://www.iana.org/assignments/relation/original>] .</pre>

Resources:

- [TWG Sandbox \(Completeness +/-9%\)](#)
- [HDRUK Gateway \(53%\)](#)
- [HealthInformationPortal.eu \(-\)](#)
- [Data.europa.eu SPARQL query \(0%\)](#)

Would you make the use of **dcat:qualifiedRelation** as

- mandatory
- recommended
- or keep it as is (optional)?

Provide any comments, suggestions or questions:

RELATED RESOURCE

This property refers to a related resource.

Example:

	[dct:relation rdfs:Resource 0..n]
Optional	dct:relation < https://www.fedlex.admin.ch/eli/cc/1995/1328_1328_1328/fr > , < https://www.fedlex.admin.ch/eli/cc/1993/2080_2080_2080/fr > , ...

Resources:

- [TWG Sandbox \(Completeness +/-18%\)](#)
- [HDRUK Gateway \(-\)](#)
- [HealthInformationPortal.eu \(-\)](#)
- [Data.europa.eu SPARQL query \(1%\)](#)

Would you make the use of **dct:relation** as

- mandatory
- recommended
- or keep it as is (optional)?

Provide any comments, suggestions or questions:

RELEASE DATE

This property contains the date of formal issuance (e.g., publication) of the Dataset.

Example:

	[dct:issued rdfs:Literal typed as xsd:date, xsd:dateTime, xsd:gYear or xsd:gYearMonth 0..1]
Optional	dct:issued "2023-08-19T18:47:54Z"^^<http://www.w3.org/2001/XMLSchema#dateTime>;

Resources:

- [TWG Sandbox \(Completeness +/-36%\)](#)
- [HDRUK Gateway \(46%\)](#)
- [HealthInformationPortal.eu \(-\)](#)
- [Data.europa.eu SPARQL query \(59%\)](#)

Would you make the use of **dct:issued** as

- mandatory
- recommended
- or keep it as is (optional)?

Provide any comments, suggestions or questions:

SAMPLE

This property refers to a sample distribution of the dataset.

Example:

	[adms:sample dcat:Distribution 0..n]
Optional	adms:sample [a dcat:Distribution ; dcat:downloadURL <https://www.healthinformationportal.eu/health-information-sources/my-life-survey>; dcat:mediaType "application/json-ld" ;] .

Resources:

- [TWG Sandbox \(Completeness +/-11%\)](#)
- [HDRUK Gateway \(-\)](#)
- [HealthInformationPortal.eu \(-\)](#)
- [Data.europa.eu SPARQL query \(0.4%\)](#)

Would you make the use of **adms:sample** as

- mandatory
- recommended
- or keep it as is (optional)?

Provide any comments, suggestions or questions:

SOURCE

This property refers to a related Dataset from which the described Dataset is derived.

Example:

	[dct:source dcat:Dataset 0..n]
Optional	dct:source < https://opendata.schleswig-holstein.de/dataset/a4c09d4b-9922-40f2-8615-4f4d89ff339f >;

Resources:

- [TWG Sandbox \(Completeness +/-16%\)](#)
- [HDRUK Gateway \(-\)](#)
- [HealthInformationPortal.eu \(-\)](#)
- [Data.europa.eu SPARQL query \(0.2%\)](#)

Metadata PURI identifiers MUST to be used with dct:hasVersion, dct:isVersionOf and dct:source.

Would you make the use of **dct:source** as

- mandatory
- recommended
- or keep it as is (optional)?

Provide any comments, suggestions or questions:

SPATIAL RESOLUTION

This property refers to the minimum spatial separation resolvable in a dataset, measured in meters.

Example:

	[dcat:spatialResolutionInMeters rdfs:Literal typed as xsd:decimal 0..1]
Optional	dcat:spatialResolutionInMeters "10"^^<http://www.w3.org/2001/XMLSchema#decimal>;

Resources:

- [TWG Sandbox \(Completeness +/-25%\)](#)
- [HDRUK Gateway \(-\)](#)
- [HealthInformationPortal.eu \(88%\)](#)
- [Data.europa.eu SPARQL query \(0.1%\)](#)

Would you make the use of **dcat:spatialResolutionInMeters** as

- mandatory
 recommended
 or keep it as is (optional)?

Provide any comments, suggestions or questions:

TEMPORAL RESOLUTION

This property refers to the minimum time period resolvable in the dataset.

Example:

	[dcat:temporalResolution rdfs:Literal typed as xsd:duration 0..1]
Optional	<pre>dcat:temporalResolution "P1D"^^<http://www.w3.org/2001/XMLSchema#duration>; dcat:temporalResolution "PT24H"^^<http://www.w3.org/2001/XMLSchema#duration>; dcat:temporalResolution "P0.14285W"^^<http://www.w3.org/2001/XMLSchema#duration>;</pre>

Resources:

- [TWG Sandbox \(Completeness +/-14%\)](#)
- [HDRUK Gateway \(-\)](#)
- [HealthInformationPortal.eu \(-\)](#)
- [Data.europa.eu SPARQL query \(0%\)](#)

Would you make the use of **dcat:temporalResolution** as

- mandatory
- recommended
- or keep it as is (optional)?

Provide any comments, suggestions or questions:

TYPE

This property refers to the type of the Dataset. A recommended controlled vocabulary data-type is foreseen.

Example:

	[dct:type skos:Concept 0..n]
Optional	<pre>dct:type <http://publications.europa.eu/resource/authority/dataset-type/SYNTHETIC_DATA>;</pre>

Resources:

- [TWG Sandbox \(Completeness +/-14%\)](#)
- [HDRUK Gateway \(-\)](#)
- [HealthInformationPortal.eu \(-\)](#)
- [Data.europa.eu SPARQL query \(14%\)](#)

The EU controlled vocabulary "[Dataset type](#)" COULD be used. (A new dataset type "personal data" could be required)

Would you make the use of **dct:type** as

- mandatory
- recommended
- or keep it as is (optional)?

Provide any comments, suggestions or questions:

VERSION

This property contains a version number or other version designation of the Dataset.

Example:

	<code>[owl:versionInfo rdfs:Literal 0..1]</code>
Optional	<code>owl:versionInfo "Statistical Bulletin 2023";</code>

Resources:

- [TWG Sandbox \(Completeness +/-20%\)](#)
- [HDRUK Gateway \(-\)](#)
- [HealthInformationPortal.eu \(-\)](#)
- [Data.europa.eu SPARQL query \(3%\)](#)

Would you make the use of **owl:versionInfo** as

- mandatory
- recommended
- or keep it as is (optional)?

Provide any comments, suggestions or questions:

VERSION NOTES

This property contains a description of the differences between this version and a previous version of the Dataset. This property can be repeated for parallel language versions of the version notes.

Example:

	[adms:versionNotes rdfs:Literal 0..n]
O p t i o n a l	adms:versionNotes "XML file v2.0 has been available since 2020-12-09. All information regarding the shortage is displayed for each medicinal package separately in a structured manner, for example dates relevant for the shortage, contact information and possible alternative treatment options."@en;

Resources:

- [TWG Sandbox \(Completeness +/-18%\)](#)
- [HDRUK Gateway \(-\)](#)
- [HealthInformationPortal.eu \(-\)](#)
- [Data.europa.eu SPARQL query \(0.06%\)](#)

Would you make the use of **adsm:versionNotes** as

- mandatory
- recommended
- or keep it as is (optional)?

Provide any comments, suggestions or questions:

WAS GENERATED BY

This property refers to an activity that generated, or provides the business context for, the creation of the dataset.

Example:

	[prov:wasGeneratedBy prov:Activity 0..n]
Optional	<pre>prov:wasGeneratedBy [a prov:Activity; dcterms:type <http://dbpedia.org/resource/Observation> ; prov:startedAtTime "2000-11-01"^^xsd:date ; prov:used dap:Parkes-radio-telescope ; prov:wasInformedBy dap:ATNF ; rdfs:label "P366 - Parkes multibeam high-latitude pulsar survey" @en ; rdfs:seeAlso <https://doi.org/10.1111/j.1365-2966.2006.10100.x>];</pre>

Resources:

- [TWG Sandbox \(Completeness +/-23%\)](#)
- [HDRUK Gateway \(77%\)](#)
- [HealthInformationPortal.eu \(-\)](#)
- [Data.europa.eu SPARQL query \(0%\)](#)

The HDRUK catalogue uses a code list: EPR, LIMS, ELECTRONIC SURVEY, PAPER BASE, MACHINE GENERATED, ...

Would you make the use of **prov:wasGeneratedBy** as

- mandatory
- recommended
- or keep it as is (optional)?

Provide any comments, suggestions or questions:



Contact

[Contact Form](#)

Health DCAT extension - TWG n°4bis 13-10-2023

Fields marked with * are mandatory.

DCAT Health extension

Technical Working Group session 13-09-2023

* Your email:

pascal.derycke@sciensano.be

Nota bene: This form remains open until Monday October 23

Introducing new properties

List of properties of the HealthInformationPortal.eu not mapped to DCAT-AP

- associated national node
- alternative title
- **acronym** (completeness 52%)
- datasource url status
- governance and legal framework
- funding
- sex
- target population
- access information
- **age range from** (completeness 25 % for Personal level data - 10 % for aggregated data)
- **age range to**
- personal identifier
- level of aggregation
- regulations for data sharing

<https://healthinformationportal.eu>

List of properties of the HDRUK Gateway not mapped to DCAT-AP

- updated at

- abstract
- **typicalAgeRange** (completeness 83%)
- physicalSampleAvailability
- followup
- pathway
- timeLag
- dataUseLimitation
- dataUseRequirements
- accessRequestCost
- deliveryLeadTime
- jurisdiction
- dataProcessor
- dataController
- **vocabularyEncodingScheme**
- derivation
- tools
- observations

<https://www.healthdatagateway.org>

Provide any comments or suggestions:

Use case: Annotating a health dataset by its acronym

Example: dct:alternative "LINK-VACC"@en

LINK-VACC stands for "Linking of registers for COVID-19 vaccine surveillance"

Do you agree introducing dct:alternative [An alternative name for the resource] as a new property of the HealthDCAT-AP dataset class?

- yes
 no

Would you make the use of the dct:alternative

- Mandatory
 Recommended
 or Optional?

Provide any comments or suggestions:

Use case: Describing the population age groups

Example: schema:typicalAgeRange "0-4"

Resource: <https://schema.org/typicalAgeRange>

Do you agree introducing schema:typicalAgeRange [The typical expected age range, e.g. '7-9', '11-'.] as a new property of the HealthDCAT-AP dataset class?

- yes
 no

Would you make the use of the schema:typicalAgeRange

- Mandatory
 Recommended
 or Optional?

Provide any comments or suggestions:

Use case: Defining the coding system in use

First option: to be used a literal

Example: schema:codingSystem "ICD-10"

Resource: <https://schema.org/codingSystem>

Property of the schema [MedicalCode type](#)

Example of values:

- MESH
- ICD-10
- LOINC ID
- SNOMED CT

- ...

Second option: to be used Wikidata ontology

Example: schema:codingSystem <<https://wikidata.org/wiki/Property:P672>>

Resource: <https://schema.org/codingSystem>

Property of the schema [Category Code](#)

Example of values:

- MESH Wikidata PURI P672
- ICD-10 Wikidata PURI P494
- LOINC ID Wikidata PURI P4338
- SNOMED CT Wikidata PURI P5806
- ...

Do you agree introducing schema:codingSystem [The coding system, e.g. 'ICD-10'.] as a new property of the HealthDCAT-AP dataset class?

- yes
 no

Do you agree introducing schema:inCodeSet [Wikidata ontology] as a new property of the HealthDCAT-AP dataset class?

- yes
 no

Would you make the use of the schema:codingSystem of schema:inCodeSet

- Mandatory
 Recommended
 or Optional?

Provide any comments or suggestions:

Contact

[Contact Form](#)

TWG session held on the 27-10-2023 - Survey n°5

Topics covered: "provide a subset – proxy data for sensitive health datasets"

- adms:sample

The TWG survey no. 5 explores the use of adms:sample for dataset distributions. It delves into use cases like providing open access subsets, ensuring meaningful use and interpretation of datasets, and how these can be presented as RDF. The survey asks for feedback on making adms:sample mandatory, recommended, or optional, and invites comments or suggestions on RDF-izing variable descriptions using vocabularies like CSVW. Additionally, it discusses interactive query tools, visualisation systems, and the Data Privacy Vocabulary (DPV) for documenting legal bases, personal data, and data processing purposes. Participants are queried on extending DCAP-AP with properties like dpv:hasLegalBasis, dpv:hasPersonalData, and dpv:hasPurpose, and their preferences for making these mandatory, recommended, or optional.

Minutes:

1. Introduction and Overview: Pascal Derycke begins the session by discussing the agenda, emphasizing the importance of understanding variables in health data sets and the use of open data standards.
2. Demonstrations and Examples: Pascal presents several demonstrations to illustrate the concepts being discussed. These include examples of accessing and visualizing data sets, handling metadata, and utilizing open data portals.
3. Discussions on Metadata and Data Access: The session dives deep into the technicalities of metadata records, data service classes, and the intricacies of accessing health data. The importance of having a unified approach to data access and the challenges of dealing with sensitive data are highlighted.
4. Interactive Q&A and Problem-Solving: Throughout the session, there's an interactive Q&A segment where participants, including Truls Korsgaard and others, engage in discussions, providing insights and raising questions about different aspects of health data management.
5. Exploration of Technical Solutions: The session explores various technical solutions for managing health data, including the use of APIs, data abstraction layers, and different data distribution methods. The use of standards like CSV on the Web and other serialization methods for data interoperability is discussed.
6. Feedback and Collaborative Decision-Making: The participants are encouraged to provide feedback on different proposals and use cases. The session aims to reach a consensus on the best practices and standards for health data management.
7. Closing Remarks and Next Steps: The session concludes with a call for further collaboration and the importance of integrating feedback into the development of health data management systems.

