

How can we help you?

How do I start using the system?

Creating a project for your data is of course the first step. One the frontpage of Steve use the link at the top to [apply for a new research project](#). After you got access from IKT to the "Sectra DMA Forskning" start menu link you can login there and see your empty project. Start by uploading data to your project following the steps in [How to send data](#). Information about these first steps are available in the [EK handbook](#) (see Forskning / Forskningsprosedyrer, 02.20.7.1 Forsknings PACS).

Where does the data come from?

Both clinical and research systems are provided as services inside the hospital system. Whereas the clinical system supports the day-to-day workflows for patient care its sister system for research provides data services on a research cohort level. For imaging data both systems receive data directly from clinical scanners and enrollment into research projects is used by the scanner operators to decide if data is send to both systems or to the research PACS only. Imaging data may also be imported from external media. Non-imaging data is captured in the research system using electronic data capture (EDC) in REDCap. Both the imaging system and the EDC secure access on the project level, provide anonymization procedures and access to the data using role based accounts. They support automated workflows for data analysis and data processing as well as data exchange with third parties.

Best practices for project setup

These are not rules, they are more like guidelines. They do may make the difference between an *ok project* and a project that is nice to work with (see FAIR data use).

- A research PACS project is more than a copy of all participant data from the clinical systems. Only transfer data explicitly covered in your REK approval - this is actually a rule, not just a recommendation. Patients might be in the hospital and receive imaging appointments for a number of different purposes. Image studies not directly related to your research project should not be transferred.
- Limit the number of coupling lists to identify participants in your project. In the best case all project members should use a single pseudonymized (numeric) identifier for each participant linking imaging data with diagnostic information. A single coupling list of participant identifying information and pseudonymized identifier is optimal as it still ensures separate storage of sensitive information from data.
- Numerical identifiers for participant ids should use leading zeros ("project_001" instead of "project_1"). This allows for a consistent alphabetic sorting of participants in the research PACS Information window. The number of leading zeroes can be derived from the maximum number of participants in the study.
- Utilize non-numeric event names if your study is longitudinal. If you assign all image data to a single dummy event you will have more work later to specify baseline assessments needed for analysis (compute values relative to the baseline assessment etc.). If your project has an open number of events a two-event setup with "baseline" for the earliest good quality DICOM study and "followup" for all other DICOM studies is ok to use. All event-based studies should assign timing-based event names like "pre-op", "post-op", "6month", etc.. The event name is visible in the research PACS if you add the "Referring physician" column to the Information window.
- In order to support clinical studies a basic REDCap project (using RIS setup) contains five data collection instruments.
 - *Basic Demography Form*: The entries in this form are used to link to the pseudonymized participant ID. All three fields usually contain the same value that is linked to the image information for PatientName and PatientID.
 - *Imaging*: The imaging instrument is automatically populated by FIONA after each data transfer into the research PACS. The basic information captured is the study instance UID, event name, (shifted) study date and the study description.
 - *Pathology*: The pathology instrument adds to the imaging instrument measures related to pathology imaging such as magnification factors, resolution and stain information based SNOMED-CT.
 - *Adverse Events, Monitoring, Record Locking, Source Data Verification*: This instrument captures information required for clinical study type data capture. For each participant in the study all found adverse events (AE), serious adverse events (SAE) and suspected unexpected serious adverse reaction (SUSAR) are captured. The instrument includes also a section on medication monitoring, documentation for record locking and a section to document a source data verification step. Not all projects, especially non-clinical drug trials will need all of these fields. Adjust the instrument for your own study as needed.
 - *e-Consent*: The template for electronic consent shows the use of signature fields to authenticate both the consentor and the consentee. Notice that HTML formatting for e-consent will be removed in the resulting PDF documenting the consent process (restriction of REDCap). Use the section headers as shown in the template file to obtain a better structured PDF version of the consent. Use images and the logo to style your consent.

Adjust instruments that you find useful in your study. Remove any instrument that you do not need.

Can data be deleted?

You will not have permission to delete data yourself - but you can request data to be deleted from the system. Send an email to Hauke with the project name and detailed information of which participant, event, study, and series should be removed. With the same workflow you may request replacement of participants for which the wrong image series where submitted.

Can I use the research information system without an ethical approval (REK number)?

