

M5.1 Draft guideline on data description – public consultation questions

1 Part A questions for generic feedback

These questions will be asked in each public consultation to provide an understanding of the recipients' demographics, the quality of the document and to gather generic feedback. Questions marked with an asterisk are mandatory.

1.1 Demography

Country*

Sweden

Type of the responder*

Academic or research institution

Sector*

Education and academia

Organisation size*

Large enterprise (250+ employees)

Professional role/function

1.2 Quality

From your perspective, how ready is the document to meet the expected needs?*

Well-developed

What is level of quality of the document?*

High

Is the document clear and easy to understand?*

4 (very clear and easy to understand)

How well does the document address the key issues and challenges related to its subject matter?*

4 (very well)

How feasible and implementable do you find the recommendations or technical specifications presented in the document?*

4 (very feasible and implementable)

1.3 Generic feedback

Do you have any suggestions for improving the document? Are there any additional topics or areas that should be covered? [Please provide feedback and ideas for enhancing the document] [max. 750 characters]

page 9, Dataset catalogue:

It would be beneficial if machine-readability is emphasised more strongly.

page 29: CSVW format

CSVW format is quite new and not yet widely spread. We suggest that the text is rephrased to: "The data dictionary must be published in a machine-readable and actionable format, such as CSVW or an equivalent format that meets these criteria."

2 Part B questions for specific feedback

These questions are specific for TEHDAS2 WP5 Task 5.1, for Milestone 5.1 on Data Description Duties for Health Data Holders.

2.1 Part 1: Role and experience

1.1 Are you a health data holder, according to the definition of the EHDS?*

Yes, future data holder as the Swedish node in 1+MG/GDI.

1.2 If yes, do you already create metadata for the health data you hold? [Yes, No]

No, not yet, but we plan to use Health-DCAT-AP.

1.3 If yes, which metadata standard are you using? [text field]

1.4 If no, why not? [text field]

1.5 Do you already share or expose your metadata?

No, not yet in production phase; awaiting legal framework.

1.6 If yes, how? [text field]

2.2 Part 2: Properties of the proposed solution HealthDCAT-AP

2.1 In your opinion, is the HealthDCAT-AP sufficient for health data holders to fully and accurately describe their datasets?*

1 (No, not sufficient)

2.2 In what ways could the HealthDCAT-AP be tailored to better support the unique characteristics of (your) health datasets?* [text field]?*

There are shortcomings, that in case HealthDCAT-AP becomes the European standard, each country need to solve on their own. For instance, it is not possible to, in a structured way, describe for which variables there are data available.

2.3 What aspects of (your) health datasets are most challenging to describe using HealthDCAT-AP? [text field]

2.4 How would you compare the HealthDCAT-AP to other metadata standards you have used/are using? What aspects stand out as strengths or weaknesses? [text field]

2.5 Feedback on the current cardinalities to describe personal electronic health datasets through HealthDCAT-AP [matrix, scale selection] *Note.* M = mandatory | R = Recommended | O = Optional

Property	Definition	Current proposal	Should be Mandatory	Should be Recommended	Should be Optional
dcat:keyword	Keywords or tags describing the dataset	M			
dct:provenance	Information about how the data was collected, including methodologies, tools, and protocols used	M			
dpv:hasPurpose	A free text statement of the purpose of the processing of data or personal data.	M			
dct:issued	The date of formal issuance (e.g.: publication) of the dataset	O		R	
dct:modified	The most recent date on which the dataset was changed or modified.	O	M		
healthdcatap:populationcoverage	A definition of the population within the dataset	R			
dct:spatial	The geographic area covered by the dataset.	M			
dct:temporal	A temporal period that the Dataset Series covers.	R			
dct:language	The language(s) of the dataset content.	R			
healthdcatap:retentionPeriod	A temporal period which the dataset is available for secondary use.	O			
dct:accrualPeriodicity	Frequency of updates to the dataset (e.g., daily, monthly, annually).	R			
dcat:spatialResolutionInMeters	The minimum spatial separation resolvable in a dataset, measured in meters.	O			
dcat:temporalResolution	The minimum time period resolvable in the dataset.	R			
dct:publisher	The organisation or entity responsible for publishing the dataset.	M			
healthdcatap:publisherType	A type of organisation that makes the Dataset available.	M			