Results of the survey:

TGW survey n°5	yes	no	M	R	O	total	Comments
adms:sample			0	2	3	5	
csvw	2	1				4	
External link	3	2				5	
dpv:hasLegalBasis	5	0	1	4	0	5	
dpv:hasPersonalData	3	2	3	1	0	5	
dpv:hasPurpose	3	2	1	2	1	5	

Survey n°5 is available at [this link](#)

Contribution ID: 7905e80e-3c2a-4874-9b93-dff702023d77

Date: 24/01/2024 16:33:23

Health DCAT extension - TWG n°5 27-10-2023

Fields marked with * are mandatory.

DCAT Health extension

Technical Working Group session 27-10-2023

* Your email:

pascal.derycke@sciensano.be

Nota bene: This form remains open until Monday November 13

SAMPLE

This property refers to a sample distribution of the dataset.

Use case: Providing open access subsets to help users to comprehend the dataset's structure and content to ensure meaningful use and interpretation

Example:

	[adms:sample dcat:Distribution 0..n]
Optional	<pre>@prefix dcat: <http://www.w3.org/ns/dcat#> . @prefix csvw: <http://www.w3.org/ns/csvw#> . <http://example.org/dataset/my-rdf-dataset> a dcat:Dataset ; adms:sample [a dcat:Distribution ; dcat:downloadURL <https://www.healthinformationportal.eu/health- information-sources/my-life-survey>; dcat:mediaType "application/rdf+xml" ;] .</pre>

Examples of subsets:

- anonymised data (i.e.:samples)

- synthetic data
- mock-up data
- data profile
- dedicated APIs / Web services
- ...

Providing open access subsets is crucial for users to understand the structure and nuances of a larger dataset without running into privacy issues at the discovery stage.

Resources:

[DCAT-AP](#) [adms:sample dcat:Distribution 0..n]

[Data.europa.eu SPARQL query](#)

Would you make the use of **adms:sample** as

- mandatoy
- recommended
- or optional?

Provide any comments, suggestions or questions:

Sample option: RDF-izing variable's descriptions

Describing dataset variables can be efficiently handled by transforming the information as RDF (see below 3 examples). This linked data presented as RDF can then be described in a DCAT distribution. This will facilitate a robust, machine-readable representation of dataset variables.

Use case: discovering the description of variables as a machine-readable information.

Examples of RDF vocabularies allowing the description of the dataset's variables:

	[adms:sample dcat:Distribution 0..n]
--	--------------------------------------

<p>csvw example</p>	<pre> <https://www.healthinformationportal.eu/health-information-sources/my-life-survey> a csvw:TableGroup ; csvw:name "MY LIFE Survey" ; csvw:description "... " ; csvw:url <https://www.healthinformationportal.eu/health-information-sources/my-life-survey> ; csvw:tableSchema [csvw:columns [{ csvw:name "var1" ; csvw:datatype "string" ; csvw:title "var1 description" ; csvw:unit " The unit of measurement given using a standard unit code or a URL " ; csvw:propertyUrl <a concept ex: http://snomed.info/id/364075005>; rdfs:comment "... " }; { csvw:name "var2" ; csvw:datatype "string" ; csvw:title "Var2 description." ; csvw:unit " A string or text indicating the unit of measurement. Useful if you cannot provide a standard unit code for unitCode. " ; } # ... Add more columns as needed]]; </pre>
<p>DataCube example</p>	<pre> qb:structure [a qb:DataStructureDefinition ; qb:component [qb:dimension [a qb:DimensionProperty; rdfs:label "var1"; rdfs:range <url of a concept>]]; qb:component [qb:dimension ...] # ... Add more dimensions as needed] . </pre>

ssn /sosa example	<pre> <https://www.healthinformationportal.eu/health-information-sources /my-life-survey> a sosa:Sample ; sosa:hasSimpleResult "MY LIFE Survey" ; rdfs:comment "..."; sosa:isSampleOf [a sosa:FeatureOfInterest ; sosa:hasProperty [a ssn:Property ; rdfs:label "var1" ; ssn:hasMeasurementCapability [a ssn:MeasurementCapability ; ssn:hasMeasurementProperty [a ssn:MeasurementProperty ; rdfs:comment "var1 description" ; ssn:hasFeatureOfInterest [a ssn:FeatureOfInterest ; rdfs:label " The unit of measurement ..." ;]]<]], [a ssn:Property ; rdfs:label "var2" ; ssn:hasMeasurementCapability [a ssn:MeasurementCapability ; ssn:hasMeasurementProperty [a ssn:MeasurementProperty ; rdfs:comment "Var2 description." ; ssn:hasFeatureOfInterest [a ssn:FeatureOfInterest ; rdfs:label " A string or text indicating the unit of measurement. ..." ;]]]] # ... Add more properties as needed] ; </pre>
-	...

Resources:

[CSVW Namespace Vocabulary Terms](#)

[The RDF Data Cube Vocabulary](#)

[Semantic Sensor Network Ontology](#)

Would you consider RDF-izing the variables' descriptions using, for instance, the csvw vocabulary?

Yes

No

What would be your preference?

Would you consider to provide an external link to access the variables' descriptions in an alternative format /solution ?

- Yes
 No

Describe your solution:

Provide any comments, suggestions or questions:

Interactive query tools, visualisations systems or Dashboards:

Are you aware of any services that allow users to make specific queries on the dataset without accessing personal level information? or any graphical representations of data that help users comprehend the dataset through visual means ; dashboards with various interactive controls allowing users to explore different facets or dimensions of a data without compromising its sensitivity?

Provide any examples:

Data Privacy Vocabulary (DPV)

The Data Privacy Vocabulary [DPV] enables expressing machine-readable metadata about the use and processing of personal data based on legislative requirements such as the General Data Protection Regulation [GDPR].

The namespace for DPV terms is <https://w3id.org/dpv#>, the suggested prefix is dpv.

dpv:hasLegalBasis

Range: dpv:LegalBasis

Description: The Legal basis used to justify processing of personal data

Note: Legal basis (plural: legal bases) are defined by legislations and regulations, whose applicability is usually restricted to specific jurisdictions.

Use case: documenting the legal ground for processing personal level data

Ref: <https://w3c.github.io/dpv/dpv/#LegalBasis>

	dpv:LegalBasis
-	<pre>@prefix dcat: <http://www.w3.org/ns/dcat#> . @prefix dpv: <https://w3id.org/dpv#> . <http://example.org/dataset/my-rdf-dataset> a dcat:Dataset ; dpv:hasLegalBasis [a dpv:LegalBasis ; dct:description "The legal basis for processing personal data in this dataset is ... CSI deliberation of April 2020"@en; dct:source <https://www.ehealth.fgov.be/ehealthplatform/file/view /AYgJ6X4OuwVJMAnC0FpM?filename=23-086-f170-ICARE4OLD.pdf> ;] .</pre>

Do you agree extending DCAP-AP with the property **dpv:hasLegalBasis**

- Yes
- No

Would you make the use of the **dpv:hasLegalBasis**

- Mandatory
- Recommended
- or Optional?

Provide your comments about the use of "dpv:LegalBasis":

dpv:hasPersonalData

Range: dpv:PersonalData

Description: Data directly or indirectly associated or related to an individual.

Use case: documenting the linkage possibilities for the dataset

Ref: <https://w3c.github.io/dpv/dpv/#PersonalData>

	[dpv:hasPersonalData dpv:PersonalData x..x]
-	<pre>@prefix dcat: <http://www.w3.org/ns/dcat#> . @prefix dpv: <https://w3id.org/dpv#> . <http://example.org/dataset/my-dataset> a dcat:Dataset ; dpv:hasPersonalData dpv-pd:Name, dpv-pd:CurrentEmployer, dpv-pd: PhysicalAddress, dpv-pd:OfficialID, dpv-pd:Salary, dpv-pd:Gender, dpv- pd:BirthDate, dpv-pd:Nationality, dpv-pd:TelephoneNumber, dpv-pd: PaymentCard, dpv-pd:PurchaseHistory, dpv-pd:Transaction, dpv-pd: CreditWorthiness ;</pre>

DPV-PD: Extended Personal Data categories for DPV

DPV-PD extends the Data Privacy Vocabulary (DPV) Specification to provide additional concepts regarding Personal Data categories. The namespace for DPV terms is <https://w3id.org/dpv/dpv-pd#>, the suggested prefix is dpv-pd.

Ref: <https://w3c.github.io/dpv/dpv-pd>

Do you agree extending DCAP-AP with the property **dpv:hasPersonalData**

- Yes
- No

Would you make the use of the **dpv:hasPersonalData**

- Mandatory
- Recommended
- or Optional?

Provide your comments about the use of "dpv:hasPersonalData":

dpv:hasPurpose

Range: dpv:purpose

Description: The purpose of processing personal data

Use case: defining and documenting the purposes of the data processing activity ; understanding the original scope of the data collection, getting the grasp of purpose limitation principles.

Ref: <https://w3c.github.io/dpv/dpv/#Purpose>

	[dpv:hasPurpose dpv:Purpose x..x]
-	<pre>@prefix dcat: <http://www.w3.org/ns/dcat#> . @prefix dpv: <https://w3id.org/dpv#> . <http://example.org/dataset/my-dataset> a dcat:Dataset ; dpv:hasPurpose [a dpv:Purpose ; dct:description "..."@en; dct:source <...> ;] .</pre>

Do you agree extending DCAP-AP with the property **dpv:hasPurpose**

- Yes
- No

Would you make the use of the **dpv:hasPurpose**

- Mandatory
- Recommended
- or Optional?

Provide your comments about the use of "dpv:hasPurpose":

Contact

[Contact Form](#)

TWG session held on the 17-11-2023 - Survey n°6

Topics covered: "providing dataset metrics"

- dct:mediator
- dct:valid
- ex:datametrics

The TWG survey no. 6 discusses various aspects of the Health DCAT extension. Key topics include the introduction of healthDCAT:HDAB to indicate agents responsible for dataset access, and the utilisation of properties like dct:valid for defining dataset time validity. The session provides a comprehensive look at the challenges and solutions in managing health data, focusing on the technical, regulatory, and practical aspects of establishing and operating health data access bodies.

It also explores the concept of dataset metrics, presenting use cases and querying examples. The survey seeks feedback on approaches to providing contact points for Health Data Access Bodies (HDAB), the extension of DCAT-AP with new properties for dataset metrics, and the use of specific properties like dcat:qualifiedRelation and dct:references for this purpose. Participants are asked to comment on the necessity and implementation of these features.

Minutes:

1. Introduction and recap: Pascal Derycke opens the session by summarising previous discussions and outlining the agenda, which includes discussing the nature of health data access bodies, metrics, and data retention periods.
2. Data Access and Regulation: The session delves into the intricacies of health data access, particularly regarding regulations and member states' obligations to establish health data access bodies. These discussions involve understanding the roles and responsibilities of these bodies in managing sensitive health data.
3. Technical Challenges and Solutions: Various technical challenges are discussed, such as implementing metadata records, handling sensitive data sets, and the use of APIs for data access. The session explores different approaches to documenting data access and retention policies.
4. Data Service Classes and Metadata: The conversation shifts to the use of data service classes and the importance of metadata in health data management. The participants discuss the need for unified ways of interacting with data services, especially in the context of sensitive data.
5. Q&A and Open Discussions: Throughout the session, there's an interactive Q&A segment where participants ask questions and provide insights into different aspects of health data management, including technical, regulatory, and operational perspectives.
6. Closing Remarks and Future Directions: The session concludes with discussions on best practices and future strategies for managing health data, with an emphasis on compliance with regulations and efficient data access mechanisms.

Results of the survey:

TGW survey n°6	New property	dct:qualifiedAttribution	healthDCAT custodian,..., PI	GeoDCAT	Other	total	Comments
HDAB	1	1		2		4	Many valuable

							comments
	yes	no	M	R	O	total	
dct:valid	4		1	1	1	4	
	essential	Nice to have	Not needed	New properties	dcat:qualifiedRelation	dct:references	
Dataset metrics		3			2	1	

Survey n°6 is available at [this link](#)

Health DCAT extension - TWG n°6 17-11-2023

Fields marked with * are mandatory.

DCAT Health extension

Technical Working Group session 17-11-2023

* Your email:

pascal.derycke@sciensano.be

Nota bene: This form remains open until Monday November 27

healthDCAT:HDAB

An agent that mediates access to the resource.

Use case: referring to an entity (HDAB) responsible for providing access to the Dataset.

"Setting up health data access bodies will ensure a predictable and simplified access to electronic health data, and a higher level of transparency, accountability and security in data processing"

Proposal for a

REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the European Health Data Space:

"On secondary use of electronic health data, researchers, innovators, policy makers and regulators would be able to have access to quality data for their work in a secure way, with a trusted governance and at lower costs than relying on consent. The common framework for secondary use would reduce the fragmentation and barriers for cross-border accesses. The preferred option requires Member States to set up one or more health data access bodies (with a coordination body), that can provide access to electronic health data to third parties, either as a new organisation or part of an existing organisation, building on the Data Governance Act. Parts of the costs will be offset through fees charged by health data access bodies."