We do accept projects without REK that are for *operational support* like scan quality control projects. As always, the project owner is responsible to ensure that such a project follows all institutional guidelines. Operational support projects need to agree to the same pseudonymization procedures as other projects.

How to handle participant data after removal of participant consent?

Participants can retract their consent to be part of a running research study at any time. One option for such data is to request a removal of the image data (send email to rDMA team/Hauke). If the data was already part of published research you as the researcher might also have an obligation to store the data in case your findings need to be verified at some point in the future. Not using data in future research and allowing for a later verification of already performed research can be difficult to implement. We suggest in this case that you use one of two approaches. i) Export the raw data that is part of your paper and store an offline copy together with your analysis scripts for any future questions that you might have to respond to. Request data where consent has been retracted to be deleted from the research PACS. All remaining data in the research PACS is therefore ok to include in the next paper. Or, ii) you can use the worklist functionality of IDS7 to create a new worklist ("Ny statistik arbetsliste") of subsets of participants. We suggest in this case that you work with three worklists, one to track participants that have removed their content - such data remains on the system but such participant data should not be used for future studies. One worklist per publication that contains references to the imaging studies that have been used. And one master worklist with participants that are ok to use in future papers by your project.

What happens at the end of the project?

The end date of a project is specified in the REK approval. We are using this information to inform you between 3 and 6 month before the end of the project. At this point you can request an extension of the project from REK. If such an extension has not been obtained the project data remains on the research PACS but access to the project will be removed by removal of the project role. The data will no longer be visible to you.

Can we send out emails to people at home?

Yes, but there are some limitations on this functionality on our REDCap on Azure system. Emails are routed through a Microsoft Exchange custom domain which limits outgoing emails from one system to at most 200 per day (30 emails per minute, 100 emails per hour). That limit is shared for all projects on REDCap. To not interfere with other projects we suggest to use a lower limit of 100 emails per day. Contact us if you need to send out more emails per day.

How to integrate with external vendors?

An external vendor might be a company that performs image analysis for you. This can be done in two basic ways - sending images to the cloud (difficult because of loss of control over data) and installing the vendor software inhouse (much easier). The process to integrate such an external vendor into the research information system includes a number of steps. Namely:

- Check against existing systems
- Budget control
- Risk assessment
- Data processing agreement
- Contractual agreements
- Data protection impact assessment

Whereas some of these steps are mandatory most are dependent on the type of integration and prior work. A working integration will allow you as a researcher to control the sending of images from the research PACS to the vendor software. The software will perform its task and any resulting images will appear back in your project in the research PACS.

How anonymous is the data in the research information system?

As copies of the image data may exist in clinical systems, research image data is considered at least indirectly identifiable personal data. Data exported from the research PACS may retain that property and should be stored on secure systems. According to GDPR this may make it necessary to carry out a Data protection impact assessment (DPIA) prior to processing.

How secure is the data, can I test that?

No. With access to this system you agree to use best-practices for data handling. This specifically includes that you value the study participants and patients right for privacy and the security of their sensitive data. It is not allowed to re-identify or attempt to re-identify participants in your project or in other projects. It is not allowed to reveal or attempt to reveal other projects data. It is not allowed to use the system in an unsupported way in order to test or trick the system.

Do you change the data in any way?

