THE SECOND JOINT ACTION TOWARDS THE EUROPEAN HEALTH DATA SPACE

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

Guideline for health data holders on their duties regarding data description

TEHDAS2 – Second Joint Action Towards the European Health Data Space

19 December 2024

This project has been co-funded by the 4th EU Health Programme (2021–2027) under Grant Agreement no 101176773.

Contents

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* [-List of countries-, European Union, Outside European Union]

Belgium

Type of the responder* [Public organisation, Private organisation, Non-governmental organisation (NGO), Academic or research institution, Industry association, Individual expert or professional, Patient representative, Citizen, Other (please specify)]

NGO

Sector* [Health care provider, Government/public administration, Research and development, Medical devices, Pharmaceutical industry, Education and academia, Information technology, Data Management/processing, Patient advocacy, Legal and compliance, Other (please specify)]

Other, healthcare professional organisation representative

Organisation size* [Micro (1–9 employees), Small to medium enterprise (10–249 employees), Large enterprise (250+ employees), Not applicable (for individual respondents)]

Micro

Professional role/function [open text field]

Senior Policy Advisor

1.2 Quality

From your perspective, how ready is the document to meet the expected needs?* [Early draft, **Advanced draft**, Well-developed, Needs finalisation only]

What is level of quality of the document?* [Low, Medium, High]

Is the document clear and easy to understand?* [Rate 1 (not clear nor easy to understand) – 4 (very clear and easy to understand)] **Not applicable**

How well does the document address the key issues and challenges related to its subject matter?* [Rate 1 (not well) – 4 (very well)] **Not applicable**

How feasible and implementable do you find the recommendations or technical specifications presented in the document?* [Rate 1 (not feasible and implementable at all) – 4 (very feasible and implementable)] **2**

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]

- Healthcare software manufacturers need to support healthcare professionals and healthcare providers by automating the creation and provision of metadata descriptions for health data access bodies. They must be addressed first.
- The duties for healthcare professionals and healthcare providers related to data descriptions and data transfers under the secondary use regime of the EHDS Regulation are seen as an administrative burden to their primary task of diagnosing and treating a patient.
- Considering the current health workforce crisis, healthcare professionals' precious time cannot be exhausted in preparing data for third parties including those which pursue private and commercial interest. CPME supports the 'once-only principle'.

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, No]

Natural or legal person, public authority, agency or other body in the healthcare or the care sectors, including reimbursement services where necessary, as well as any natural or legal person developing products or services intended for the health, healthcare or care sectors, developing or manufacturing wellness applications, performing research in relation to the healthcare or care sectors or acting as a mortality registry, as well as any Union institution, body, office or agency, that has either:

(i) *the right or obligation, in accordance with applicable Union or national law and in its capacity as a controller or joint controller, to process personal electronic health data for the provision of healthcare or care or for the purposes of public health, reimbursement, research, innovation, policy making, official statistics or patient safety or for regulatory purposes; or*

(ii) *the ability to make available non-personal electronic health data through the control of the technical design of a product and related services, including by registering, providing, restricting access to or exchanging such data.*

Please note that the following categories of health data holders shall be exempt from the obligations on health data holders: (a) natural persons, including individual researchers; (b) legal persons that qualify as microenterprises.

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

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? [Yes, No]

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?* [Rate 1 (No not sufficient) – 4 (Yes very sufficient)]

Not applicable

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

Not applicable

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

- The maturity and capability of health data holders is different, and this difference is not addressed in the deliverable.
- Requesting health data holders, which includes healthcare professionals, to provide dataset descriptions in the HealthDCAT-AP standard, and update the information regularly, is disproportionate. HCP should not be burdened with this standardisation task.

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			
dct:modified	The most recent date on which the dataset was changed or modified.	O			
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			
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?* [Rate 1 (Not adequate) – 4 (Very adequate)] Not applicable

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

- The HealthDCAT-AP standard does not include any reference to ethical requirements or risk to medical ethics.
- Mandatory metadata descriptions should be minimal.
- As a health data holders, I should not use extra resources to describe metadata that will not be used. Not all metadata descriptions are useful.
- As a data holder, I should be able to describe metadata in free text format, and it should be up to the health data access body to translate into the appropriate standard. The tools and methods to do so should be sorted between Member States and the healthcare software industry, which needs to be addressed first.

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

Not applicable

3.4 On a scale of 1 to 4, how clear did you find the guideline on data description?* [Rate 1 (not clear) – 4 (very clear)]

Not applicable

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

- The maturity and capability of health data holders is different, and this difference is not addressed in the deliverable.
- The deliverable needs to be coordinated with deliverable M5.3. It should be clearer in deliverable M5.3 who should transcribe the meta data description into the HealthDCAT-AP standard (section 7.1. in deliverable M5.3). HCP should not be burdened with this standardisation task.
- Doctors should be required to code only once for the continuity of care in the EHR. CPME welcomes current international initiatives which aim to explore seamless data conversion and linkages of different coding terminologies for users. European doctors call for further engagement from coding organisations to enable a common language worldwide.

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

- Health data holders require specific training on data management which currently they (healthcare providers) do not have or receive.
- The role of health data intermediaries is not addressed. To what extend would be their role and how can they support HCPs which fit into the description of health data holders? Doctors should not be data harvesters for other users and interests.

- Secondary use data should be generated from structured primary data. The EHR system should provide for multiple uses across uses, including billing and statistical reporting.