	DCAT-AP
--	---------

[dcat: contact Point vcard: Kind 0..n]	This property contains contact information that can be used for sending comments about the Dataset.
[dct: publisher foaf: Agent 0..1]	This property refers to an entity (organisation) responsible for making the Dataset available.
[dct: creator foaf: Agent 0..n]	This property refers to the entity responsible for producing the dataset.
[prov: qualifiedAttribution prov: Attribution 0..n]	This property refers to a link to an Agent having some form of responsibility for the resource.
[dcat: landing Page foaf: Document 0..n]	This property refers to a web page that provides access to the Dataset, its Distributions and/or additional information. It is intended to point to a landing page at the original data provider, not to a page on a site of a third party, such as an aggregator.

	geoDCAT-AP
[geodcat: custodian foaf: Agent]	Party that accepts accountability and responsibility for the data and ensures appropriate care and maintenance of the resource [ISO-19115].

[geodcat: distributor foaf: Agent]	Party who distributes the resource [ISO-19115].
[geodcat: originator foaf: Agent]	Party who created the resource [ISO-19115].
[geodcat: principalInvesti gator foaf: Agent]	Key party responsible for gathering information and conducting research [ISO-19115].
[geodcat: processor foaf: Agent]	Party who has processed the data in a manner such that the resource has been modified [ISO-19115].
[geodcat: resourceProvid er foaf:Agent]	Party that supplies the resource [ISO-19115].
[geodcat:user foaf:Agent]	Party who uses the resource [ISO-19115].

Example N°1:

	[geodcat:custodian foaf:agent 0..n]
Optional	<pre><http://data.europa.eu/930/custodian> [a foaf:Organization; locn:address [a locn:Address; locn:adminUnitL1 "Polska"; locn:postCode "83-240"; locn:postName "Lubichowo"; locn:thoroughfare "ul. Zblewska 8"]; foaf:homepage <http://www.lubichowo.pl>; foaf:mbox <mailto:ug@lubichowo.pl>; foaf:name "Urząd Gminy Lubichowo"; foaf:phone < tel:++2B48585885221>];</pre>

Example N°2:

	[prov:qualifiedAttribution prov:Attribution 0..n]
Optional	<pre> prov:qualifiedAttribution [a prov:Attribution; dcat:hadRole <http://inspire.ec.europa.eu/metadata-codelist /ResponsiblePartyRole/custodian>; prov:agent [a foaf:Organization; locn:address [a locn:Address; locn:adminUnitL1 "Denmark"; locn:adminUnitL2 "K"; locn:postCode "1050"; locn:postName "Copenhagen"; locn:thoroughfare "Kongens Nytorv 6"]; foaf:mbox <mailto:info@eea.europa.eu>; foaf:name "European Environment Agency"]]; </pre>

Resources:

[Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the European Health Data Space COM/2022/197 final](#)

[Responsible party role - INSPIRE registry \(europa.eu\)](#)

SPARQL query "[geodcat:custodian](#)"

SPARQL query "[prov:qualifiedAttribution](#)"

INSPIRE "ResponsiblePartyRole" Code list:

Label	▲ Governance level
Author	Legal (EU)
Custodian	Legal (EU)
Distributor	Legal (EU)
Originator	Legal (EU)
Owner	Legal (EU)
Point of Contact	Legal (EU)
Principal Investigator	Legal (EU)
Processor	Legal (EU)
Publisher	Legal (EU)
Resource Provider	Legal (EU)
User	Legal (EU)

Which of the following option for providing the contact point of the HDAB, would you choose?

- create a standalone property [healthDCAT:HDAB foaf:agent 1:1]
- use dct:qualifiedAttribution with an associated code list
- according to the business logic of geoDCAT, create the following healthDCAT properties custodian, distributor, originator, principal investigator, ...
- reuse the geoDCAT properties, custodian, distributor, ...
- other solution

Provide any comments, suggestions or questions:

dct:valid

Use case: defining time validity of the dataset

Resource:

[Dublin Core](#)

Do you agree extending DCAP-AP with the property **dct:valid**

- Yes
- No

Would you make the use of **dct:valid**

- mandatory
- recommended
- or optional?

Provide your comments about the use of "dct:valid":

healthDCAT:datasetMetrics

Use case: providing dataset metrics

Data on the Web Best Practices

Best practice 32: Provide complementary presentations

Best Practice 32: Provide Complementary Presentations

Enrich data by presenting it in complementary, immediately informative ways, such as visualizations, tables, Web applications, or summaries.

Why

Data published online is meant to inform others about its subject. But only posting datasets for download or [API](#) access puts the burden on consumers to interpret it. The Web offers unparalleled opportunities for presenting data in ways that let users learn and explore without having to create their own tools.

Intended Outcome

Complementary data presentations will enable human consumers to have immediate insight into the data by presenting it in ways that are readily understood.

Possible Approaches to Implementation

One very simple way to provide immediate insight is to publish an analytical summary in an HTML page. Including summative data in graphs or tables can help users scan the summary and quickly understand the meaning of the data.

If you have the means to create interactive visualizations or Web applications that use the data, you can give consumers of your data greater ability to understand it and discover patterns in it. These approaches also demonstrate its suitability for processing and encourage reuse.

EXAMPLE 32

The MyCity transit agency publishes detailed data about all its transit lines through an [API](#), but it also has many users who are not Web developers and who want to know how to use the system to move about the city. The transit agency could build a Web application that allows users to enter a departure address and a destination and receive step-by-step directions for making their journey via public transit.

How to Test

Check that the dataset is accompanied by some additional interpretive content that can be perceived without downloading the data or invoking an [API](#).

Evidence

Relevant requirements: [R-DataEnrichment](#)

Benefits



Examples:

StatDCAT:

```
<http://data.europa.eu/m8g/numSeries> a rdf:Property ; rdfs:label "number of data series"@en ;  
vann:usageNote "Additional optional property. Cardinality [0..n]. This property contains the number of data
```

series contained in the Dataset. This property should be defined as `rdfs:Literal` typed as `xsd:date` or `xsd:dateTime`. "@en ; rdfs:range rdfs:Literal .

HealthInformationPortal.eu:

- Age range
- Sex


Other examples:

- Number of records
- Number of unique individuals
- ...

Augmented data analytics

Providing metrics and insights thanks to an data abstraction layer: "... we don't have to think about how we're going to use the data before we access it. It's just there for us to query..."

Example: Beacon API



The Beacon project started by the Global Alliance for Genomics and Health (GA4GH) in 2014, initially to test the willingness of international data owners to make available the smallest amount of information. A Beacon allows an anonymous query as to whether the dataset contains a genome with a given base at a particular coordinate, with a simple yes/no response. The Beacon Project had processed more than 1.5 million queries against at least 100,000 genomes held by more than 25 different organisations.

Beacon

Example of a query: Have you seen deletions in this region of Chr.9 in Glioblastomas from a juvenile patient, in a dataset with unrestricted access?

Other use cases

Utility indicators:

- No poverty
- Zero hunger
- Good health and well-being
- ...

Quality indicators:

- % of completeness

• ...

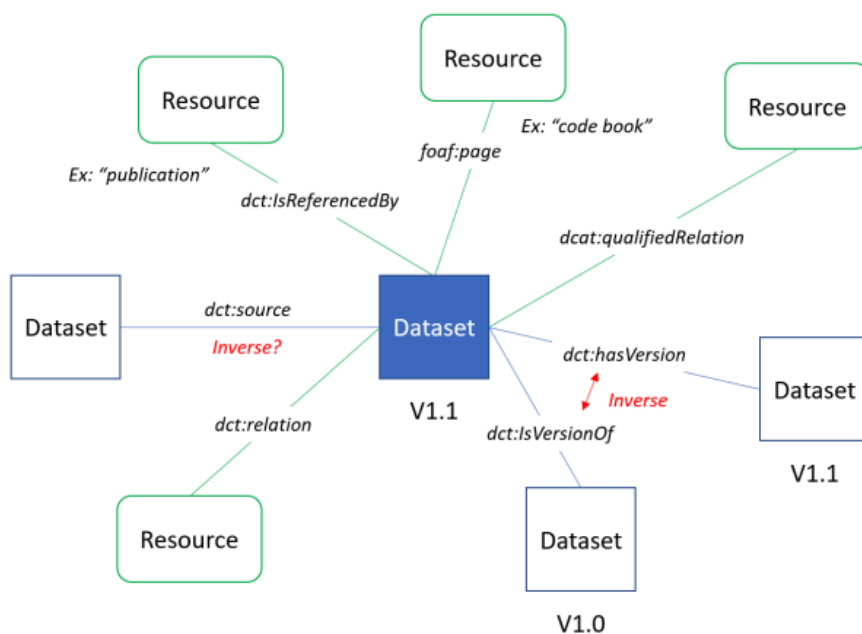
Resources:

- [Data on the Web Best Practices](#)
- [Beacon API](#)
- [Sustainable development indicators](#)
- [IANA Link relation types](#)
- [Create dataset metrics](#)

Possible approaches to implementation:

1. Create custom properties: Extend the DCAT vocabulary by creating new custom properties for dataset metrics.
2. Use of `dcat:qualifiedRelation` specifying that the related resource is a metrics dataset.
3. Linking with external metric platforms. Use of `dct:references` (inverse property of `dct:isReferencedBy`)
4. ...

DCAT-AP metadata properties describing a relation to another dataset or a resource



dct:source

This property refers to a related Dataset from which the described Dataset is derived.

dcat:qualifiedRelation

This property provides a link to a description of a relationship with another resource.

dct:relation

This property refers to a related resource.

dct:isReferencedBy

A related resource that references, cites, or otherwise points to the described resource.

dct:references

A related resource that is referenced, cited, or otherwise pointed to by the described resource.

Do you consider that providing dataset metrics are:

- essential
- nice to have
- not needed?

Which of the following option for providing dataset metrics, would you choose?

- extend DCAT-AP with a list of new properties [healthDCAT:metric1,2,3...]
- use dcat:qualifiedRelation with an associated code list
- use dct:references
- other solution.

Provide any comments, suggestions or questions:

Contact

[Contact Form](#)

TWG session held on the 24-11-2023 - Survey n°7

Topics covered: "metadata fit for AI"

- SeTA presentation by the JRC-Ispra
- dct:provenance, ex:populationtype, dpv :purpose

The session from November 24th highlighted a Natural Language Processing AI tool for health data set searches. Vidas Daudaravicius from the European Commission's Joint Research Centre in Ispra, Italy introduced the 'SeTA Knowledge Explorer' tool, developed by JRC, focusing on semantic text analysis. The session emphasized the tool's applications in data processing, policy research, and health datasets, showcasing its capabilities in enhancing data findability.

The TWG survey no. 7 explores the metadata properties conducive to NLP/AI applications, like dct:description and dct:provenance, and considers their mandatory or optional status. The survey also discusses the dpv:hasPurpose property to document the purpose of personal data processing. Additionally, it covers health:populationCoverage for dataset population types and seeks feedback on extending DCAP-AP with these properties. The survey includes detailed use cases and invites specific comments on their implementation.

Minutes:

1. Introduction: Pascal Derycke opens the session, introducing the topic of knowledge exploration tools, particularly in the context of natural language processing AI and technical working groups.
2. Presentation on Semantic Text Analysis Tool: Vidas Daudaravicius presents a tool, emphasizing its capabilities in semantic text analysis. He discusses the challenges of handling large volumes of data in documents, scientific papers, websites, and social media. The tool's ability to process both structured and unstructured data using advanced technologies is highlighted.
3. Tool's Functionality and Use Cases: The session delves into the tool's functionality, including its full-text search capabilities, narrowing down searches to specific text parts, and concept search. The tool's utility in searching for specific keywords and phrases like "sustainable Europe" is demonstrated.
4. Technical Discussion: The discussion shifts to more technical aspects, including the use of neural networks, metadata, and augmented analytics. The tool's approach to document indexing and search, including the use of hybrid search approaches, is explained.
5. Live Demo and Interaction: A live demonstration of the tool is conducted, showing its practical application. This includes showcasing the tool's ability to enrich queries with related keywords automatically and its capability to stage and export data for further analysis.
6. Questions and Answers Session: The floor is opened for questions, leading to discussions on various topics like API integration, the tool's language capabilities, and its application in different contexts, including health data sets and policy research.
7. Closing Remarks: The session concludes with further clarifications and discussions about the tool's potential applications, integration with other systems, and future developments.

Results of the survey:

TGW 7 (AI UC)	yes	no	M	R	O	total	Comments
dct:provenance				5		5	
dpv:purpose			2	3		5	
health:populationCoverage	5		2	3		5	
	UC1	UC2	UC3	UC4	UC5		
Search simulation	2	2	2		2		

Survey n°7 is available at [this link](#)

Health DCAT extension - TWG n°7 24-11-2023

Fields marked with * are mandatory.

DCAT Health extension

Technical Working Group session 24-11-2023

* Your email:

pascal.derycke@sciensano.be

Nota bene: This form remains open until Monday December 4

Metadata fit for the purpose of NLP/AI

JRC report "Semantic Text Analysis tool: SeTA"

"5.3.3 Results and preliminary conclusions"

Use case: "It was proven that the primary source of information about each of the data set was the abstract that can be mined with ML techniques to often achieve more consistent results compared to human classification, e.g INSPIRE themes keyword."

Resource:

[Semantic Text Analysis tool: SeTA](#)

Examples of metadata properties which can be mined:

dct:description

Range: rdfs:Literal

Description: This property contains a free-text account of the Dataset. This property can be repeated for parallel language versions of the description.

Ref: [DCAT official Application Profile \(DCAT-AP 2.1.1\)](#)

[dct:description rdfs:Literal 1..n]

M a n d a t o r y	dct:description "The LINK-VACC project, organized by Sciensano, involves the communication of pseudonymized data related to COVID-19 vaccines. The data includes demographic information (age, gender, postal code, date of death) of individuals who have received at least one dose of a COVID-19 vaccine in Belgium, details about the vaccinator, vaccination location, administered vaccine, and observed side effects. Additionally, the project encompasses data from the HealthData COVID-19 database and Clinic database, which provide demographic and clinical details of individuals tested for or hospitalized with COVID-19, as well as information from the Common Base Registry for HealthCare Actor (CoBRHA), STATBEL, and the Agency for Health Insurance Fund (AIM) database"@en;
---	--

dct:provenance

Range: dct:ProvenanceStatement

Description: This property contains a statement about the lineage of a Dataset.