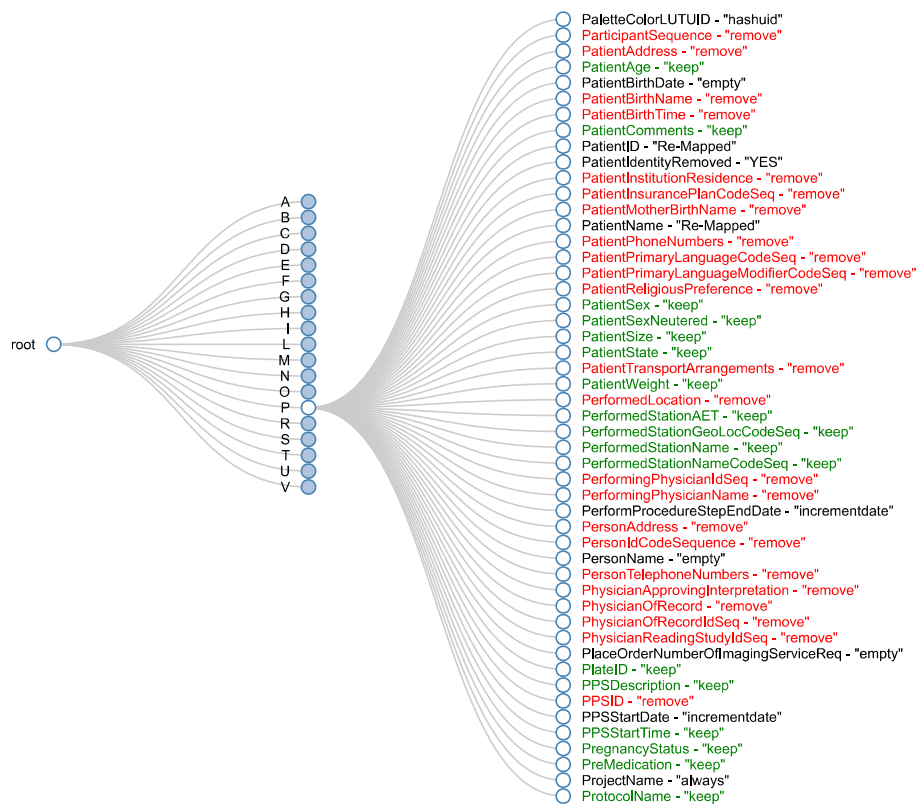
Yes. With input from the project we attempt to anonymize all data forwarded into the project space. This includes changes to the meta-data section and changes to burned in image information of some of the incoming data (secondary captures). These data processing steps are implemented to ensure an anonymization of the data with respect to the Steve system and a pseudonymization of the data towards the project as they may retain a coupling list.

Why are all the study dates wrong?

The study date is one of the easiest to obtain information in order to link imaging studies between the clinical and the research PACS. This re-identification of participants is discouraged for anyone who is not in possession of the projects coupling list. Accurate timing information of imaging studies may also be required to analyze image data. In order to serve both the need to keep study participant information private and the need to allow for good science we opted to *shift* data collection dates in a consistent way per project. Relative timing between imaging events is as accurate as in the clinical PACS. It needs to be stressed that this only prevents a direct path to re-identification. Data export using FIONA's "Export" application can be used to undo study date pseudonymization for data sharing that requires correct dates.

Is there a list of DICOM tags changed during import?

Yes, a list of about 270 tags inspected during import is available as part of the source code of the anonymization tool [DICOMAnonymizer](#). Below a tree-view of all the touched DICOM tags sorted by first letter. Tags listed with "remove" are deleted, tags listed with "keep" are kept etc..



Tags not listed above are untouched by the pseudonymization tool.

Can I export to TSD/Safe?

TSD supports data upload links. This API is expected by our system to allow a direct submission of data folders (zip-format) to your TSD storage space. This feature has to be setup for your projects, contact us to receive more information. There is no comparable technology for Safe yet. Contact Christine Stansberg to request such an interface.

The following information from your TSD project on (<https://data.tsd.usit.no/i/>) are required:

- TSD group name:
- TSD ID: e0b0c0e-abcd-abcd-abcd-a0b0c0d0e0f0 (example)
- TSD user name:

Can I export to clinical PACS?

Yes, export to clinical PACS is possible using "NoAssign". Mostly this option allows pseudonymized data to be forwarded to other institutions using clinical PACS to PACS features like OneConnect.

In order to send to clinical PACS use the NoAssign application of fiona. You may need "Export" permissions for your project to use this application. The application will list all studies currently found in quarantine on fiona. Specify the project, participant, event information and the workflow type "FIONA anonymization". Select the examination you want to forward and "Export...". A dialog "Are you sure?" will allow you to select a destination in the final step. Both "CDRobot" and "clinical PACS" are supported destinations.

Note: Additionally to the standard pseudonymization done by fiona files will have a fake Date of Birth (0010,0030) DICOM attribute value of "19000101". This may be required if receiving PACS systems expect valid clinical data. By default the value of this attribute is empty inside research PACS. Only exporting data using NoAssign will add the dummy value.

PACS to PACS connectivity: If images pseudonymized on FIONA are forwarded to another PACS inform them on how to find your pseudonymized images. Tell them:

- The AccessionNumber (Undersøkelse-ID) DICOM tag will start with the letters "FIONA" followed by some random letters and numbers.
- The PatientName and PatientID tags will be the same (entered on FIONA, can be something like <project>_<numeric_id>, e.g. "TOBE_0022").
- The ReferringPhysician DICOM tag will contain the name of the imaging event (e.g. "Eventname:baseline").
- Further information on the pseudonymization procedure can be found here: <https://github.com/mmiv-center/DICOMAnonymizer>.

