Medical Informatics Initiative

Supporting Project – Central Office of the National Steering Committee

MII Metadata on Data Availability, Analysis Opportunities, and Cooperation Options

Last revised: March 23, 2017

DISCLAIMER: The official document has been adopted in German language.

Table of Contents

1. Introduction and Objectives

In connection with the Medical Informatics Initiative (MII), data is to be exchanged between data integration centers (DICs) and data recipients. In order to be able to determine whether the data from a location are relevant for an undertaking and can be made available to the undertaking in an appropriate form, information on the data and on the underlying conditions for data sharing is required.

There are a number of areas of intersection with the work of some of the other working groups and drafting groups of the MMI's monitoring body. With regard to "typical" metadata, i.e., metadata that describe syntactic, structural, and semantic attributes of individual interfaces, data elements, or data collections, efforts should be made toward coordinating with the results from the Core Data Set Drafting Group. Data provenance aspects allow (with certain limitations) for inferences regarding the quality of the data and contain information on data controllers, rights to the data, and authorship. In this regard it will be necessary to coordinate with the work of the Data Sharing Working Group, in particular with the uniform use and access policy developed by that group. Key underlying conditions for data sharing will also be defined by way of the actual construction of the consent forms. Coordination with the uniform consent form developed in the Data Sharing Working Group is therefore absolutely necessary.

The underlying conditions developed by the working groups can and must first lay down simple, minimum underlying conditions, which will not reflect all options for data use. For instance, higher resolution data or particular types of data (image data, genetic data) might be made available only in compliance with specific security precautions (safe setting). However, these specific forms of data sharing can only be used on a widespread basis once they have been systematized and – together with all of the relevant underlying conditions – rendered transparent in structured form by means of metadata. Another key aspect is that the underlying conditions for data sharing may also have a direct impact on the content of the data shared. So, for example, consent forms granting varying degrees of consent may mean that data on some patients is not able to be shared. In such cases, having corresponding information – for example how many data sets are not able to be returned in response to a query for which reasons (lack of consent vs. lack of data collected) – available as metadata would be a welcome development.

For these reasons, developing structured, standardized, and automatically analyzable metadata on data availability appears to be a necessary element of the MII. The possibility of being able to provide and check a machine-readable representation at granular levels for both large volumes of data and individual data subsets, thereby avoiding time-consuming manual processes or simplifying and

reducing organizational complexity, is a key motivational factor for the development of such a representation. Over the long term, achieving an extensively automated form of data sharing, even in complex conditions (think “smart contracts”) might actually be possible.

2. Methodology

An iterative, multi-stage approach will be arranged for continuing work. For this process, the workflow below can also be run through multiple times:

1.) The first step is to coordinate which types and items of information are to be taken into consideration in connection with the structuring and standardization undertaking. This step will also include prioritizing by timing and content. One key prerequisite for coordination is an analysis of the use cases and queries foreseen in the consortia and at the national level. The dates for conducting corresponding analyses will therefore be guided by the Roadmap milestones (see also Section 3).
2.) With regard to selected types and items of information, research is to be conducted into the extent to which international standards for representation, communication, and processing already exist. Generally, work undertaken in this regard that is not aligned with international standards and developments should be avoided. Where necessary, existing standards should be adapted based on predetermined paths for further development, or national extensions should be established. In some areas, guidance will be able to be taken, if need be, from informally standardized best practices. And equally important, the possibility of the MMI’s results being linked at a later date at the international level will also depend on this step being implemented with great care.
3.) In the final step, the adapted and developed solutions chosen are to serve as the basis for implementation. Versioning is to be implemented for all results. This also includes a technically, fully backwards compatible implementation of all approved versions in order to ensure flexible communications across locations and consortia.

3. Relevant Information

The following list constitutes an initial compilation – by no means definitive or exhaustive – of relevant information on availability that should prospectively be made available in the form of structured, standardized, and automatically analyzable metadata. However, this compilation is to be refined and adjusted over the course of the Medical Informatics Initiative within the context of an iterative approach. Initial pointers regarding relevant prep work or standards that can serve as examples have been included, although again without any claim to being exhaustive.