Ref: [DCAT official Application Profile \(DCAT-AP 2.1.1\)](#)

	[dct:provenance dct:ProvenanceStatement 0..n]
O p t i o n a l	dct:provenance "The data for the LINK-VACC project is sourced from several existing databases, including Vaccinnet+, HealthData COVID-19 database (Contact tracing and Clinic database), CoBRHA, STATBEL, and the AIM database. These databases collectively provide comprehensive demographic, clinical, and socio-economic data relevant to the project's objectives "@en;

Considering AI requirements, would you make the use of **dct:provenance** as

- mandatoy
- recommended
- or optional?

Provide any comments, suggestions or questions:

dpv:hasPurpose

Range: dpv:purpose

Description: The purpose of processing personal data

Use case: defining and documenting the purposes of the data processing activity ; understanding the original scope of the data collection, getting the grasp of limitation principles.

Ref: <https://w3c.github.io/dpv/dpv/#purpose>

	[dct:hasPurpose]
-	<code>dpv:purpose "The primary objective of Sciensano's LINK-VACC project is to monitor COVID-19 vaccines post-authorization and evaluate the public health value of prioritizing vaccination for people with comorbidities. This involves assessing the vaccines' effectiveness and safety in the broader population context, beyond the limited scope of clinical trials, and determining future vaccination policies in public health emergencies such as epidemics or pandemics"@en;</code>

Considering AI requirements, would you make the use of **dct:purpose** as

- mandatoy
- recommended
- or optional?

Provide any comments, suggestions or questions:

health:populationCoverage

Range: rdfs:Literal

Description: The type of population common to all subjects of the dataset

Use case: Indicating the type of population common to all subjects/records of the dataset. For all associated subjects, it can be characterised by age-specific population, gender-specific population, race, disease, life factors etc.

Resources:

[Schema:epidemiology](#) :

The characteristics of associated patients, such as age, gender, race etc.

[Schema:PopulationType](#):Indicates the populationType common to all members of a StatisticalPopulation or all cases within the scope of a StatisticalVariable.

[Schema:StatisticalPopulation](#): A StatisticalPopulation is a set of instances of a certain given type that

satisfy some set of constraints. The property `populationType` is used to specify the type. Any property that can be used on instances of that type can appear on the statistical population. For example, a `StatisticalPopulation` representing all Persons with a `homeLocation` of East Podunk California would be described by applying the appropriate `homeLocation` and `populationType` properties to a `StatisticalPopulation` item that stands for that set of people. The properties `numConstraints` and `constraintProperty` are used to specify which of the populations properties are used to specify the population. Note that the sense of "population" used here is the general sense of a statistical population, and does not imply that the population consists of people.

	[health:populationCoverage rdfs:Literal]
-	health:populationCoverage "The population targeted by the LINK-VACC project comprises all individuals in Belgium who have received a COVID-19 vaccine, undergone testing for COVID-19, or have been hospitalized with a confirmed diagnosis of COVID-19. The project also considers healthcare professionals and the general Belgian population for understanding vaccination coverage and effectiveness, especially among those with comorbidities and varying socio-economic backgrounds "@en;

Do you agree extending DCAP-AP with a property **health:populationCoverage**?

- Yes
- No

Considering AI requirements, would you make the use of **health:populationCoverage** as

- mandatoy
- recommended
- or optional?

How would you search for health datasets?

The pilot project `healthData@EU` is conducting 5 Use Cases, and you are responsible for one of them. To understand your approach, we're interested in how you would search for the necessary data and your process in doing so.

Please explore the <https://data.europa.eu> metadata catalogue, which contains about 19,000 health-related records, and share your experience:

- What keywords or phrases did you use in your search?
- Which links did you choose to explore, and why did you choose them?
- Identify the datasets and their sources that seem suitable for your Use Case.
- Also, mention any datasets or sources you expected to find but didn't. Any other information that might be relevant to your search for data.

Your insights will help us understand your approach and how to improve the usability of a metadata catalogue for a health project

U C t i t l e	Surveillance of antimicrobial resistance
O b j e c t i v e s	The primary objective of the Use case is to feasibility, for ECDC, to conduct its surveillance of antimicrobial resistance (AMR) mission through the EHDS. The team will test the sharing of communicable disease surveillance data from the Member States to ECDC through the solutions developed by the Pilot and will compare performance with systems in place (TESSy). The Use case will also aim to identify and address potential legal issues.
D a t a	Invasive isolates (blood and cerebrospinal fluid) of 8 bacterial pathogens commonly causing infections in humans

Based on the Use Case's objectives, please, describe in detail your dataset quest:

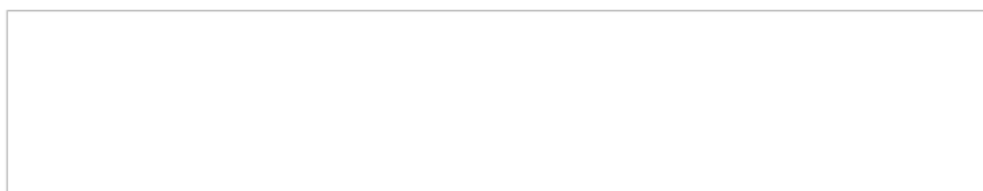
U C t i t l e	Natural history of coagulopathy (blood clotting) related events in COVID-19 patients and risk factors
---------------------------------	---

Objectives	The Use case will aim to address 5 research questions of growing complexity: estimate the incidence of venous and arterial thromboembolic events among 1/the general population; 2/patients with COVID-19; 3/patients with SARS-CoV-2 vaccination ; estimate 4/the impact of clinical risk factors and prior SARS-CoV-2 vaccination on the incidence of venous and arterial thromboembolic events among patients with COVID-19 and worsening of COVID-19, as well as 5/ the incidence rate ratios for such events among patients with COVID-19 during the period when Omicron was the dominant variant and people vaccinated against SARS-CoV-2, compared to background rates as estimated in objectives. 1, 2 and 3.
Data	Coagulopathy related events, Use of thrombotic agents, Risk factors, Vaccination, Covid-19 worsening

Based on the Use Case's objectives, please, describe in detail your dataset quest:

U C t i t l e	Population uptake metrics: COVID-19 test positivity, vaccination and hospitalization
O b j e c t i v e s	This use case aims to measure the uptake of tests, hospitalisation and vaccination in the general population and in vulnerable subpopulations, in order to get a European overview of the situation and to compare between Member States. Vulnerable subpopulations will be defined through the use of socio-economic indicators (income, level of education, migratory background...)
D a t a	Vaccination, Tests, Hospitalisation, Socio-economic information

Based on the Use Case's objectives, please, describe in detail your dataset quest:



U C t i t l e	Comparing Nationwide Health trajectories to evaluate European Health Data interoperability: an application to cardiometabolic diseases
O b j e c t i v e s	This Use case aims to address two questions: Are health trajectories leading to cardiometabolic diseases (i.e. cardiovascular diseases, type 2 diabetes and stroke) comparable across countries? Can we leverage longitudinal disease trajectories to forecast risk for cardiometabolic diseases
D a t a	Diagnosis, Procedures, Drugs, Cause of death.

U C t i t l e	Genomic data linked to health data, with a focus on cancer
O b j e c t i v e s	The Use case will draw on genomic data to interpret complex mutational patterns and medical trajectories of metastatic colorectal cancer patients in the context of associated clinical data. The objective is to confirm known gene signatures and unveil new ones beyond other confounding factors like origin and socio-economics factors as well as confirming new data-driven hypotheses on having different signatures depending on disparate factors like tumour localization or age.
D a t a	Whole Genome Sequencing / Whole Exome sequencing, Gene panels, Phenoclinical data, Socio-economic data

Based on the Use Case's objectives, please, describe in detail your dataset quest:

Contact

[Contact Form](#)

TWG session held on the 08-12-2023 - Survey n°8

Topics covered: "Data classification"

- dpv:dataController, ...
- Resource type, data type, funding type, ...

The TWG survey no. 8 focuses on the implementation of quality metrics in datasets, exploring options such as dataset quality rating and provision of quality metrics. The survey provides examples from StatDCAT-AP and GeoDCAT-AP, detailing how these approaches handle quality annotations and measurements. It also outlines QUANTUM's objectives, a project aimed at developing a data quality and utility label, and its potential implementation in the HealthData@EU infrastructure. The survey also focuses on the importance of control vocabularies in standardising data descriptions and ensuring their machine readability and actionability. The need for effective governance and maintenance of these vocabularies was emphasized. The group considered using existing resources like WikiData as ontologies for medical domains and the potential for creating new control vocabularies based on specific requirements. The survey seeks feedback on extending DCAT-AP with quality-related properties and the use of controlled vocabularies for metadata description.

Minutes:

1. Introduction and Agenda Setting: Pascal Derycke opens the session, outlining the focus on data set quality indicators and control vocabularies. He also mentions the agenda to discuss the survey results related to these topics.
2. Review of Past Sessions and Survey Results: Pascal presents the outcomes from sessions 6 and 7, and discusses the survey results, focusing on health data access, retention periods, data set metrics, and quality indicators.
3. Discussion on Data Set Quality Indicators: The session explores different approaches to expressing data quality in metadata records. Derycke discusses the distinction between rating data set quality and providing quality metrics, along with examples from various sources.
4. Control Vocabularies and Metadata Enrichment: The discussion shifts to control vocabularies and their role in enriching metadata records. Derycke explains the importance of these vocabularies in providing a consistent way to describe data, particularly in the context of health data sets.
5. Technical Insights and Examples: Technical aspects of metadata management are highlighted, including the use of specific vocabularies and examples of how they are implemented in metadata records.
6. Interactive Questions and Clarifications: The session includes a Q&A segment where participants, such as Thinsz Zoltan and Bert Van Nuffelen, ask questions and provide comments, leading to further clarification on topics like metadata quality assessment tools and data set quality metrics.
7. Future Plans and Closing Remarks: The session concludes with Pascal outlining future steps, including the presentation of a draft application profile for health data and plans to test and validate this profile based on pilot use cases.

Results of the survey:

TGW survey n°8	hasQuality Annotation	hasQuality Measurement	Both properties	none	M	R	O	total	Comments
Dataset quality metrics	1		3	1		4	1	5	QUANTUM > hasQualityAnnotation

Milestone M6.2 - Technical working group on the transition from existing metadata templates to HealthDCAT-AP - « Working group minutes »

										New class "Analytics" provides a mean to associate with the dataset, quality measurements Consider EMA quality dimensions as input for QUANTUM.
Review of the proposed controlled vocabularies	Out of the 5 received surveys, the participants are mostly in favour of using the proposed vocabularies and metadata properties. They suggested some code list for art.33, the theme categories and publisher types									Suggestion for Biobanks, to have a "material type" property and an associated controlled vocabulary.
Classification systems	Drugs: CIP, UCD, ATC, SNOMED-CT, RXNORM, meddra diagnosis : ICD-10, ICD-11, SNOMED-CT medical acts: snomed, CCAM, NABM, CSARR, CPT-4 medical device: snomed, LPP lab: LOINC Orphanet MeSH BMO, SMASH, FOBI, HPO, ATC, RxNorm, EphMRA, ALT, DrugBank , ICD-9, ICD-10, ICD, ICD-10-CM, ICD-9-CM, ICPC, ICPC-1, ICPC-2, HCPCS, Read, SNOMED, SNOMED CT, MedDRA, OPCS, CCS, EDC, CPT, dm+d, Orphacode, Orphanet Rare, Disease Ontology (ORDO), Human Phenotype Ontology (HPO), ICD-11, OGG, FG, GO, EGO, SOPHARM, PHARE, HGNC, ART 57, IFA GmbH, EDQM, SPN, MTHSPL, SNOMED, AIC, CNF, DIN, Gemsript, dm+d, Z-index (G-standard), GTIN, NDC, NDF, RxNorm, ATC level 5, WHODrug)									

Survey n°8 is available at [this link](#)

Health DCAT extension - TWG n° 8 08-12-2023

Fields marked with * are mandatory.

DCAT Health extension

Technical Working Group session 08-12-2023

* Your email:

pascal.derycke@sciensano.be

Nota bene: This form remains open until Monday December 18

healthDCAT:hasQualityMetrics/Label

- Option1: Rating the dataset's quality
- Option2: Providing quality metrics

Example of StatDCAP-AP

Rating the dataset's quality

Definition: A statement related to quality of the Dataset, including rating, quality certificate, feedback that can be associated to datasets or distributions.

Comment: The information may include quality aspects such as accuracy, reliability, comparability, coherence, relevance, timeliness etc.

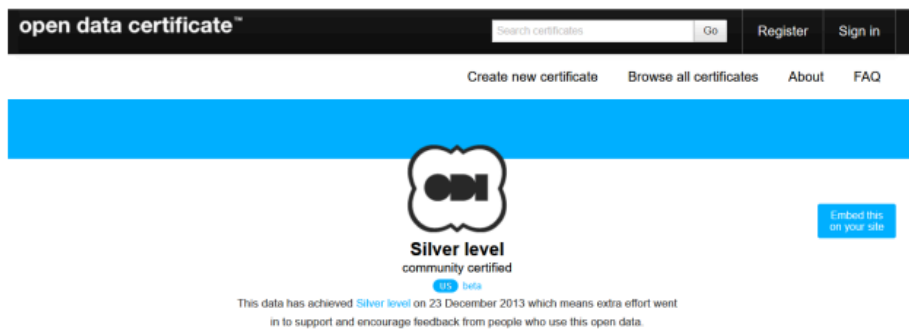
Usage note: The annotation requires the provision of information about the motivation of the annotation (oa:motivation), and an explicit link to the resource being annotated (oa:hasTarget) together with either a link to a resource that contains the annotation (oa:hasBody) or text filed (oa:bodyText).