What other types of data can you store in PACS?

Our PACS system can store image data from radiology, cardiology, urology, oncology (DICOM) and pathology (whole-slide image format). Other types of files can be embedded into DICOM and stored that way. For example, the Siemens Spectroscopy (DICOM) format (.ima files) can be stored and exported again. These files can be read successfully by spectroscopy software packages like [OXSA](#). The Siemens TWIX format (.dat, .rda) are not suitable for PACS storage, use the .ima format instead.

Some of the spectroscopy DICOM files are *non-image* files. PACS viewers might not show them in the interface. In order to verify that they are stored correctly (other than downloading them again using Export) the FIONA system will add a secondary capture image that lists the hidden non-image objects including their size and series description.

The generation of the secondary capture image is currently limited to Siemens non-image files (SOPClassUID = 1.3.12.2.1107.5.9.1). Contact your FIONA team if you want to include other files.

Why the name "Steve Project"?

So that you are less afraid of adopting a new workflow. The "Over the Hedge" movie from 2006 had this scary hedge, everyone was afraid of it, it was new and looked big and scary. They suggested to call the hedge "Steve" and it was not so scary anymore. For the same reason our research information system portal is also called Steve - its a nice name and makes the system much less scary to use.

Research information system

We have created a research information system in response to common issues faced in integrating research algorithms into clinical practise. We started with a system that required many people to work together to provide access to research data, which does not sound like a bad thing, research is based on good cooperation between many people with diverse backgrounds. Looking at the type of things that needed to happen you realize that highly skilled hospital staff hand-carried a bag filled with 80 individual DVDs from one hospital area to another. Those DVDs each contained individually de-identified radiological images exported from an MRI machine where such a process may take up to 10 minutes per disk.

Based on these experiences we realized that many research tasks required for the successful running of a medical research study like data identification, data export, de-identification are not well supported if research institutions are setup as external entities to the health-care enterprise they are supposed to benefit.

The purpose of the Steve Project is to create a research information system that no-one is afraid of using and that provides an interface between hospital procedures that generate data and research institutions that consume them. The system focuses on supporting two aspects of medical data - all the lab samples, questionnaires, diagnosis reports and clinical history and the medical image data in the form of DICOM images.

Safety first - Separation of hospital and research

Our research information system is independent from the clinical systems at our hospitals. It is setup as a shadow system that connects to all hospital infra-structure and that has the overall shape and appearance of the clinical system but it is specifically geared to serve the needs of research projects.

How similar are the hospital and the research systems? Both hospital and research system use the same user accounts and permission services (active directory). This allows us to provide access to our research services with the same user-names and passwords as for the clinical system. Both the hospital and the research system use the same version of a vendor neutral archive and image viewing software (PACS). Whereas the instances of the clinical and the research system are separate and data storage is independent features of the clinical system like modality specific hanging protocols, image annotation tools and keyboard shortcuts are shared. This provides access to commercial image viewing software to researchers for data inspection and quality control which is essential for machine learning projects. For their clinical partners it provides a familiar interface to rate the products of research algorithms. Most importantly it removes the gap between the quality of data generated by research tools and the quality and level of automation that needs to be provided if they want to be evaluated for clinically use. This allows researchers to act as solution providers towards the hospital without the need for the integration of the research tools in commercial software applications first. Such a safe solution for the clinically relevant accelerated evaluation of novel solutions can help to understand the limitations of novel systems and limit the risks involved in the development of commercial solution.

Safety first - Separation of research projects

In a clinical setting a health region will share a single clinical system which helps limiting the costs of such systems. Each hospital will be setup to see parts of the data such as all information from the hospital itself but not necessarily the information from patients at other hospitals. Often this is not a true separation but it is enforced by individual worklists and role based permissions. A general patient search at one institution will still turn up patients scans at the connected hospitals in the health region. Whereas this is a feature for a clinical system a research information system needs to be more restrictive as access to data is more restricted by regional institutional review boards that allow for the use of research data in approved projects only.

