



An EU-Canada joint infrastructure
for next-generation multi-Study Heart research

Deliverable D3.4

Guidelines and protocol for data deposition

Reference	D3.4_euCanSHare EMC_30112020
Lead Beneficiary	EMC
Author(s)	Marcel Koek
Dissemination level	Public
Type	Report
Official Delivery Date	30 November 2020
Date of validation by the WP Leader	26/11/2020
Date of validation by the Coordinator	27/11/2020
Signature of the Coordinator	





Version Log

Issue Date	Version	Involved	Comments
13/11/2020	1	Marcel Koek	1 st draft
17/11/2020	2	Marcel Koek	Updated 1 st draft
18/11/2020	3	Katharina Heil, Karim Lekadir	Review and feedback to 1 st draft
26/11/2020	4	Marcel Koek	Processed feedback and polishing
26/11/2020	4	Aad van der Lugt	Approval by WP leader
27/11/2020		Katharina Heil, Karim Lekadir	Final review and finalization of the deliverable
			Revised and corrected final version.

Executive Summary

This document describes the guidelines and protocols for medical data deposition for the Euro-BioImaging medical imaging archive and is referred to as Deliverable 3.4 titled: “Guidelines and protocol for data deposition”.

The Euro-BioImaging Medical imaging data deposition goes through 3 phases: Preparation, Execution and Completion. In the Preparation phase a data collection request is submitted at the support portal, users and access is provisioned and de-identification protocols are defined. In the Execution phase the data is uploaded, and the data organisation is managed. The dataset is curated and ready for interaction in the Completion phase.

The data is securely stored and transported to and from the Euro-BioImaging medical imaging Archive. Software and standard operating procedures have been defined and implemented to protect the privacy of the patient and ensure security of the data.



Table of Contents

1	Data deposition overview	4
1.1	Preparation phase	4
1.1.1	Request collection and user account	5
1.1.2	De-Identification Process Decisions	5
1.1.3	Install CTP-client at submitting sites	7
1.2	Execution phase	8
1.2.1	Uploading images	8
1.2.2	Manage study	9
1.3	Completion phase	10
1.3.1	Data Cleaning	10
1.3.2	Data extraction/interaction	10
1.3.3	Data storage	10
2	XNAT Data structure	11
2.1	General data structure	11
2.2	Experiment data structure	11
3	Euro-Biolmaging XNAT Security and Privacy	12

Acronyms

CTP = Clinical Trail Processor

PHI = Protected Health Information

DICOM = Digital Imaging and Communications in Medicine

Health-RI = Health Research Infrastructure (Dutch Research Infrastructure organisation)

TraIT = Translational IT (Project of Health-RI)

CT = Compute Tomography (Imagin)

MR = Magnetic Resonance (Imaging)

PET = (Positron Emission Tomography)

DEXA = Dual-Energy X-ray Absorptiometry

HIPAA = Health Insurance Portability and Accountability Act (US Law)

RSNA = Radiological Society of North America

VRE = Virtual Research Environment

REST API = Representational state transfer Application Programming Interface



1 Data deposition overview

Uploading data to the Euro-BiolImaging Archive is divided in 3 phases:

- the Preparation phase,
- Execution phase and
- Completion phase.

This section explains each phase in detail. The Euro-BiolImaging Archive is operated by the Health-RI TraIT platform¹. The Euro-BiolImaging Archive is running XNAT as archiving solution².

1.1 Preparation phase

In the preparation phase user accounts and collections are being created as well as de-identification of the data if needed. See Figure 1 for an overview of the workflow in the preparation phase. In the following sections each of the steps is described. The de-identification process focusses mainly on DICOM data, because that can contain protected health information (PHI) data. Derived data and converted image formats like nifti are discussed in the execution phase.