StatDCAT-AP [dqv:hasQualityAnnotation oa:Annotation 0..n]

Optional	<pre> <https://certificates.theodi.org/en/datasets/393> a dcat:Dataset ; dqv:hasQualityAnnotation :myDatasetQA . : myDatasetQA a dqv:QualityCertificate ; oa:hasTarget <https://certificates.theodi.org/en/datasets/393> ; oa:hasBody <https://certificates.theodi.org/en/datasets/393 /certificate> ; oa:motivatedBy dqv:qualityAssessment </pre>
----------	--

Resource:

<https://certificates.theodi.org/en/datasets/393/certificate>



Example of GeoDCAP-AP

Providing quality metrics

Refers to the performed quality measurements.

In GeoDCAT-AP, this property is used to specify "spatial resolution", as defined in [INSPIRE-MD-REG], [ISO-19115], and [ISO-19115-1].

	GeoDCAT-AP [dqv:hasQualityMeasurement dqv:QualityMeasurement 0..n]
Optional	<pre> dqv:hasQualityMeasurement [a dqv:QualityMeasurement ; dqv:isMeasurementOf <http://data.europa.eu/930 /spatialResolutionAsScale>; dqv:value f0.00002]; </pre>

Geospatial quality metrics:	Health quality metrics (examples)
-----------------------------	-----------------------------------

<ul style="list-style-type: none"> • Spatial resolution as angular distance • Spatial resolution as distance • Spatial resolution as equivalent scale • Spatial resolution as vertical distance <p>Resource: GeoDCAT vocabulary</p>	<ul style="list-style-type: none"> • Accuracy • Completeness • Consistency • Timeliness • Reliability • Granularity • Validity • Accessibility • Interoperability • Ethical Considerations
---	--

The “StatDCAT” OPTION is planned as an implemented solution for the HealthDATA@EU infrastructure

HORIZON-HLTH-2023-TOOL-05 "QUALITY, UTILITY AND MATURITY MEASURED; DEVELOPING A DATA QUALITY AND UTILITY LABEL FOR HEALTHDATA@EU"

About QUANTUM: QUANTUM aims at providing guidance on a quality and utility label, out of the testing of a data quality, utility and maturity labelling mechanism in a substantive number of data holders. Ideally, this labelling mechanism could be adopted in the HealthData@EU as foreseen in Article 56 in the current legislative European Health Data Space for secondary use proposal (still in discussion).

QUANTUM will focus on the following objectives:

1. Conceptualise and develop a data quality and utility label in the context of data holder maturity model;
2. Design, developing and testing the labelling of the data sets' quality and utility, and data holders' maturity;
3. Analyse the implementation challenges with a view to enable the transferability and sustainability of the labelling mechanism as part of HealthData@EU; and,
4. Develop a capacity building and outreach program that allows a broad engagement of the data quality professional community.

[Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the European Health Data Space](#)

Article 56

Data quality and utility label

1. Datasets made available through health data access bodies may have a Union data quality and utility label provided by the data holders.
2. Datasets with electronic health data collected and processed with the support of Union or national public funding shall have a data quality and utility label, in accordance with the principles set out in paragraph 3.
3. The data quality and utility label shall comply with the following elements:
 - (a) for data documentation: meta-data, support documentation, data model, data dictionary, standards used, provenance;
 - (b) technical quality, showing the completeness, uniqueness, accuracy, validity, timeliness and consistency of the data;
 - (c) for data quality management processes: level of maturity of the data quality management processes, including review and audit processes, biases examination;
 - (d) coverage: representation of multi-disciplinary electronic health data, representativity of population sampled, average timeframe in which a natural person appears in a dataset;
 - (e) information on access and provision: time between the collection of the electronic health data and their addition to the dataset, time to provide electronic health data following electronic health data access application approval;
 - (f) information on data enrichments: merging and adding data to an existing dataset, including links with other datasets;
4. The Commission is empowered to adopt delegated acts in accordance with Article 67 to amend the list of principles for data quality and utility label. Such delegated acts may also amend the list set out under paragraph 3 by adding, modifying or removing requirements for data quality and utility label.
5. The Commission shall, by means of implementing acts, set out the visual characteristics and technical specifications of the data quality and utility label, based on the elements referred to in paragraph 3. Those implementing acts shall be adopted in accordance with the advisory procedure referred to in Article 68(2). Those implementing acts shall take into account the requirements in Article 10 of Regulation [...] [AI Act COM/2021/206 final] and any adopted common specifications or harmonised standards supporting those requirements.

Data on the Web Best Practices: [Data Quality Vocabulary](#)

StatDCAT approach:

4.15 Property: Has Quality Annotation

RDF Property:	dqv:hasQualityAnnotation
Definition:	Refers to a quality annotation. Quality annotation can be applied to any kind of resource, e.g., a dataset, a linkset, a graph, a set of triples. However, in the DQV context, this property is generally expected to be used in statements in which subjects are instances of dcat:Dataset or dcat:Distribution .
Range:	dqv:QualityAnnotation

4.8 Class: Quality Certificate

RDF Class:	dqv:QualityCertificate
Definition:	An annotation that associates a resource (especially, a dataset or a distribution) to another resource (for example, a document) that certifies the resource's quality according to a set of quality assessment rules.
Subclass of:	dqv:QualityAnnotation

GeoDCAT approach

4.13 Property: Has Quality Measurement

RDF Property:	dqv:hasQualityMeasurement
Definition:	Refers to the performed quality measurements. Quality measurements can be performed to any kind of resource (e.g., a dataset, a linkset, a graph, a set of triples). However, in the DQV context, this property is generally expected to be used in statements in which subjects are instances of dcat:Dataset or dcat:Distribution .
Range:	dqv:QualityMeasurement
Inverse property:	dqv:computedOn

4.1 Class: Quality Measurement

RDF Class:	dqv:QualityMeasurement
Definition:	Represents the evaluation of a given dataset (or dataset distribution) against a specific quality metric.
Subclass of:	qb:Observation
Equivalent to:	daq:Observation
DQV usage note:	The unit of measure in quality measurement should be specified through the property sdmx-attribute:unitMeasure as recommended by RDF Data Cube [Vocab-Data-Cube]. The Ontology of units of Measure (OM) [RijgersbergETAl] provides a list of HTTP dereferenceable unit of measures, which can be exploited as values for sdmx-attribute:unitMeasure .

The following properties can be used with instances of this class: [dqv:isMeasurementOf](#), [qb:dataSet](#), [dqv:computedOn](#), [dqv:value](#).

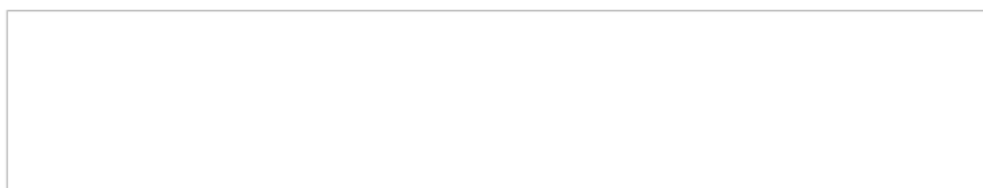
Do you agree extending DCAP-AP with the property

- hasQualityAnnotation
- hasQualityMeasurement
- both properties
- none of these properties

Would you make the use of the **quality property**

- Mandatory
- Recommended
- or Optional?

Provide your comments:



EU (and non EU) controlled vocabularies

<https://op.europa.eu/en/web/eu-vocabularies>

Controlled vocabularies provide a consistent way to describe data. They are standardised and organised arrangements of words and phrases presented as alphabetical lists of terms or as thesauri and taxonomies with a hierarchical structure of broader and narrower terms.

	Controlled vocabularies in DCAT-AP maintained by the Publications Office
dct:spatial	http://publications.europa.eu/resource/authority/country/
dcat:theme	http://publications.europa.eu/resource/authority/data-theme/
dct:accessRights	http://publications.europa.eu/resource/authority/access-right/
dct:accrualPeriodicity	http://publications.europa.eu/resource/authority/frequency/
dct:language	http://publications.europa.eu/resource/authority/language/
dct:type (Dataset)	http://publications.europa.eu/resource/authority/dataset-type/(PERSONAL_DATA)
dct:type (Distribution)	http://publications.europa.eu/resource/authority/file-type/

	Other controlled vocabularies in DCAT-AP
dcat:mediaType	http://www.iana.org/assignments/media-types/text/

SKOS "Simple Knowledge Organization System"

SKOS is a common data model for sharing and linking knowledge organization systems via the Web. It represents, for instance, thesauri within the framework of the Semantic Web. It makes thesauri machine readable and actionable.

Resources:

[Wikipedia](#)

[W3C](#)

Example:

<pre>dct:publisher <https://org.belgif.be/id/CbeRegisteredEntity/0207367489>;</pre>
<pre><https://org.belgif.be/id/CbeRegisteredEntity/0207367489> a foaf:Organization , foaf:Agent; dct:type <http://purl.org/adms/publishertype/LocalAuthority>; foaf:name "Vorst"@nl , "Forest"@fr , "Forest"@en .</pre>
<pre>dct:type <http://purl.org/adms/publishertype/LocalAuthority>;</pre>

```

<owl:NamedIndividual rdf:about="http://purl.org/adms/publishertype/1.0">
<rdf:type rdf:resource="&skos;ConceptScheme"/>
<rdfs:label xml:lang="en">Publisher Type</rdfs:label>
<skos:hasTopConcept rdf:resource="http://purl.org/adms/publishertype
/Academia-ScientificOrganisation"/>
<skos:hasTopConcept rdf:resource="http://purl.org/adms/publishertype
/Company"/>
<skos:hasTopConcept rdf:resource="http://purl.org/adms/publishertype
/IndustryConsortium"/>
<skos:hasTopConcept rdf:resource="http://purl.org/adms/publishertype
/LocalAuthority"/>
<skos:hasTopConcept rdf:resource="http://purl.org/adms/publishertype
/NationalAuthority"/>
<skos:hasTopConcept rdf:resource="http://purl.org/adms/publishertype
/NonGovernmentalOrganisation"/>
<skos:hasTopConcept rdf:resource="http://purl.org/adms/publishertype
/NonProfitOrganisation"/>
<skos:hasTopConcept rdf:resource="http://purl.org/adms/publishertype
/PrivateIndividual(s)"/>
<skos:hasTopConcept rdf:resource="http://purl.org/adms/publishertype
/RegionalAuthority"/>
<skos:hasTopConcept rdf:resource="http://purl.org/adms/publishertype
/StandardisationBody"/>
<skos:hasTopConcept rdf:resource="http://purl.org/adms/publishertype
/SupraNationalAuthority"/>
</owl:NamedIndividual>

```

	Proposed controlled vocabularies in healthDCAT-AP
healthdcatap:healthCategory [TWG session n°2b]	Based on Art.33 (See example of High-value dataset categories)
healthdcatap:healthTheme [TWG session n°3 & 4]	Wikidata (" Wikidata: A large-scale collaborative ontological medical database ") (See example of MobilityDCAT-AP)
healthdcatap:hasCodeValues [TWG session n°4b]	Create a SKOS-based structure to represent relevant parts of the thesauri: LOINC, SNOMED-CT, ICD-10, FASTA,...
healthdcatap:publisherType [TWG session n°2]	Registry, Biobank, Public Health Institute, Hospital, ...

dct:conformsTo [TWG session n°3 & 4]	Wikidata (See: https://medical-data-models.org)
prov:qualifiedAttribution dcat:hadRole [TWG session n°3 & 4]	Ex: http://inspire.ec.europa.eu/metadata-codelist/ResponsiblePartyRole/ (processor)
dcat:qualifiedRelation dcat:hadRole [TWG session n°3 & 4]	Ex: http://www.iana.org/assignments/relation/ (related)

Provide your comments or suggestions on the use of controlled vocabularies:

What medical terminology, classification systems, drug and pharmaceutical vocabularies, etc., do you know or use in your domain? Are they represented with SKOS?

Alignments between controlled vocabularies

Alignments are common in database interoperability projects and tasks. It is in the nature of controlled vocabularies that a given concept from one vocabulary can have a level of correspondence with a concept in a different vocabulary, no matter if these controlled vocabularies comply with the same or different ontologies.

[EuroVoc Alignment WikiData](#)

Alignement example: Eurovoc <-> wikidata

```
<rdf:Description rdf:about="http://eurovoc.europa.eu/5104">
<rdf:type rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#about"/>
<closeMatch xmlns="http://www.w3.org/2004/02/skos/core#" rdf:resource="http://www.wikidata.org/entity/Q12078"/>
</rdf:Description>
```

Contact

[Contact Form](#)

TWG session held on the 22-12-2023 - Survey n°9 & mockups:

Topics covered:

- Draft version of the healthDCAT-AP (LINK-VACC dataset example)
- Mockups of the healthdata.europa.eu catalogue

The last TWG session focused on reviewing the draft Health DCAT Application Profile. The TWG survey N°9 showcases a metadata record for the "LINK-VACC" dataset, detailing how the healthDCAT application profile is applied to represent health dataset metadata, including free text properties, distributions, faceted search properties, and other metadata elements. The TWG survey N°9 includes mockups related to the EU Health Data Portal. Key features include a comprehensive search functionality, a Health Data Access Body Register, various filters like categories, health themes, coding systems, age range, population size, time period, and update frequency. It also covers the Quality and Utility Label and a Personal Data Filter. The survey solicits comments on these features. Participants are invited to provide feedback on these aspects. The survey n°9 will remain open for feedback until the end of the project.

