Project title: Implementing verifiable oncological imaging by quality assurance and optimisation

Grant Agreement: 101056832

Call identifier: EU4H-2021-PJ

Topic: EU4H-2021-PJ-03 Action grants for a project on the quality and safety of radiation technology in diagnosis and treatment of cancer

Deliverable D5.1 Data Management Plan

Lead partner: UMC-Mainz

Author(s): Peter Mildenberger (UMC-Mainz)

Work Package: 5

Estimated delivery: Month 4 (31.12.2022)

Actual delivery: Month 4 (30.12.2022)

Type: Report

Dissemination level: Public

This project is co-funded under the EU4Health Programme 2021-2027 under grant agreement no. 101056832.

Co-funded by the European Union

Table of Contents

Disclaimer

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 the European Health and Digital Executive Agency (HaDEA). Neither the European Union nor the granting authority can be held responsible for them.

Abbreviations

ABR	Abbreviation
BC	Bronchial Cancer
CRC	Colorectal Cancer
CT	Computed Tomography
DICOM	Digital Imaging and Communication in Medicine
DoA	description of action
GB	gigabyte
HTTPs	Hypertext Transfer Protocol Secure
IHE	Integrating the Healthcare Enterprise
IT	Information Technology
MB	megabyte
MRRT	Management of Radiology Reporting Templates
PACS	Picture Archiving and Communication System
RAID	Redundant Array of Independent Disks
SC	Stomach Cancer
SO	Specific Objectives
SQL	Standardized Query Language
UID	Unique Identifier
VR	Value Representation

List of figures

List of tables

1. PREAMBLE

The EU project i-Violin aims to achieve quantifiable optimization of image quality and radiation exposure. For this purpose, CT examinations for three oncological entities will be recorded, qualitatively and quantitatively analysed in connection with relevant metadata. This document is the first Data Management Plan (DMP) submitted at the beginning of the data collection for i-Violin as Deliverable D5.1.

2. DATA SUMMARY

Provide a summary of the data addressing the following issues:

- **State the purpose of the data collection/generation**
- **Explain the relation to the objectives of the project**
- **Specify the types and formats of data generated/collected**
- **Specify if existing data is being re-used (if any)**
- **Specify the origin of the data**
- **State the expected size of the data (if known)**
- **Outline the data utility: to whom will it be useful**

The i-Violin project aims to address qualitative and quantitative parameters for optimizing image quality and radiation exposure. For this purpose, corresponding image data from CT examinations and the associated metadata are required. (SO1, SO3, SO4)

CT data from three different entities are needed for the project. The image data will be transferred and stored in the DICOM format. A structured template based on IHE's MRRT profile will be used for the metadata.

Image data will be acquired from clinically indicated examinations, and no additional radiation exposure will be required for the project.

Image data will be generated by six clinical partners in the project (OVGU, UMC-Mainz, UoC, UCHD, UMCL, UCHD).

The size of a single CT image is typically 0.5 MB, CT examinations in i-Violin will probably have around 1000 images per examination, in total 600 examinations are needed for the project according to the current status. For the metadata, there are only small volumes per dataset, these will be in the order of a few megabytes. Therefore, the total data volume is assumed to be in the range of 300-500 GB.

The results of the project are relevant for all persons dealing with oncological imaging and radiation exposure.

Clinical Entity	Datatype and Format	No. of data sets	Images/Study	Total size
Bronchial Cancer (BC)	CT images and dose reports using DICOM objects	200	1000 (or 500 MB)	100 GB
Stomach Cancer (SC)		200	1500 (or 750 MB)	150 GB
Colorectal Cancer (CRC)		200	1500 (or 750 MB)	150 GB