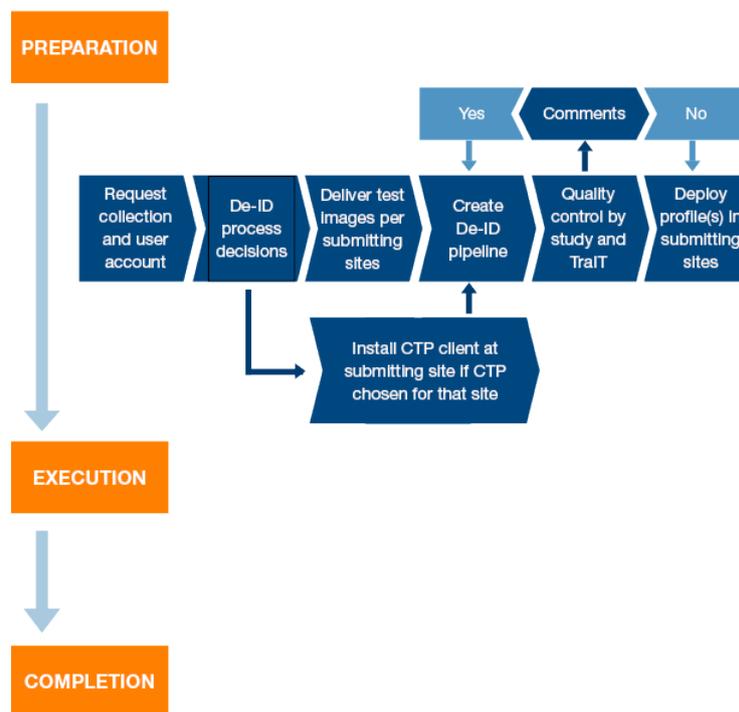


Figure 1: Overview of the preparation phase.

1_ <https://trait.health-ri.nl/>

2_ <https://trait.health-ri.nl/trait-tools/xnat>



1.1.1 Request collection and user account

To start using the Euro-BioImaging Archive for your study, images have to be uploaded to a new project. To do so, you first need to request a new collection and user accounts. In the euCanSHare project, the Euro-BioImaging partners can do this for the partners or a partner can request a Health-RI Service Desk account and request user accounts and collections themselves. The service desk is located at <https://trait.topdesk.net> and a service account can be requested via servicedesk@health-ri.nl.

Using the service desk account, a new collection can be requested by filling out the form: “[XNAT New Collection Request](#)”. This request also includes an account for the person who will manage the XNAT project on XNAT and receives the role of Collection Administrator by default. This collection administrator (or study coordinator) has also a coordinating role to prepare the study on the Euro-BioImaging Archive.

When requesting a new collection, the accessibility level of the data need to be indicated, the following choices are available:

- Public: anyone, also non-registered users, can browse and download the data, but cannot upload and delete data. Only registered users can change the project’s data, based on their access rights.
- Protected: the project (general information) is shown on the home page. Users have to request access to browse and change the project’s data.
- Private: the project is not shown on the home page but listed in “Other Projects”. Users have to request access to browse and change the project’s data.

The study coordinator will receive a variety of documents to guide the procedure, a.o. a [responsibility document](#) that shows, per step, who is responsible for its execution. Because most probably for each of the steps there will be different people delegated to execute it, we would request the coordinator to fill in the names of these contact persons (e.g. local CTP administrator) and send it back to Health-RI.

If you want to add more users in the study you will need to fill out either the “[XNAT New User Request](#)” (for a XNAT study) forms in the Health-RI Self Service Desk.

There are different user roles within XNAT. The roles at project level are:

- Owner: Full Access,
- Member: Full access except deleting,
- Collaborator: Only read access.

1.1.2 De-Identification Process Decisions

Not every research study and hospital is the same. Therefore, there are several decisions which need to be made upfront. These decisions can be split into organisational and technical decisions, as explained below.

Organisational decisions

These decisions are mostly influenced by the processes and collaboration efforts of participating centres in your study.



Collecting imaging data: upload directly or collect centrally before upload?

As a coordinating centre, the first question is whether you want to collect all data before uploading, or to have centres upload the data directly. The first option gives more control and overview for the coordinating centre. Whereas the second option is faster in terms of data transfer, however it introduces more work and responsibilities for participating centres. Our experience is that direct upload by participating centres is preferable, as data upload can be integrated in the participating centre's workflow. We are aware that installation of software in local hospitals is often cumbersome and not always feasible. Therefore, the option for de-centralised upload is often difficult or not possible.

Technical Decisions

These decisions are mostly influenced by the needs of the study regarding the information that is kept in the uploaded files after de-identification and can be guided by Euro-BioImaging.