The session included plans for future work (Task 6.3 & 6.4), such as testing and validating the application profile based on real health metadata records, with an aim to refine the design and implementation of the profile. The TWG sessions were a step towards creating a more standardized, efficient, and legally compliant approach to data set management and metadata handling, in the context of health data and the healthData@EU infrastructure envisioned by DG Santé of the European Commission.

Health DCAT extension - TWG n°9 LINK-VACC example 22-12-2023

Fields marked with * are mandatory.

A HealthDCAT-AP Metadata record

LINKING OF REGISTERS FOR COVID-19 VACCINE SURVEILLANCE (LINK-VACC)
<https://www.sciensano.be/en/projects/linking-registers-covid-19-vaccine-surveillance>

The following survey presents a draft version of the healthDCAT-AP taking the example of the LINK-VACC dataset which in use in the Sciensano pilot Use Case. If the RDF code is in *italic*, it means that the code is an example that doesn't apply to the LINK-VACC dataset.

* Your email:

pascal.derycke@sciensano.be

Nota bene: This form will remain open and can be shared in order to collect feedback

DCAT-AP has 2 mandatory properties: "Title" and "Description".

TITLE

This property contains a name given to the Dataset. This property can be repeated for parallel language versions of the name.

	DCAT-AP [dct:title rdfs:Literal 1..n]
Mandatory	dct:title "Linking of registers for COVID-19 vaccine surveillance"@en;

DESCRIPTION

This property contains a free-text account of the Dataset. This property can be repeated for parallel language versions of the description.

	DCAT-AP [dct:description rdfs:Literal 1..n]
M a n d a t o r y	dct:description "The LINK-VACC project, organised by Sciensano, involves the communication of pseudonymized data related to COVID-19 vaccines. The data includes demographic information (age, gender, postal code, date of death) of individuals who have received at least one dose of a COVID-19 vaccine in Belgium, details about the vaccinator, vaccination location, administered vaccine, and observed side effects. Additionally, the project encompasses data from the HealthData COVID-19 database and Clinic database, which provide demographic and clinical details of individuals tested for or hospitalized with COVID-19, as well as information from the Common Base Registry for HealthCare Actor (CoBRHA), STATBEL, and the Agency for Health Insurance Fund (AIM) databaset"@en;

METADATA IDENTIFIER

This property contains the main identifier for the Dataset, e.g. the URI or other unique identifier in the context of the Catalogue.

	DCAT-AP [dct:identifier rdfs:Literal 0..n]
HEALTHDCA T-AP Mandatory PURI	dct:identifier "https://fair.healthdata.be/dataset/d43a158e-7d13-4660-bbc3-9d3f8d5501e5"^^<http://www.w3.org/2001/XMLSchema#anyURI>;

Provide any comments, suggestions or questions:

ACRONYM

Alternative title of the dataset such as an acronym

	HEALTHDCAT-AP [dct:alternative rdfs:Literal 0..n]
Optional	dct:alternative "LINK-VACC"@en;

Provide any comments, suggestions or questions:

KEYWORD

This property contains a keyword or tag describing the Dataset.

	DCAT-AP [dcat:keyword rdfs:Literal 0..n]
Recommended	dcat:keyword "COVID-19"@en , "SARS-CoV-2"@en, "corona virus"@en, "vaccine"@en, "vaccine effectiveness"@en, "surveillance"@en;

Provide any comments, suggestions or questions:

CONTACT POINT

This property contains contact information that can be used for sending comments about the Dataset.

	DCAT-AP [dcat:contactPoint vcard:Kind 0..n]
	<pre>dcat:contactPoint [a vcard:Organization; vcard:fn "Sciensano"; vcard:hasAddress [a vcard:Address;</pre>

Recommended	<pre> vcard:country-name "BEL"; vcard:locality "Elsene - Ixelles"; vcard:postal-code "B-1050"; vcard:street-address "Rue Juliette Wytzmanstraat 14"]; vcard:hasURL <https://sciensano.be/>; vcard:hasEmail <mailto:covacsurv@sciensano.be>]; </pre>
-------------	---

Provide any comments, suggestions or questions:

GEOGRAPHICAL COVERAGE

This property refers to a geographic region that is covered by the Dataset.

	<p>DCAT-AP [dct:spatial dct:Location 0..n] OP VOCABULARY http://publications.europa.eu/resource/authority/country/</p>
Recommended HEALTHD CAT-AP Mandatory	<pre> dct:spatial <http://publications.europa.eu/resource/authority/country/BEL>; dct:spatial [a dct:Location ; locn:geometry "POLYGON((2.3889137 51.5516667,6.408097 51.5516667,6.408097 49.4969821,2.3889137 49.4969821, 2.3889137 51.5516667))"^^gsp:wktLiteral ; locn:geometry "<gml:Envelope srsName=\http://www.opengis.net/def/crs/DGC/1.3/CRS84\> <gml:lowerCorner>49.4969821 2.3889137</gml:lowerCorner> <gml:upperCorner>51.5516667 6.408097</gml:upperCorner> </gml:Envelope> "^^gsp:gmlLiteral] ; locn:geometry "{\ "type": \"Polygon\", \"crs\": {\ "type\": \"name\", \"properties\": {\ "name\": \"urn:ogc:def:crs:OGC:1.3:CRS84\"}}, \"coordinates\": [[[2.3889137, 51.5516667], [6.408097, 51.5516667], [6.408097, 49.4969821], [2.3889137, 49.4969821], [2.3889137, 51.5516667]]]"^^ https://www.iana.org/assignments/media-types/application/vnd.geo+json]. </pre>

Provide any comments, suggestions or questions:

PUBLISHER

This property refers to an entity (organisation) responsible for making the Dataset available.

	DCAT-AP [dct:publisher foaf:Agent 0..1]
Recommended NSIP Mandatory	<pre>dct:publisher [a foaf:Organization; locn:address [a locn:Address; foaf:name "Sciensano"; foaf:mbox <mailto:info@sciensano.be>; foaf:homepage <https://sciensano.be>;];]; #dct:publisher <https://org.belgif.be/id/CbeRegisteredEntity/0693876830>;</pre>

Comment: healthDCAT-AP introduces 2 new properties:

- healthcatap:publisherType to be used with a controlled vocabulary
- healthcatap:publisherNote

PUBLISHER TYPE

	healthDCAT-AP [healthcatap:publisherType 0..1]
Recommended HEALT HDCAT- AP	<pre>healthcatap:publisherType <http://healthdataportal.eu/resource/authority/publisher-type/NationalPublicHealthInstitut>;</pre>

PUBLISHER NOTE

	healthDCAT-AP [healthdcatap:publisherNote rdfs:Literal 0..1]
R e c o m m e n d e d	healthdcatap:publisherNote "Sciensano is a research institute and the national public health institute of Belgium. It is a so-called federal scientific institution that operates under the authority of the federal minister of Public Health and the federal minister of Agriculture of Belgium."@en;
H E A L T H D C A T - A P	

Provide any comments, suggestions or questions:

HEALTH DATA ACCESS BODY

This property refers to the Health Data Access Body, supporting access to electronic health data in the Member State.

	HEALTHDCAT-AP [healthdcatap:hdab foaf:Agent 1..1]
	#healthdcatap:hdab <https://healthdata.europa.eu/hdab/00000000>; healthdcatap:hdab [a foaf:Organization; locn:address [a locn:Address;

Mandatory	<pre> locn:adminUnitL1 "BEL"; locn:postCode "1210"; locn:postName "Saint-Josse-ten-Noode"; locn:thoroughfare "Galileelaan 5, Bus 2"]; foaf:mbox <mailto:info@hda.fgov.be>; foaf:homepage <https://www.hda.belgium.be>; foaf:name "Belgian Health Data Agency"; </pre>
-----------	--

Recommendation: The HealthData@EU infrastructure maintains a register of the EU HDABs

APPLICABLE LEGISLATION

The legislation that mandates the creation or management of the Data Service.

	HVDDCAT-AP [dcatap:applicableLegislation rdfs:Resource 1..n]
Mandatory (Legal resource: European Legislation Identifier (ELI))	dcatap:applicableLegislation <http://data.europa.eu/eli/reg/2022/868/oj>;

TEMPORAL COVERAGE

This property refers to a temporal period that the Dataset covers.

	DCAT-AP [dct:temporal dct:PeriodOfTime 0..n]
Recommended	<pre> dct:temporal [a dct:PeriodOfTime; #dcat:endDate "2024-12-31"^^<http://www.w3.org/2001/XMLSchema#date>; dcat:startDate "2021-01-01"^^<http://www.w3.org/2001/XMLSchema#date>]; </pre>

Provide any comments, suggestions or questions:

RETENTION PERIOD

This property refers to a temporal period during which the dataset is available for secondary use.

	HEALTHDCAT-AP [healthdcatap:retentionPeriod dct:PeriodOfTime 0..1]
Optional	<pre>healthdcatap:retentionPeriod [a dct:PeriodOfTime; dcat:endDate "2034-12-31"^^<http://www.w3.org/2001 /XMLSchema#date>; dcat:startDate "2020-03-01"^^<http://www.w3.org/2001 /XMLSchema#date>; rdfs:comment: "Provide complementary information"@en];</pre>

THEME

This property refers to a category of the Dataset. A dataset may be associated with multiple themes.

	DCAT-AP [dcat:theme, subproperty of dct:subject skos:Concept 0..n] OP vocabulary: http://publications.europa.eu/resource/authority/data-theme
Recommended HEALTHDCAT-AP Mandatory	<pre>dcat:theme <http://publications.europa.eu/resource /authority/data-theme/HEAL>;</pre>

Provide any comments, suggestions or questions:

HEALTH CATEGORY

This property refers to the datasets as described in Commission Regulation on the European Health Data Space laying down a list of minimum categories of electronic data for secondary use, Art. 33. A dataset may be associated with multiple top categories.

	HEALTHDCAT-AP [healthdcatap:healthCategory, subproperty of dct:subject skos:Concept 1..n]
--	---

	OP vocabulary: http://publications.europa.eu/resource/authority/healthdata-theme (art.33 of Regulation proposal)
HEALTHDCAT-AP Mandatory	healthdcatap:healthCategory < <a);"="" href="http://healthdata.ec.europa.eu/.../(o)">http://healthdata.ec.europa.eu/.../(o)");

HEALTH THEME

This property refers to the themes associated with the Dataset. A dataset may be associated with multiple themes.

	HEALTHDCAT-AP [healthdcatap:healthTheme, subproperty of dct:subject skos:Concept 0..n] Vocabulary: Wikidata
Recommended	healthdcatap:healthTheme < https://www.wikidata.org/wiki/Q58624061 > , < https://www.wikidata.org/wiki/Q7907952 > ;

HEALTH CODES

This property refers to health classifications and their codes associated with the Dataset. A dataset may be associated with multiple health classifications.

	HEALTHDCAT-AP [healthdcatap:hasCodeValues, skos:Concept 0..n] Thesauri: LOINC, SNOMED-CT, ICD-10, ...
	<pre> healthdcatap:hasCodeValues [a skos:Concept; skos:inScheme [a skos:ConceptScheme; dct:identifier "https://www.wikidata.org/wiki/Property:P494" ^^<http://www.w3.org/2001/XMLSchema#anyURI>; skos:prefLabel "International Classification of Diseases, 10th Revision (ICD-10)"@en; skos:definition "ICD-10 is a medical classification list by the World Health Organization."@en; skos:notation "ICD-10"; owl:versionInfo "Version:2019"]; dct:identifier "https://icd.who.int/browse10/2019/en#/Y59.0" ^^<http://www.w3.org/2001/XMLSchema#anyURI>; skos:notation "Y59.0"; skos:prefLabel "Viral vaccines"@en]; </pre>

Recommended	<pre> healthdcatap:hasCodeValues [a skos:Concept; skos:inScheme [a skos:ConceptScheme; dct:identifier "https://www.wikidata.org/wiki/Property:P494" ^^<http://www.w3.org/2001/XMLSchema#anyURI>; skos:prefLabel "International Classification of Diseases, 10th Revision (ICD-10)"@en; skos:definition "ICD-10 is a medical classification list by the World Health Organization."@en; skos:notation "ICD-10"; owl:versionInfo "Version:2019"]; dct:identifier "https://icd.who.int/browse10/2019/en#/U07.1" ^^<http://www.w3.org/2001/XMLSchema#anyURI>; skos:notation "U07.1"; skos:prefLabel "COVID-19, virus identified"@en]; </pre>
-------------	--

ACCESS RIGHTS

This property refers to information that indicates whether the Dataset is open data, has access restrictions or is not public.

	DCAT-AP [dct:accessRights dct:RightsStatement 0..1] OP Vocabulary: http://publications.europa.eu/resource/authority/access-right
Optional NSIP Mandatory	dct:accessRights < http://publications.europa.eu/resource/authority/access-right/RESTRICTED >;

Provide any comments, suggestions or questions:

CONFORMS TO

This property refers to an implementing rule or other specification.

--	--

	DCAT-AP [dct:conformsTo dct:Standard 0..n] Wikidata
Optional	dct:conformsTo < https://www.wikidata.org/wiki/Q19597236 >;

Provide any comments, suggestions or questions:

CODING SYSTEMS

This property refers to coding systems in use (e.g.: ICD-10-CM, DRGs, SNOMED CT, ...).

	HealthDCAT-AP [healthdcatap:hasCodingSystem dct:Standard 0..n] Wikidata
Optional	healthdcatap:hasCodingSystem < https://www.wikidata.org/wiki/Property:P1690 >, < https://www.wikidata.org/wiki/Property:P4229 >;

CREATOR

This property refers to the entity responsible for producing the dataset.

