# MMIV Research PACS/RIS Design

Hauke Bartsch, Dr. rer. nat.

# Target workflows

**November 25, 2019**

## Overview

In order to structure a research PACS[1] it is useful to list target workflows that the system is expected to support. Actually "research PACS" is only half of the system we will need. The other half is a research RIS[2] system that contains the project information not directly related to imaging data. Together these two systems will allow for consistent data tracking.

Examples for PACS data include DICOM image data, overlays. Examples for project information are information on the project design (longitudinal / cross-sectional / interventional / observational), event definition and the associated collection of assessments. At a minimum this contains information on the arm of the study for each participant and event (control or disease group) and general demographic information relevant to the interpretation of the image data such as sex at birth, age and  handedness.

The following is a list of workflows that the research PACS/RIS (i.e. rDMA) is expected to support at Helse-Vest. This list was compiled from the user requirement gathering process at the startup of the project. The current implementation and list of issues is available in the "PACS project structure" document.

## Automated import and export of study data

Images can be acquired on hospital systems at some point in the past or going forward, or images can be imported from outside vendors. Each time we want to use a consistent tracking of image data in the system with a consistent assignment of data to research projects. This

[1] Picture Archive and Communication System. The whole system will be called rDMA.
[2] Radiology Information System

requires us to import from a number of DICOM nodes in the hospital as well as providing web-based import facilities to allow users to import external USB/DVD media data.

## Reuse of image data in multiple projects

Image data obtained from control subjects is frequently used in more than one project. Also, projects might continue and extend the work of other projects sharing participants. In order to obtain the true number of participants and to allow a project specific removal of image data (at the end of the project) we need to be able to assign the same image data to more than one project. This might involve the creation of project specific identifiers (see project structure document) but should not involve the creation of new participant IDs such as participant ID and participant name.

## GDPR support

The general data protection regulations (GDPR) contain two relevant provisions that the rDMA will support. The right to be informed and the right to be forgotten. Both of these rights are supported by the electronic data capture system that is playing the role of the RIS in the research PACS.

## Study consent tracking

A participant might retract consent. In this case all DICOM studies that contain data from this participant need to be removed. This includes images and metadata. We will not remove data from backups! StudyInstanceUIDs of these participants will be stored in a banned-list and the system will refuse to import such studies in the future.

## Export packaged version of project data

A packaged version of the project data consists of all the projects image data in a directory structure with a spreadsheet that describes the data. There are two options for this export that need to be supported

1. **Export as de-identified package:** In this case all project data (DICOM and spreadsheet) shall contain the identifiers used in the research PACS. Usually those id's are de-identified IDs for which a coupling exists outside the system.
2. **Export as anonymized package:** In this case all project data (DICOM and spreadsheet) shall contain newly created anonymized IDs.

## Support reproducible research in image sampling

A task frequently done on clinical and research image data is the extraction of (semi-) quantitative measures from regions of interest in the data. This protocol for this data extraction is as far as possible automated with usually at least two measurements of diseased and healthy tissue. The process involves the placement of a target point on one of the images. At a given distance around the indicated point the image intensity is measured as a) minimum intensity in the region, b) maximum intensity in the area, c) average intensity in the area, d) standard deviation of the image intensity in the area, and e) the area of the region of sampling. A predefined area is expected to be measured.

## Allow for direct statistical analysis and machine learning

Data represented in the Research PACS and research RIS should be amenable to statistical methods like classification and regression analysis. This includes the possibility to filter the data, to impute missing data, to sample training and test sets, to learn a model, and to use the model to predict on novel data.

## Study level export of derived data

Data generated from the images in the PACS by automated or manual segmentation steps need to be exported as structured information on the level of study. As a study might contain longitudinal data for many participants the extraction and export of such data should be automated and result in a standard file format such as a CSV at arbitrary points during the study.

# Anonymization of burned-in image information

Some modalities have frequent burned-in image information (US). Usually this appears in secondary capture images. Any personally identifying information (PII) should be removed from these images by changing the pixel data in rectangular regions. Overlay elements are not sufficient to hide this information due to the flexibility of the display programs and the anonymized export functionality of rDMA.