Using one strict or multiple tailored de-identification protocol?

Although the DICOM standard defines the structure and placeholders (tags) for metadata, it also supports additional (non-standardised) additional information. This non-standardised information can be different for different manufacturers, modalities (e.g. CT/MR/PET scanners), or even software versions of the same modalities and could potentially contain private information.

To be sure that XNAT does not contain patient-related information (retrieved from the DICOM metadata), Euro-BioImaging can support the creation of de-identification pipelines. In general, Euro-BioImaging provides a generic (strict) anonymisation pipeline (compliant to DICOM supplement 142, following US HIPAA laws) which should be applicable to every study. The downside of this strict pipeline is possibly the loss of detailed information specifically needed for certain research questions and/or data management. Many patient-related information (e.g. weight) is missing, which limits applicability for research questions. For example:

- a) Computation of PET Standard Uptake Values (SUV) or DEXA or MRI measures of body composition, which are based on body weight, body surface area or lean body mass.
- b) The data stored in private tags by some vendors are required for specific image analysis. E.g. for PET-Scans from Philips two private tags needed to be preserved for a particular study.

To overcome these problems, we support tailoring of anonymisation protocols, however this means that every submitting site must have its own, tailored anonymisation protocol. This propagates into the next steps, as several tasks should be done for every participating centre (e.g. defining de-identification protocols).

As a standard, Euro-BioImaging requests studies to follow a standard de-identification policy. It is up to the study to define variations to this policy and justify these choices.



The steps to define and implement the de-identification protocols are described in Table 1.

Table 1: Steps for defining and implementing the de-identification protocols.

Steps	Short description	Actor	Long description
a)	Define patient id replacement	Study coordinator	Define a CTP look-up table for patient id replacement
b)	Create CTP collection de-identification template	Euro-Biolmaging	Create CTP collection de-identification template based on the approved Collection De-identification protocol
c)	Implement patient id replacement	Submitting site	Implement patient id replacement in CTP
d)	Implement CTP collection de-identification template	Submitting site	Implement CTP collection de-identification template based on the provided CTP collection de-identification template
e)	Clean burned pixel data	Submitting site	
f)	(Random) visual inspection	Submitting site	(Random) visual inspection of tags for each submitting site

If the strict pipeline does not comply with the needs of the study, the following also needs to be done:

g)	Define collection de-identification profile	Study coordinator	Define deviations from DICOM-142 supplement
h)	Approve collection de-identification profile	Principal Investigator	Approval of collection de-identification profile
	Repeat steps a-f		

1.1.3 Install CTP-client at submitting sites

The Clinical Trial Processor (CTP) is the proposed tool for de-identifying and submitting imaging data. Every submitting site needs to install CTP locally to submit images to the Archive.

CTP is a stand-alone Java program (developed by the Radiological Society of North America, RSNA) for de-identification and submission of imaging data to a central archive. CTP has the following key features:

- Easy installation.
- Support for multiple data pipelines.
- Processing pipelines supporting multiple configurable stages.



- Support for multiple quarantines for data objects which are rejected during processing.
- Pre-defined implementations for key components:
 - HTTP / DICOM Import.
 - DICOM Anonymizer.
 - HTTP(s) / DICOM / FTP(s) Export.
- Web-based monitoring of the application's status, including configuration, logs, quarantines, status.

More information about CTP is available from the [RSNA](#).

A couple of configuration templates are available on [this GitHub repository](#), namely a strict anonymiser according to DICOM 142 standards, and a customised profile.

CTP needs to be installed at each site wishing to submit images to XNAT. Images are then sent to a central CTP server, which then directs them to XNAT.

1.2 Execution phase

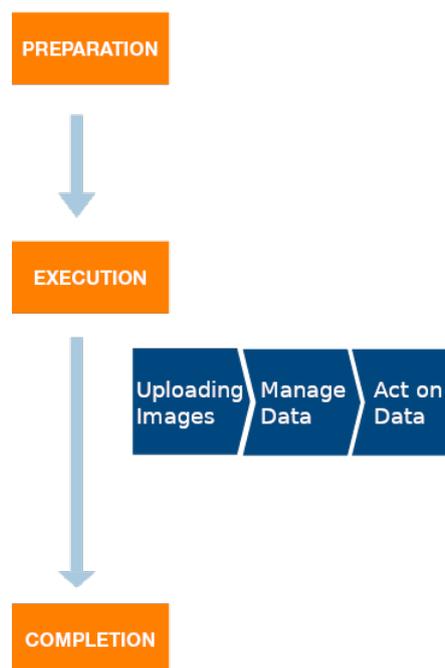


Figure 2: Overview of the execution phase.