	[dct:creator foaf:Agent 0..n]
Optional	dct:creator < https://org.belgif.be/id/CbeRegisteredEntity/0693876830 >;

Provide any comments, suggestions or questions:

DOCUMENTATION

This property refers to a page or document about this Dataset.

	DCAT-AP [foaf:page foaf:Document 0..n]
Optional	foaf:page < https://www.sciensano.be/en/projects/linking-registers-covid-19-vaccine-surveillance >;

Comment: A best practice involves making the documentation accessible as a PURI

Provide any comments, suggestions or questions:

FREQUENCY

This property refers to the frequency at which the Dataset is updated.

	DCAT-AP [dct:accrualPeriodicity foaf:Document 0..1] OP vocabulary: http://publications.europa.eu/resource/authority/frequency/
Optional	dct:accrualPeriodicity < http://publications.europa.eu/resource/authority/frequency/DAILY >, < http://publications.europa.eu/resource/authority/frequency/UPDATE_CONT >;

Provide any comments, suggestions or questions:

HAS VERSION

This property refers to a related Dataset that is a version, edition, or adaptation of the described Dataset.

	DCAT-AP [dcat:hasVersion dcat:Dataset 0..n]
Optional	<code>dcat:hasVersion <https://opendata.schleswig-holstein.de/dataset/91ca0886-0090-497a-8fc0-bd3055d72191>;</code>

Comment:

Metadata PURI identifiers MUST to be used with dcat:hasVersion, dct:isVersionOf and dct:source.

Provide any comments, suggestions or questions:

IS REFERENCED BY

This property is about a related resource, such as a publication, that references, cites, or otherwise points to the dataset.

	DCAT-AP [dct:isReferencedBy rdfs:Resource 0..n]
Optional	<code>dct:isReferencedBy <https://doi.org/10.1186/s13690-021-00709-x>;</code>

Provide any comments, suggestions or questions:

IS VERSION OF

This property refers to a related Dataset of which the described Dataset is a version, edition, or adaptation.

	DCAT-AP [dcat:isVersionOf dcat:Dataset 0..n]
--	--

Optional	<code>dcat:isVersionOf <https://opendata.schleswig-holstein.de/dataset/a4c09d4b-9922-40f2-8615-4f4d89ff339f>;</code>
----------	--

Provide any comments, suggestions or questions:

LANDING PAGE

This property refers to a web page that provides access to the Dataset, its Distributions and/or additional information. It is intended to point to a landing page at the original data provider, not to a page on a site of a third party, such as an aggregator.

	DCAT-AP [dcat:landingPage foaf:Document 0..n]
Optional	<code>dcat:landingPage <https://sciensano.service-now.com/sp>;</code>

Provide any comments, suggestions or questions:

LANGUAGE

This property refers to a language of the Dataset. This property can be repeated if there are multiple languages in the Dataset.

	DCAT-AP [dct:language dct:LinguisticSystem 0..n] OP vocabulary: http://publications.europa.eu/resource/authority/language/
Optional	<code>dct:language <http://publications.europa.eu/resource/authority/language/ENG>;</code>

Provide any comments, suggestions or questions:

MODIFICATION DATE

This property contains the most recent date on which the Dataset was changed or modified.

	DCAT-AP [dct:modified rdfs:Literal typed as xsd:date, xsd:dateTime, xsd:gYear or xsd:gYearMonth 0..1]
Optional	dct:modified "2024-12-31T18:47:54Z"^^<http://www.w3.org/2001/XMLSchema#dateTime>;

Provide any comments, suggestions or questions:

OTHER IDENTIFIER

This property refers to a secondary identifier of the Dataset, such as MAST/ADS , DataCite , DOI , EZID or W3ID.

	DCAT-AP [adms:identifier adms:Identifier 0..n]
Optional	adms:identifier [a adms:Identifier; skos:notation " https://www.healthinformationportal.eu/health-information-sources/linking-registers-covid-19-vaccine-surveillance "^^xsd:anyURI ; adms:schemaAgency "Health Information Portal"] adms:identifier [a adms:Identifier ; skos:notation "HDBP0250"^^ex:type ; adms:schemaAgency "Health Data" ; dcterms:issued "2021-01-01"^^xsd:date]

Provide any comments, suggestions or questions:

PROVENANCE

This property contains a statement about the lineage of a Dataset.

	DCAT-AP [dct:provenance dct:ProvenanceStatement 0..n]
Q u e r y	<pre>dct:provenance [a dct:ProvenanceStatement; rdfs:label "The data for the LINK-VACC project is sourced from several existing databases, including Vaccinnet+, HealthData COVID-19 database (Contact tracing and Clinic database), CoBRHA, STATBEL, and the AIM database. These databases collectively provide comprehensive demographic, clinical, and socio-economic data relevant to the project's objectives"@en];</pre>

Provide any comments, suggestions or questions:

QUALIFIED ATTRIBUTION

This property refers to a link to an Agent having some form of responsibility for the resource.

	DCAT-AP [prov:qualifiedAttribution prov:Attribution 0..n]
Optional	<pre>prov:qualifiedAttribution [a prov:Attribution; dcat:hadRole <https://standards.iso.org/iso/19115/resources/Codelists /gml/CI_RoleCode.xml#processor>; prov:agent [a foaf:Organization; locn:address [a locn:Address; locn:adminUnitL1 "Belgium"; locn:adminUnitL2 "Brussels capital"; locn:postCode "1050"; locn:postName "Elsene - Ixelles"; locn:thoroughfare "Rue Juliette Wytzmanstraat 14"] ; foaf:homepage <https://healthdata.be>; foaf:mbox <mailto:healthdata@sciensano.be>; foaf:name "healthdata.be (Sciensano)"; foaf:phone <tel:+3227930142>]];</pre>

Provide any comments, suggestions or questions:

QUALIFIED RELATION

This property provides a link to a description of a relationship with another resource.

	DCAT-AP [dcat:qualifiedRelation dcat:Relationship 0..n]
	<pre>dcat:qualifiedRelation dcat:qualifiedRelation [a dcat:Relationship ; dct:relation <https://www.vaccinnet.be/Vaccinnet/welkom.do> ; dcat:hadRole <http://www.iana.org/assignments/relation/related>];</pre>

O
p
t
i
o
n
a
l

```
dcat:qualifiedRelation [  
  a dcat:Relationship ;  
  dct:relation <https://www.ehealth.fgov.be/ehealthplatform/fr/cobrha> ;  
  dcat:hadRole <http://www.iana.org/assignments/relation/related> ] ;  
  
dcat:qualifiedRelation [  
  a dcat:Relationship ;  
  dct:relation <https://statbel.fgov.be> ;  
  dcat:hadRole <http://www.iana.org/assignments/relation/related> ] ;  
  
dcat:qualifiedRelation [  
  a dcat:Relationship ;  
  dct:relation <https://ima-aim.be/-Nos-banques-de-donnees-?lang=fr>,  
  <https://www.ima-aim.be/-Gezondheidsdata-?lang=nl> ;  
  dcat:hadRole <http://www.iana.org/assignments/relation/related> ] ;  
  
dcat:qualifiedRelation [  
  a dcat:Relationship ;  
  dct:relation <https://www.sciensano.be/en/sciensano-and-covid-19-  
  data>, <https://www.corona-tracking.info/wp-content/uploads/2020/10/  
  /Samenwerkingsakkoord.pdf>;  
  dcat:hadRole <http://www.iana.org/assignments/relation/related> ] ;
```

Provide any comments, suggestions or questions:

RELATED RESOURCE

This property refers to a related resource.

	DCAT-AP [dct:relation rdfs:Resource 0..n]
Optional	<i>dct:relation <https://www.fedlex.admin.ch/eli/cc/1993/2080_2080_2080/fr>;</i>

Provide any comments, suggestions or questions:

RELEASE DATE

This property contains the date of formal issuance (e.g., publication) of the Dataset.

	DCAT-AP [dct:issued rdfs:Literal typed as xsd:date, xsd:dateTime, xsd:gYear or xsd:gYearMonth 0..1]
Optional	dct:issued "2023-08-19T18:47:54Z"^^<http://www.w3.org/2001/XMLSchema#dateTime>;

Provide any comments, suggestions or questions:

DISTRIBUTION

This property links the Dataset to an available Distribution.

The class Distribution has the following properties:

ACCESS URL [dcat:accessURL rdfs:Resource 1..n]

This property contains a URL that gives access to a Distribution of the Dataset. The resource at the access URL may contain information about how to get the Dataset.

AVAILABILITY [dcatap:availability skos:Concept 0..1]

This property indicates how long it is planned to keep the Distribution of the Dataset available.

DESCRIPTION [dct:description rdfs:Literal 1..n]

This property contains a free-text account of the Distribution. This property can be repeated for parallel language versions of the description.

FORMAT [dct:format dct:MediaTypeOrExtent 0..1]

This property refers to the file format of the Distribution.

LICENSE [dct:license dct:LicenseDocument 0..1]

This property refers to the licence under which the Distribution is made available.

ACCESS SERVICE [dcat:accessService dcat:DataService 0..n]

This property refers to a data service that gives access to the distribution of the dataset

BYTE SIZE [dcat:byteSize rdfs:Literal typed as xsd:decimal 0..1]

This property contains the size of a Distribution in bytes.

CHECKSUM [spdx:checksum spdx:Checksum 0..1]

This property provides a mechanism that can be used to verify that the contents of a distribution have not changed. The checksum is related to the downloadURL.

COMPRESSION FORMAT [dcat:compressFormat dct:MediaType 0..1]

This property refers to the format of the file in which the data is contained in a compressed form, e.g. to reduce the size of the downloadable file. It SHOULD be expressed using a media type as defined in the official register of media types managed by IANA.

DOCUMENTATION [foaf:page foaf:Document 0..n]

This property refers to a page or document about this Distribution.

DOWNLOAD URL [dcat:downloadURL rdfs:Resource 0..n]

This property contains a URL that is a direct link to a downloadable file in a given format.

HAS POLICY [odrl:hasPolicy odrl:Policy 0..1]

This property refers to the policy expressing the rights associated with the distribution if using the ODRL vocabulary

LANGUAGE [dct:language dct:LinguisticSystem 0..n]

This property refers to a language used in the Distribution. This property can be repeated if the metadata is provided in multiple languages.

LINKED SCHEMAS [dct:conformsTo dct:Standard 0..n]

This property refers to an established schema to which the described Distribution conforms.

MEDIA TYPE [dcat:mediaType, subproperty of dct:format dct:MediaType 0..1]

This property refers to the media type of the Distribution as defined in the official register of media types managed by IANA.

PACKAGING FORMAT [dcat:packageFormat dct:MediaType 0..1]

This property refers to the format of the file in which one or more data files are grouped together, e.g. to enable a set of related files to be downloaded together. It SHOULD be expressed using a media type as defined in the official register of media types managed by IANA.

RELEASE DATE [dct:issued rdfs:Literal typed as xsd:date, xsd:dateTime, xsd:gYear or xsd:gYearMonth 0..1]

This property contains the date of formal issuance (e.g., publication) of the Distribution.

RIGHTS [dct:rights dct:RightsStatement 0..1]

This property refers to a statement that specifies rights associated with the Distribution.

SPATIAL RESOLUTION [dcat:spatialResolutionInMeters rdfs:literal typed as xsd:decimal 0..1]

This property refers to the minimum spatial separation resolvable in a dataset distribution, measured in meters.

STATUS [adms:status skos:Concept 0..1]

The status of the distribution in the context of maturity lifecycle. It MUST take one of the values Completed, Deprecated, Under Development, Withdrawn.

TEMPORAL RESOLUTION [dcat:temporalResolution xsd:duration 0..1]

This property refers to the minimum time period resolvable in the dataset distribution.

TITLE [dct:title rdfs:Literal 1..n]

This property contains a name given to the Distribution. This property can be repeated for parallel language versions of the description.

UPDATE/MODIFICATION DATE [dct:modified rdfs:Literal typed as xsd:date, xsd:dateTime, xsd:gYear or xsd:gYearMonth 0..1]

This property contains the most recent date on which the Distribution was changed or modified.

NSIP REQUIREMENTS

At least one distribution MUST be provided with the following mandatory properties:

- dcat:accessURL
- dct:format
- dcat:byteSize
- dct:rights

HEALTHDCATAP REQUIREMENTS

- dcatap:applicableLegislation
- dcat:accessURL MUST be the landing page of the Health Data Access Body presenting the dataset
- dct:rights MUST be a citation of the EHDS2 regulation on access conditions

	DCAT-AP [dcat:distribution dcat:Distribution 0..n]
Optional NSIP Mandatory (1)	dcat:distribution < https://www.healthinformationportal.eu/health-information-sources/link-vacc/nsip_distribution >;

	< https://www.healthinformationportal.eu/health-information-
--	---

	<code>sources/link-vacc/nsip_distribution> a dcat:Distribution;</code>
HEALTHDC AT-AP Mandatory	<code>dcatap:applicableLegislation <http://data.europa.eu/eli/reg/2022/868/oj>;</code>
NSIP Mandatory	<code>dcat:accessURL <https://hda.be> ; dct:format <http://publications.europa.eu/resource/authority/file-type/XML>; dcat:byteSize "800000"; dct:rights [a dct:RightsStatement; rdfs:label "Access to data is conditional on the issuance of a permit by the HDAB after submission of a data request application"@en];</code>
	<code>dcat:mediaType <http://www.iana.org/assignments/media-types/text/tab-separated-values>.</code>

SAMPLE

This property refers to a sample distribution of the dataset.