- **Informed consent information sheets**
    - Aspects **(time frame: Roadmap M 1.2 (Q4/18, Q4/20)**)

        - Specimens vs. data or data types
        - Purpose restrictions and/or use restrictions
        - Ability to be recontacted
        - Transparency requirement
    - Standards
        - IHE BPPC, APPC, HL7 FHIR (e.g., ConsentDirective resource)
    - Prep work
        - "Automatable Discovery and Access Matrix" from the Global Alliance for Genomics and Health[1]
- **Underlying organizational conditions**
    - Provenance aspects **(time frame: Roadmap M 5.3 (Q3/20)**)
        - What are the source departments for the data
        - Rights to the data/authorship
        - Involved/relevant IT systems
    - Use aspects **(time frame: Roadmap M 3.2 (Q3/19)**)
        - Depiction of key aspects of the use and access policy
        - Depiction of relevant committees/decision-making processes
    - Prep work
        - Organizational structure in the hospitals' quality reports (Quality Report Directive, "Qb-R")[2,3]
        - "Automatable Discovery and Access Matrix" from the Global Alliance for Genomics and Health[1]
- **Context of data use (time frame: Roadmap M 3.2 (Q3/19), M 4.2. (Q3/19))**
    - Underlying conditions applicable for a specific query
        - E.g., attributes of the query, of the querying institution, authentication of the querying party, preliminary agreements, existing approvals and committee decisions
- **Data types (e.g., clinical phenotype data vs. genome data) (time frame: Roadmap M 5.1 (Q4/17, Q4/18, and Q4/19)**)
    - Reference to results from the Core Data Set Drafting Group
    - Prospectively, also metadata on the availability of biospecimens (in coordination with the German Biobank Alliance, GBA, and the information from the core data set).

[1] https://genomicsandhealth.org/work-products-demonstration-projects/automatable-discovery-and-access-matrix

[2] https://www.g-ba.de/institution/themenschwerpunkte/qualitaetssicherung/qualitaetsdaten/qualitaetsbericht/servicedateien/ (XML schema)

[3] https://www.g-ba.de/downloads/62-492-1235/Qb-R_2016-07-21_iK_2016-08-06.pdf (attachment 1)

- **Type of access to the data/queries**
  - Aspects **(time frame: Roadmap M 7.1 (Q4/18), M 4.2. (Q3/19))**
    - Safe setting (e.g., query interface)
    - Safe data (e.g., turning over microdata)
  - Prep work
    - Directives from the German Federal and State Statistics Offices[4]
    - Directives from the German Institute of Medical Documentation and Information (DIMDI) on the data transparency process
- **Data quality**
  - Provenance aspects **(time frame: Roadmap M 5.3 (Q3/20)**)
    - Purpose for data collection (studies, treatment, etc.)
    - Preprocessing work (e.g., for data privacy purposes)
  - Ultimately, data quality always depends on the nature of the query **(time frame: Roadmap M 7.1 (Q4/18)**)
    - Typical queries may need to be defined
  - Prep work
    - Guidelines on registry and cohort data quality (Volume 4 of the TMF series)
    - German Research Foundation (DFG) project on the practical assessment of data quality in cohort studies (directed by Mr. C. Schmidt from Greifswald)
    - In the concept phase, it is the role of the monitoring body to author a catalogue of requirements for an opinion paper on data quality and evaluation of usability of bodies of data (part of the TMF application to the monitoring body). The paper itself is to be drawn up during the setup phase.
    - Experiences with data quality in connection with external quality assurance pursuant to §137 of Volume V of the Social Security Statute Book (SGB V)
- **Options for reporting information back (time frame: Roadmap M 3.2 (Q3/19), M 1.2 (Q4/18 and Q4/20)**)
  - The following occurrences call for an appropriate response to be sent back to the location or the IT system of the primary record/enrichment/curation of data
    - Errors are detected in the data turned over
    - As the data turned over is used, indications of serious illnesses are discovered, and it cannot be determined based on the current analysis scenario/use case that a corresponding diagnosis was made.
  - In such cases, the data recipient/analyzing party is under no obligation to perform analyses/evaluations of the data turned over that extend beyond the immediate intended use.

[4] http://www.forschungsdatenzentrum.de/datenzugang.asp

- **Representativeness (inclusion/exclusion criteria)**
  - DIC-related aspects **(time frame: Roadmap M 7.1 (Q4/18)**
    - Period of data provision
    - Percent without consent forms
    - etc.
  - Query-related aspects **(time frame: Roadmap M 7.1 (Q4/18)**
    - Which data are available but are unable to be provided for the query, e.g., due to lack of sufficient consent

Drafting Group Members and Contributors

Johannes Drepper
Thomas Ganslandt
Josef Ingenerf
Fabian Prasser (Leader, Metadata on Data Availability, Analysis Opportunities, and Cooperation Options Drafting Group)
Ulrich Sax
Thomas Wendt
Sven Zenker