The execution phase is the active runtime of the project. In this phase the data is uploaded, managed and acted upon. See an overview of the execution in Figure 2.

1.2.1 Uploading images

During this step the images are uploaded by the participating sites in the XNAT Project

DICOM Images should be uploaded via the CTP software installed in the participating site, which is managed by local IT of the participating site. All other types of images or derived data can be uploaded via the XNAT REST API.

Section 2 describes the data organization in XNAT.

Upload via DICOM

For uploading data to XNAT via the DICOM protocol, the Clinical Trial Processor (CTP) should be used. CTP anonymizes the imaging data and sends it to XNAT. Figure 3 depicts the architecture of the interface to XNAT. The imaging data is sent from the local PACS, other workstation, PC or laptop to the hospital CTP. This client CTP automatically sends the data to the server-side Health-RI CTP, which then sends it to XNAT. Data that is sent from the client CTP to the Health-RI CTP uses an encrypted (https) connection. It is the responsibility of the CTP administrator to configure CTP such that all imaging data is sent anonymously to Health-RI CTP.

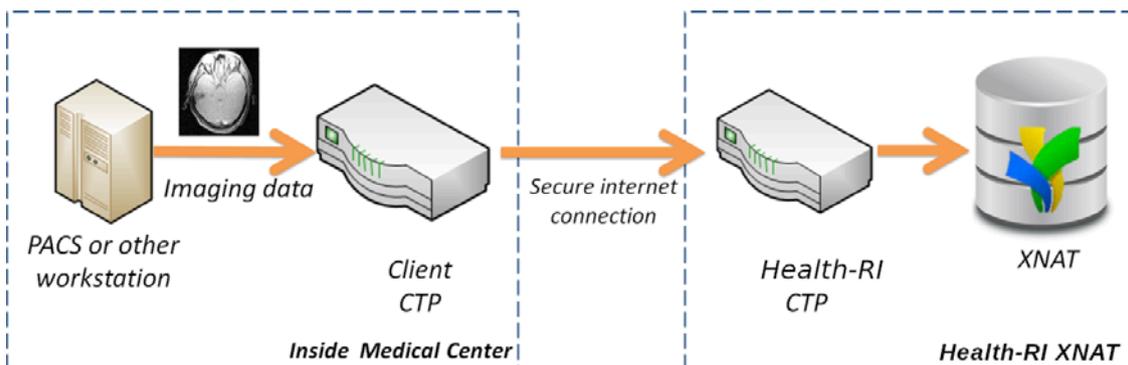


Figure 3: Dicom upload data flow.

Upload via XNAT REST API

Data, other than DICOM, can be uploaded using the REST API of XNAT. In order to do this, some expertise in programming and API's is required. The XNAT REST API is documented here: <https://wiki.xnat.org/documentation/the-xnat-api>.

It is recommended to use the Python client library [xnatpy](#) to interact with XNAT through the REST API. The documentation for xnatpy can be found here: <https://xnat.readthedocs.io>. By using this library, the interaction with XNAT can be integrated in a programmatic/automated environment more fluently. For instance, for connecting the euCanShare VRE.

1.2.2 Manage study

The study coordinator is responsible for the management of the XNAT project during the execution phase.

The tasks for this person include amongst others the following:

- Monitor the usability of the uploaded images for the envisaged research purposes.
- Modify the de-identification profile when required.

To avoid late stage problems, it is very important to test changes of an ongoing study (e.g. modifications of the de-identification profile). When testing is neglected, it could cause serious problems after years of data collection!

XNAT allows for more than just archiving, viewing and downloading of images. Through the REST API additional functionality can be implemented allowing for more advanced analysis of the data, as well as adding additional meta-data to your project.



1.3 Completion phase

In the completion phase the data can be cleaned/curated, extracted for downstream use and archived when the project ends.