	DCAT-AP [adms:sample dcat:Distribution 0..n]
Optional HEALTHDCATAP Mandatory or Recommended	<code>adms:sample [a dcat:Distribution ; dct:description "Synthetic data of the HealthPilot Use Case"@en; dcat:downloadURL <https://github.com/CAVDgit/EHDS2_UC_Sciensano/blob/main/use_case_1_synthetic_data_10K_individuals.csv>; dcat:mediaType <http://www.iana.org/assignments/media-types/text/tab-separated-values> ;]; adms:sample [a dcat:Distribution ; dct:description "Structural metadata expressed using the csvw RDF vocabulary"@en; dcat:downloadURL <https://www.healthinformationportal.eu/health-information-sources/link-vacc>;</code>

	<pre>dcat:mediaType "text/turtle" ;];</pre>
--	--

Recommended distribution for the structure of the dataset:

Use the csvw RDF vocabulary to provide structural metadata.

	<pre><https://www.healthinformationportal.eu/health-information- sources/link-vacc>;</pre>
	<pre><LINKVACC> a csvw:TableGroup ; csvw:table <STATBEL>, <VACCINET+>, <COBRHA> . <STATBEL> a csvw:Table ; dcterms:title "Statbel"@en ; dcat:keyword "Statbel"@en ; csvw:url <http://example.org/tree-ops-ext.csv> ; csvw:column [a csvw:Column ; csvw:name "CD_RN_STATUS" ; csvw:datatype xsd:float ; csvw:describes "Patient status: Whether patient is deceased, has migrated or is de-registered" ;], [a csvw:Column ; csvw:name "CD_COD_COVID" ; csvw:datatype xsd:string ; csvw:describes "COVID-19 specific death Patients deceased from COVID-19" ;], [a csvw:Column ; csvw:name "CD_EDU" ; csvw:datatype xsd:string ; csvw:describes "Patient educational level using ISCED classification" ;], [a csvw:Column ; csvw:name "HH_TYPE_LIPRO" ; csvw:datatype xsd:string ; csvw:describes "Patient household status" ;] . <VACCINET+> a csvw:Table ; dcterms:title "COVID-19 vaccine related data from vaccine registry"@en ; dcat:keyword "VACCINET+"@en ;</pre>

HEALTH
DCATAP
Mandatory
or
Recommended

```
csvw:url <http://example.org/tree-ops-ext.csv> ;
csvw:column [
  a csvw:Column ;
  csvw:name "Patient identifier" ;
  csvw:datatype xsd:integer ;
  csvw:describes "NISS of vaccinated person" ;
], [
  a csvw:Column ;
  csvw:name "Year of birth" ;
  csvw:datatype xsd:date ;
  csvw:describes "Year of birth, except for children <24 months
where month and year required" ;
], [
  a csvw:Column ;
  csvw:name "Date of death" ;
  csvw:datatype xsd:date ;
  csvw:describes "Patient date of death" ;
], [
  a csvw:Column ;
  csvw:name "Vaccine dose" ;
  csvw:datatype xsd:integer ;
  csvw:describes "Dose number (1st dose, 2nd dose...)" ;
], [
  a csvw:Column ;
  csvw:name "Vaccine Identification Code" ;
  csvw:datatype xsd:integer ;
  csvw:describes "CNK/ATK vaccine code : CNK code if known,
otherwise ATC code. CNK and ATC codes can also be used as
separate fields will be delivered." ;
], [
  a csvw:Column ;
  csvw:name "Date of administration" ;
  csvw:datatype xsd:date ;
  csvw:describes "Date of vaccine administration" ;
], [
  a csvw:Column ;
  csvw:name "Record date" ;
  csvw:datatype xsd:date ;
  csvw:describes "Date of recording in the registry" ;
] .

<COBRHA> a csvw:Table ;
dcterms:title "Common Base Registry for HealthCare Actor"@en ;
dcat:keyword "COBRHA"@en ;
csvw:url <http://example.org/tree-ops-ext.csv> ;
csvw:column [
  a csvw:Column ;
  csvw:name "IndividualID" ;
```

```

csvw:datatype xsd:integer ;
csvw:describes "Unique individual ID" ;
], [
a csvw:Column ;
csvw:name "HCW_Gender" ;
csvw:datatype xsd:string ;
csvw:describes "Gender of the healthcareworker" ;
], [
a csvw:Column ;
csvw:name "HCPProfessional" ;
csvw:datatype xsd:string ;
csvw:describes "Healthcare profession of the individual. Can be
more than one and is given by the authentic source 'FPS_PH' or
NIHII" ;
], [
a csvw:Column ;
csvw:name "Speciality" ;
csvw:datatype xsd:string ;
csvw:describes "Speciality performed by the individual" ;
], [
a csvw:Column ;
csvw:name "Professional_status" ;
csvw:datatype xsd:string ;
csvw:describes "Dinction between active and non-active
professionals NIHII Status : active regarding NIHII End date :
Active regarding FPS_PH" ;
] .

```

Provide any comments, suggestions or questions:

ANALYTICS

HealthDCAT-AP defines a new class of range Distribution to link the dataset to an available technical reports, quality measurements, usability indicators,...

	HEALTHDCATAP [healthdcat:analytics dcat:Distribution 0..n]
	healthdcatap:analytics [a dcat:Distribution;

R e c o m m e n d ed	<pre> dcatap:applicableLegislation <http://data.europa.eu/eli/reg/2022/868 /oj>; dct:title "Technical report number of unique study subjects available by environment for project HDBP0250"@en; dcat:accessURL <https://fair.healthdata.be/sites/default/files /distribution/d43a158e-7d13-4660-bbc3-9d3f8d5501e5 /Technical_report_number_of_unique_study_subjects_available_by_enviro nment_for_project_HDBP0250.csv> ; dct:format <http://publications.europa.eu/resource/authority/file-type /CSV>; dcat:mediaType <http://www.iana.org/assignments/media-types/text/tab- separated-values>.]; </pre>
---	--

Comment: Any API that functions as a data abstraction layer, such as the BEACON API, can be described using this new class.

Provide any comments, suggestions or questions:

SOURCE

This property refers to a related Dataset from which the described Dataset is derived.

	DCAT-AP [dct:source dcat:Dataset 0..n]
O p t i o n al	<pre> dct:source <https://fair.healthdata.be/dataset/e703dd1e-0bca-438d-a7cb- d3429dc6d118>, <https://fair.healthdata.be/dataset/6777a689-7fd7-49a0- adc7-d28fef161843>; </pre>

Comment:

Metadata PURI identifiers MUST to be used with dcat:hasVersion, dct:isVersionOf and dct:source.

Provide any comments, suggestions or questions:

SPATIAL RESOLUTION

This property refers to the minimum spatial separation resolvable in a dataset, measured in meters.

	[dcat:spatialResolutionInMeters rdfs:Literal typed as xsd:decimal 0..1]
Optional	<code>dcat:spatialResolutionInMeters "10"^^<http://www.w3.org/2001/XMLSchema#decimal>;</code>

Provide any comments, suggestions or questions:

TEMPORAL RESOLUTION

This property refers to the minimum time period resolvable in the dataset.

	[dcat:temporalResolution rdfs:Literal typed as xsd:duration 0..1]
Optional	<code>dcat:temporalResolution "P1D"^^<http://www.w3.org/2001/XMLSchema#duration>;</code>

Provide any comments, suggestions or questions:

TYPE

This property refers to the type of the Dataset. A recommended controlled vocabulary data-type is foreseen.

	DCAT-AP [dct:type skos:Concept 0..n] OP vocabulary: http://publications.europa.eu/resource/authority/dataset-type/
Optional HEALTHDCATAP Recommended	dct:type < http://publications.europa.eu/resource/authority/dataset-type/PERSONAL_DATA >;

Provide any comments, suggestions or questions:

VERSION

This property contains a version number or other version designation of the Dataset.

	[owl:versionInfo rdfs:Literal 0..1] DCAT-AP 3.0 [dcat:version rdfs:Literal 0..1]
Optional	owl:versionInfo "Project HDBP0250";

Provide any comments, suggestions or questions:

VERSION NOTES

This property contains a description of the differences between this version and a previous version of the Dataset. This property can be repeated for parallel language versions of the version notes.

	[adms:versionNotes rdfs:Literal 0..n]
Optional	adms:versionNotes "Dataset continuously updated"@en;

Provide any comments, suggestions or questions:

WAS GENERATED BY

This property refers to an activity that generated, or provides the business context for, the creation of the dataset.

	DCAT-AP [prov:wasGeneratedBy prov:Activity 0..n]
Optional	<pre>prov:wasGeneratedBy [a prov:Activity; dct:type <http://dbpedia.org/resource/Record_linkage> ; foaf:page <https://www.sciensano.be/sites/default/files /information_letter_linkvacc_20221010_en.pdf>; rdfs:label "Pseudonymisation and Anonymisation"@en ; rdfs:seeAlso <https://www.ehealth.fgov.be/ehealthplatform/fr/service- codage-anonymisation-et-ttp> prov:startedAtTime "2021-01-01"^^xsd:date ; prov:wasAssociatedWith [a prov:Agent; foaf:givenName "Dr. Joris van Loenhout"^^xsd:string; foaf:mbox <mailto:Joris.VanLoenhout@sciensano.be>; foaf:homePage <https://www.sciensano.be/fr/people/joris-van- loenhout>; prov:actedOnBehalfOf [a prov:Agent, prov:Organization; foaf:name "Sciensano"]];];</pre>

Provide any comments, suggestions or questions:

POPULATION COVERAGE

This property provides a definition of the population within the dataset.

	HEALTHDCATAP [healthdcatap:populationCoverage rdfs:Literal 1..n]
Mandatory	healthdcatap:populationCoverage "The population targeted by the LINK-VACC project comprises all individuals in Belgium who have received a COVID-19 vaccine, undergone testing for COVID-19, or have been hospitalized with a confirmed diagnosis of COVID-19. The project also considers healthcare professionals and the general Belgian population for understanding vaccination coverage and effectiveness, especially among those with comorbidities and varying socio-economic backgrounds"@en

Provide any comments, suggestions or questions:

MIN AND MAX AGE

HealthDCAT-AP introduces two new properties to specify the minimum and maximum ages of the population within the dataset, expressed in years.

	HEALTHDCATAP [healthdcatap:min/maxTypicalAge rdfs:Integer 1..1]
Mandatory	healthdcat:minTypicalAge: "18"; healthdcat:maxTypicalAge: "64";

Provide any comments, suggestions or questions:

DATASET DIMENSIONS

HealthDCAT-AP introduces two new properties: one defining the size of the dataset in terms of the number of records, and the other specifying the number of records per unique individual.

	HEALTHDCATAP [healthdcatap:numberOfRecords rdfs:Integer 1..1] HEALTHDCATAP [healthdcatap:numberOfUniqueIndividuals rdfs:Integer 1..1]
Mandatory	healthdcatap:numberOfrecords: "124866488"; healthdcatap:numberOfUniqueIndividuals: "8914722";

Provide any comments, suggestions or questions:

LEGAL BASIS

The legal basis used to justify processing of personal data

	HEALTHDCATAP [dpv:hasLegalBasis dpv:LegalBasis 1..n]
Mandatory	<pre> dpv:hasLegalBasis [a dpv:LegalBasis ; dct:description "CSI Deliberation no. 21/028 of february 18, 2021, last amended on june 18, 2021, relating to the communication of data to pseudonymized personal character relating to the health of vaccinnnet+, healthdata covid-19 database i and ii, healthdata covid-19 clinical database, cobrha, statbel and the agency intermutualist in sciensano, as part of the link-vacc project and the subsequent processing of personal data pseudonymised by the federal drug agency in view monitoring the safety of covid-19 vaccines"@en; dct:source <https://www.ehealth.fgov.be/ehealthplatform/file/view /AXkNfdPml9vUufvGGfJr?filename=21-028-f212-AFMPS-vaccinnnet-modifi%C3%A9e%20le%2018%20juin%202021.pdf>, <https://www.ehealth.fgov.be /ehealthplatform/file/view/AX_-9sZSuwVJMAnc0ENo?filename=21-028-f166- LINK-VACC-modifi%C3%A9e%20le%205%20avril%202022.pdf> ;] ; </pre>

Provide any comments, suggestions or questions:

PERSONAL DATA

This property enumerates the key elements that represent an individual in the dataset.

	HEALTHDCATAP [dpv:hasPersonalData dpv:PersonalData 0..n]
Recommended	<code>dpv:hasPersonalData dpv-pd:Gender, dpv-pd:Age, dpv-pd:Location, dpv-pd:Nationality, dpv-pd:Education, dpv-pd:HealthRecord;</code>

Provide any comments, suggestions or questions:

PURPOSE

This property contains a free text statement of the purpose of the processing of personal data.

	HEALTHDCATAP [dpv:hasPurpose dpv:Purpose 1..n]
Manifestation	<pre>dpv:hasPurpose [a dpv:Purpose ; dct:description " The primary objective of Sciensano's LINK-VACC project is to monitor COVID-19 vaccines post-authorization and evaluate the public health value of prioritizing vaccination for people with comorbidities. This involves assessing the vaccines' effectiveness and safety in the broader population context, beyond the limited scope of clinical trials, and determining future vaccination policies in public health emergencies such as epidemics or pandemics "@en;] ;</pre>

Provide any comments, suggestions or questions:

QUALITY LABEL

This property refers to a quality and utility label as defined by the QUANTUM health data quality framework.

	HEALTHDCATAP [dqv:hasQualityAnnotation dqv:QualityCertificate 1..1]
Mandatory	<pre>dqv:hasQualityAnnotation [a dqv:QualityCertificate ; oa:hasTarget <https://certificates.theodi.org/en/datasets/393> ; oa:hasBody <https://certificates.theodi.org/en/datasets/393/certificate> ; oa:motivatedBy dqv:qualityAssessment];</pre>

Provide any comments, suggestions or questions:

Contact

[Contact Form](#)