Table 1: Overview on entities and data volume

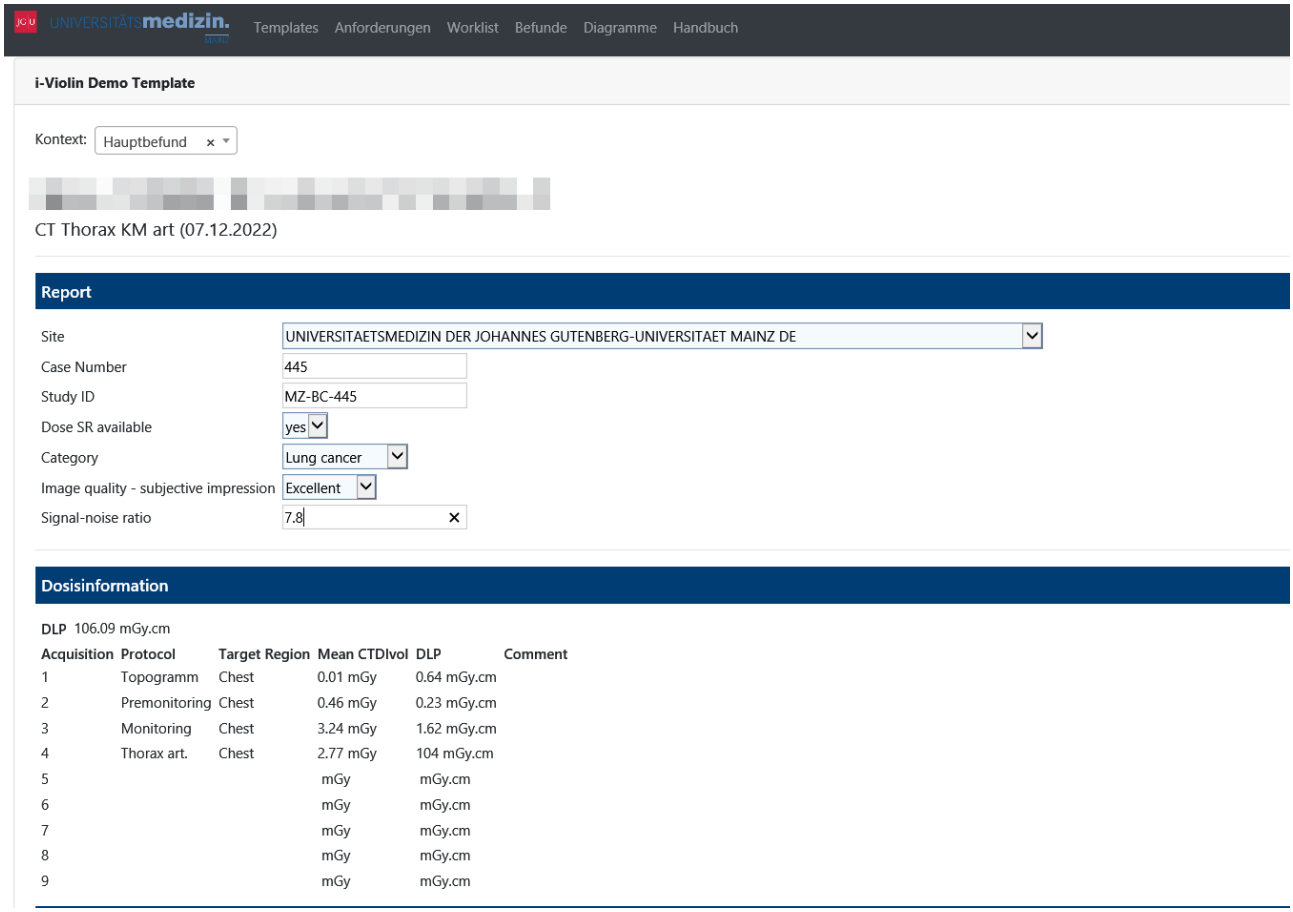

Figure 1: Demo – template for additional meta-data registration using MRRE (Mainz Radiology Report Engine) – the final templates are still under development according to the different WPs

3. FAIR DATA

Making data findable, including provisions for metadata

- **Outline the discoverability of data (metadata provision)**
- **Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?**
- **Outline naming conventions used**
- **Outline the approach towards search keyword**
- **Outline the approach for clear versioning**
- **Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how**

The identification and description of data sets is based on the DICOM standard procedure established in clinical and scientific circles. This is used to generate unique data sets. The image data are anonymized at the point of origin according to the DICOM standard. A standardized procedure is used for the designation of the institution and the identifier of the clinical entity (BC, SC, CRC) as well as the respective data set. The additionally captured metadata are uniquely assigned via the DICOM UIDs. All data captured in the structured templates are recorded in a database and are to be analysed here according to SQL criteria. In addition, versioning is supported for the use of the templates. The templates are based on the IHE profile MRRT. This means that they can in principle be exchanged with other research institutions.

Making data openly accessible

- **Specify which data will be made openly available? If some data is kept closed provide rationale for doing so**
- **Specify how the data will be made available**
- **Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?**
- **Specify where the data and associated metadata, documentation and code are deposited**
- **Specify how access will be provided in case there are any restrictions**

Within the i-Violin project, no funds are foreseen to build a publicly accessible repository. In principle, the data can later be transferred to a corresponding European project that offers a corresponding database for image studies. The exchangeability of the data is ensured by the use of the standards. The prerequisite for this is the respective clarification of data protection and ethical aspects, e.g. the consent of the patients, if this is required in the clinical institutions carrying out the study for participation.

Making data interoperable

- **Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability**
- **Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?**

The image data itself is captured and stored in the DICOM standard. This is used worldwide and is highly interoperable due to its specification. The additionally captured metadata is recorded in structured templates that comply with the IHE profile MRRT and thus also fulfil interoperability requirements. Categorization is performed according to the standards and profiles used, predominantly with Snomed-CT and RadLex.

Increase data re-use (through clarifying licenses)

- **Specify how the data will be licensed to permit the widest reuse possible**
- **Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed**
- **Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why**
- **Describe data quality assurance processes**
- **Specify the length of time for which the data will remain re-usable**

- No funds for long-term data storage were requested as part of the project application. Therefore, the availability of data after the end of the project depends on local conditions of the project partners. For the centrally stored image data and evaluations, longer-term availability is possible.