1.3.1 Data Cleaning

Data cleaning is the process of detecting and removing corrupt or inaccurate images.

Quality control should begin before the actual images are uploaded and continue until the end of the data collection process. Images in XNAT are not editable. They can only be removed by the collection administrator or flagged as "questionable" or "unusable for analysis".

1.3.2 Data extraction/interaction

Data can be downloaded via the XNAT web interface, the XNAT desktop client and through the REST API.

Interaction with the data is made possible through a REST API. Applications like the euCanShare Virtual Research Environment, External Image Viewers and High Performance Computing environments can be connected to the XNAT in order to retrieve data and upload data back to XNAT. Not only image data, but also meta-data can be uploaded to XNAT.

1.3.3 Data storage

When data is no longer being collected in XNAT and all outstanding queries have been resolved, then direct access to the study data is either no longer required or limited to occasional read only access.

All study data will remain on the XNAT server and therefore there is no need for archival to disc or to another system. This could nevertheless be considered.

In the unlikely event that this policy is changed, it will be guaranteed that study data are provided to the PI of the study.

2 XNAT Data structure

2.1 General data structure

The XNAT data model is a hierarchy consisting of the following layers (see Figure 4 for a visual representation):

- 1) Project: a project for a specific research project.
- 2) Subject: a patient.
- 3) Session (also known as Experiment): a set of scans for a specific modality (CT/MR/PET).
- 4) Scan: a set of images performed during a session on a patient for a project.

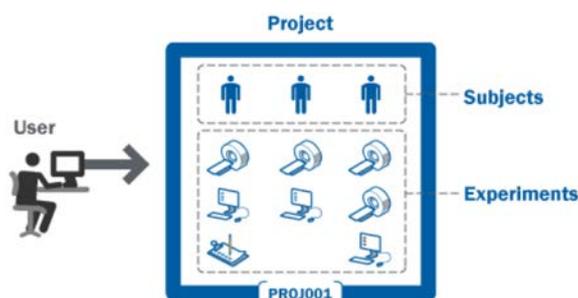
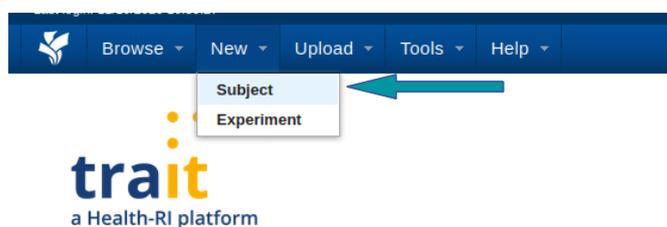


Figure 4: XNAT general data structure.

A new subject and session can be created using the top menu on the XNAT website → “New” (see Figure 5). For imaging data that are uploaded via the website; the project and subject should be created in advance. For data uploaded via CTP, only the project has to be created in advance. The subject and session will be created during archiving.



BMIAXNAT currently contains 218 projects, 18910 subjects, and 53925 imaging sessions.

Figure 5: Creating a new subject on XNAT

2.2 Experiment data structure

Beside a scan, three other types of material can be uploaded:

- Reconstruction: for derived files (images only) based on the scans e.g. an average.
- Assessment: for derived files based on analysis e.g. logs, spreadsheets.
- (Additional) Resources: any other data e.g. small video's, documents.

These additional types, further described on the [XNAT wiki](https://wiki.xnat.org/display/XNAT17/Strategies+for+XNAT+Image+Data+Storage)³, can be uploaded via the REST API.

3 <https://wiki.xnat.org/display/XNAT17/Strategies+for+XNAT+Image+Data+Storage>



3 Euro-BioImaging XNAT Security and Privacy

The Euro-BioImaging XNAT is operated by Health-RI TraIT. Euro-BioImaging operators are part of the Health-RI TraIT staff for the maintenance and support of the XNAT service.

The protection of the privacy and security of the data are of paramount importance for Health-RI TraIT. The Health-RI TraIT infrastructure, software and procedures have been defined and implemented to protect the privacy of the patient and ensure security of the data.

Health-RI TraIT has taken a number of security measures (see [Technical and organisational control measures⁴](#)).

Health-RI TraIT uses Standard Operating Procedures (SOPs) for creating new user accounts in order to comply to the data owner's data access policies and to prevent unauthorized access to study data.