healthdcatap:publishernote	A description of the publisher activities	O			
healthdcatap:hdab	Health Data Access Body supporting access to data in the Member State.	M			
dct:creator	An entity responsible for producing the dataset.	O			
dcat:contactPoint	Contact information for inquiries about the dataset.	M			
foaf:page	A page or document about this Dataset.	R			
dcat:landingPage	A web page that provides access to the Dataset, its Distributions and/or additional information.	R			
dcat:qualifiedRelation	A description of a relationship with another resource.	O			
prov:qualifiedAttribution	An Agent having some form of responsibility for the resource.	O			
prov:wasGeneratedBy	An activity that generated, or provides the business context for, the creation of the dataset.	O			
dqv:hasQualityAnnotation	Dataset, including rating, quality certificate, feedback that can be associated to the dataset.	R			
dcat:hasVersion	A related Dataset that is a version, edition, or adaptation of the described Dataset.	O			
dcat:version	The version indicator (name or identifier) of a resource.	O			
adms:versionnotes	A description of the differences between this version and a previous version of the Dataset.	O			
dcat:isVersionOf	inverse of dcat:hasVersion	O			
dct:relation	A related resource	R			

dct:isReferencedBy	A related resource, such as a publication, that references, cites, or otherwise points to the dataset.	R			
dct:source	A related Dataset from which the described Dataset is derived.	R			
dpv:hasLegalBasis	The legal basis used to justify processing of personal data	R			
adms:identifier	A secondary identifier of the Dataset, such as MAST/ADS17, DataCite18, DOI19, EZID20 or W3ID21	O			
dcat:theme	The theme or category of the dataset: Health (M by default, with the option to add other themes)	M			
dcatap:applicableLegislation	The legislation that mandates the creation or management of the Dataset.	M			
healthdcatap:healthCategory	The health category to which this dataset belongs as described in the EHDS Article 51	M			
healthdcatap:healthTheme	A category of the Dataset or tag describing the Dataset.	M			
dct:type	A type of the Dataset: for health datasets containing personal level information, the type of the dataset MUST take the value "personal data".	M			
dpv:hasPersonalData	Key elements that represent an individual in the dataset. Example: Age Exact, Blood type, Current Employment, etc.	R			
dct:conformsTo	An implementing rule or other specification.	R			
healthdcatap:hasCodingSystem	Coding systems in use (ex: ICD-10-CM, DGRs, SNOMED=CT, ...)	R			
healthdcatap:minTypicalAge healthdcatap:maxTypicalAge	Minimum/maximum typical age of the population within the dataset	R			
healthdcatap:numberOfRecords	Size of the dataset in terms of the number of records.	R			

healthdcatap:numberOfUniqueIndividuals	Number of records for unique individuals.	R			
dcat:distribution	An available distribution for the dataset. For personal electronic health data, a distribution must include the landing page of the Health Data Access Body supporting data access.	M			
adms:sample	A sample distribution of the dataset.	M		R	
healthdcatap:analytics	An analytics distribution of the dataset.	R			
dct:identifier	The main identifier for the Dataset, e.g. the URI or other unique identifier in the context of the Catalogue.	M			

2.3 Part 3: Usability of the proposed solution HealthDCAT-AP

3.1 How adequate do you find the HealthDCAT-AP standard for describing (your) health datasets?* 3

3.2 Are there specific challenges you will face in the implementation of HealthDCAT-AP?

We foresee some challenges in implementing HealthDCAT-AP for smaller datasets, e.g. survey data collected in academic studies, where resources and expertise for open and shareable data are not always available.

3.3. How do you foresee using the HealthDCAT-AP standard to improve data discoverability and interoperability?* [text field]

3.4 On a scale of 1 to 4, how clear did you find the guideline on data description?*

3

3.5 What improvements would this guideline need in order to be clear and useful?* [text field]

3.5 What kind of additional support, such as training, documentation, or tailored tools, would help you better adopt the standard? [text field]