- Since the image data are medical data, the requirements of data protection and ethics committees of the respective image-generating institutions will have to be taken into account. The process of clarifying these aspects is currently ongoing.

- It is expected that a contract regulating the ethical and data protection aspects may be required for use by third parties.

- On the part of the project partners, no specific licensing requirements are envisaged; if a publication is planned on the basis of the data collected, an embargo would be intended until the work has been accepted.

4. ALLOCATION OF RESOURCES

Explain the allocation of resources, addressing the following issues:

- Estimate the costs for making your data FAIR. Describe how you intend to cover these costs
- Clearly identify responsibilities for data management in your project
- Describe costs and potential value of long-term preservation

No funding has been requested for the IT infrastructure as part of the project. Existing IT systems of the partners are used for the implementation of the project. The central archiving of the image data as well as the hosting of the server for the structured templates will be taken over by the IT of UM-Mainz. The storage of the data generated in the project can be carried out over the duration of the project.

5. DATA SECURITY

Address data recovery as well as secure storage and transfer of sensitive data

The same standards and procedures used in clinical imaging are used for data security. For storage, this means the use of a fail-safe PACS cluster and backup of image data with Worm functionality. Automated, standardized backup procedures are used for the databases, which also enables fast and efficient recovery (snapshot, RAID systems ...).

The transmission of image data is encrypted using the established DICOM e-mail procedure. Metadata acquisition is also performed using encrypted web technologies (HTTPs).

The original DICOM objects will be modified on site before uploading into the i-Violin repository. The de-identification will be strictly following the DICOM standard (Table E.1-1- Application Level Confidentiality Profile Attributes from Part 15 – Security and System Management Profiles) with following items:

Attribute Name	Tag	Basic Profile	Retain UIDs	Retain Patient Characteristics	Clean Descriptors
Patient's Address	(0010,1040)	X			
Patient's Age	(0010,1010)	X		K	
Patient's Birth Date	(0010,0030)	Z			
Patient's Birth Name	(0010,1005)	X			
Patient's Birth Time	(0010,0032)	X			
Patient's Institution Residence	(0038,0400)	X			
Patient's Insurance Plan Code Sequence	(0010,0050)	X			
Patient's Mother's Birth Name	(0010,1060)	X			
Patient's Name	(0010,0010)	Z			
Patient's Primary Language Code Sequence	(0010,0101)	X			

Patient's Primary Language Modifier Code Sequence	(0010,0102)	X			
Patient's Religious Preference	(0010,21F0)	X			
Patient's Sex	(0010,0040)	Z		K	
Patient's Sex Neutered	(0010,2203)	X/Z		K	
Patient's Size	(0010,1020)	X		K	
Patient's Telecom Information	(0010,2155)	X			
Patient's Telephone Numbers	(0010,2154)	X			
Patient's Weight	(0010,1030)	X		K	
Patient Comments	(0010,4000)	X			C
Patient ID	(0010,0020)	Z			
Patient Setup Photo Description	(300A,0794)	X			C
Patient Setup UID	(300A,0650)	U	K		
Patient State	(0038,0500)	X		C	C
Patient Transport Arrangements	(0040,1004)	X			
Patient Treatment Preparation Method Description	(300A,0792)	X			C
Patient Treatment Preparation Procedure Parameter Description	(300A,078E)	X			C

Table 2: De-identification of DICOM items

Following actions will be applied:

Z	replace with a zero-length value, or a non-zero length value that may be a dummy value and consistent with the VR
X	remove
K	keep (unchanged for non-sequence attributes, cleaned for sequences)
C	clean, that is replace with values of similar meaning known not to contain identifying information and consistent with the VR
U	replace with a non-zero length UID that is internally consistent within a set of Instances
X/Z	X unless Z is required to maintain IOD conformance (Type 3 versus Type 2)

Table 3: De-identification rules

If no other value is specified, C is treated as Z.

6. ETHICAL ASPECTS

To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables.

No personal data is collected as part of the i-Violin project. The participating partners are each locally, decentrally responsible for obtaining an ethics approval. In this context, it must be determined whether patient consent is required for the use of the anonymized image data within the framework of the project. The responsibility for this also lies with the respective partners.

Contact person for questions regarding this topic will be the scientific coordinator of i-Violin.

7. OTHER

Not applicable

8. CONCLUSION

For the i-Violin project, image data from three oncological entities will be used to test parameters for assessing image quality using newer methods. The results will be used to develop suggestions for optimizing examination protocols. For this, a dataset of about 600 CT examinations is needed. This will be supplemented by metadata collected using structured templates. Anonymized data will be used for data processing, which will be carried out in established systems and with teleradiological methods that are established in clinical routine and meet data protection requirements and IT security requirements.