The Health-RI TraIT XNAT environment is hosted by Vancis. Vancis meets the ISO 27001 standard for information security management. The network communication with the TraIT XNAT web-application (incl. the upload process) is secured via the HTTPS protocol, meaning that hospitals do not have to open ports on their firewall (assuming that the basic HTTPS port is open to the outside world). Our certificate guarantees that the Health-RI TraIT XNAT user is really communicating with the Health-RI TraIT XNAT environment. Our infrastructure architecture has been optimized for data security. The XNAT database is backed up every night and every other year we perform a disaster recovery test.

Health-RI TraIT takes security seriously at the application level. It works with controlled access through user accounts and authorization at the study and role level.

XNAT in turn has similar authorization options. For each study (project in XNAT), the study owner decides which users to add and with which role. For more information, please see: <https://wiki.xnat.org/display/XNAT17/Understanding+User+Roles+and+Permissions>.

Furthermore, the visibility of a project/collection can be set as:

- Public: all project visible to outside world.
- Protected: project description visible to other users, but image (derived) data not. Users can request access.
- Private: a user should be a member of the project to see it after login.

By default, a project is created with private visibility. The study owner can change this at the creation of the project, or at a later moment (see Figure 6).

The data owner (data controller), generally the principal investigator of the research project, decides who gets access to the study and indicates the role/authorization that should be applied to the user account. The institute or investigator that supplies the data to Health-RI TraIT remains the owner of the data and access to the data is controlled by the owner. In order to comply to privacy regulations, it is only allowed to store anonymized or pseudonymized data on the Health-RI TraIT XNAT server.

Security at transport level, especially with regards to data collection/ingestion and data consumption, is covered by using standard ciphering protocols to provide encrypted communications over the Internet.

4 <https://trait.health-ri.nl/about-trait/scope>



The screenshot shows the XNAT web interface for data access management. The browser address bar shows the URL: `https://xnat.bmia.nl/app/template/XDATScreen_report_xnat_projectData.v`. The interface has tabs for 'Details', 'Access', 'Manage', and 'Pipelines', with 'Access' selected. A 'Show Deactivated Users' checkbox is present. Below it is a table with columns for Username, First Name, Last Name, Email, Group, and Remove. Filter buttons are provided for each column. The table contains four rows of data, with the first three rows having 'Owners' as the group and the last row having 'Members'. To the right of the table is an 'Actions' menu with options: Add, Upload Images, View Prearchive, Scan Type Cleanup, Add to Favorites, Download XML, Download Images, and Manage Files. Below the table are sections for 'Add/Invite User' (with a text input, dropdown, and 'Add User' button) and 'Project Access Settings' (with 'Manage Groups' and 'Manage Features' buttons). The 'Define Default Project Accessibility' section has three radio buttons: 'Private' (unselected), 'Protected' (selected), and 'Public' (unselected). Each option has a description of its access level. A 'Set Accessibility' button is at the bottom of this section. The 'Project Invitations' section contains a table with columns: Date Created, Invited User, Project Role, and Status. It lists three invitations from 2016 and 2017, all with the role of 'owner'.

Username	First Name	Last Name	Email	Group	Remove
[Redacted]	[Redacted]	[Redacted]	[Redacted]	Owners	X
[Redacted]	[Redacted]	[Redacted]	[Redacted]	Owners	X
[Redacted]	[Redacted]	[Redacted]	[Redacted]	Owners	X
[Redacted]	[Redacted]	[Redacted]	[Redacted]	Members	X

Date Created	Invited User	Project Role	Status
10/5/2017, 11:48:56 AM	[Redacted]	owner	
10/4/2017, 5:35:14 PM	[Redacted]	owner	
5/12/2016, 9:27:44 AM	[Redacted]	owner	

Figure 6: Data access management in XNAT.



List of tables

Index of Tables

Table 1: Steps for defining and implementing the de-identification protocols..... 7

List of figures

Table of Figures

Figure 1: Overview of the preparation phase 4

Figure 2: Overview of the execution phase 8

Figure 3: Dicom upload data flow 9

Figure 4: XNAT general data structure..... 11

Figure 5: Creating a new subject on XNAT 11

Figure 6: Data access management in XNAT..... 13