Our system uses project access restrictions to provide a full separation of project data from each other. This includes project specific data identifiers in the VNA that allow project data to be used and deleted without interfering with other projects that might use the same patients data. Only users of the research PACS that are part of the projects role will have access to list the data and to see the assessments and images.

Safety first - Moving data between hospital and research

Data is transferred only from the hospital to the research information system, not the other way around. This limitation is not technical but operational. Only personnel with access to clinical data can forward such data to be added to a project in the research system. Such data transfers are possible from the clinical PACS as well as from modality systems at all Helse Vest hospitals that collect data as part of research studies.

We decided to store all image data and assessment data in a de-identified form before any analysis. As data is often collected by a research study as part of a general clinical workflow those data are labelled with patient identifying information. This includes names and identification numbers as well as sensitive information on where and when the data was collected. Such information is part of the medical file formats meta-data, burned into the pixel information in files and part of spreadsheets used to track the data.

The data de-identification is implemented as an automated process that connects the clinical systems of a hospital with the research system using an edge-device called FIONA. This system acts as an intermediary translation service that maps sensitive data to de-identified data in the research information system. Input required by the translation service is only a mapping of the clinical data to the particular project, the de-identified patient identifier and the event the data belongs to.

A system aware of study design

Whereas clinical system structures each patient individually in a research system participants are grouped on more levels for example by *project*, *research arm*, *assessment event*, *imaging study*, *image series*, and *individual image*. Such complex hierarchies allow for group level analysis of hundreds and thousands of participant data simultaneously without the need to individually export and handle the data. Especially the introduction of an associated event name to collected data allows for many features of statistical analysis. Our system includes these classifications in a central location for both the assessment as well as the image data. This limits the amounts of decision and assumptions that have to be made by various researchers in the structure of their analysis.

Study management and study tracking

Our research information system allows all participant data to be forwarded into the system at the time of data collection as part of a clinical study. There is no time-lag between when data is collected at a scanner and the time that image data is accessible in the research system. If technicians that collect the data are aware of the need to transfer the data into a specific project they will forward the data once to the clinical system - if that is required for safety reads - and a second time into the research systems edge device called FIONA. In a second step they need to assign the project, de-identified patient id and event name to the forwarded data. Such an identification step is the only requirement to map clinical data to the complex structure of research projects. After minutes the data becomes accessible to the research project in a proper de-identified manner.

Data translations from clinical to research system

The list of meta-data tags that are removed by the FIONA edge system when data is assigned to a project is long. We document which tags are changed by making the source code and the process publicly accessible at our [DICOMAnonymizer](#) page. We hope that this will improve the quality and security of our solution by allowing other groups to evaluate the software in their settings. This includes for example the need to evaluate the de-identified images generated by the software with other image distribution and viewing solutions and the need to test the data interpretation of the tool with new DICOM image files.

Some image data generated in the hospital setting will include textural information about the patient as burned in pixel information. This is common in some modalities such as ultra-sound but also appears in secondary capture images generated by specialized workstations applications for perfusion, diffusion and molecular imaging. Our research information system detects such images and attempts to automatically remove burned in image information by overwriting the detected areas with rectangles of a uniform color. Due to the fully automated process secondary capture images can be forwarded into the research PACS and are safe to use after a review by the project.

Specialized applications

To provide access to the feature of the research PACS we provide web-applications for data submission, project setup and configuration, review and data export. All of these features are accessible on the home page of the Steve project page at the institution. Based on your role you will need to use only some of these applications.

Assign

The Assign application is the entry page for project data. The application lists incoming data that is in quarantine and allows the user to select the appropriate project, de-identified participant name and the event name of the imaging study. This is sufficient for a manual assignment of captured data as it is acquired in the hospital setting. For legacy data and external data in large quantities several automated import strategies are available. If data is de-identified outside of the research information system by writing a new patient ID such files are recognized by the edge system using either the send destination (AETitle of the addressed service on the edge FIONA) or by the pattern used in the patient ID. This detection of incoming data is used to detect the destination research project and trigger the de-identification step without another manual identification step. Additionally to such automated data routing the Assign application also provides a mapping table upload that can be used to identify project and event based on the data's accession number.

Export

As data is already in de-identified format in the research PACS exporting them for the use in external systems is straight forward. The VNA system for example allows users to export individual imaging studies with an embedded image viewer in the same way as clinical systems do. To allow for greater flexibility in data export capabilities the Export web-application allows user to export image data for a project in a variety of file formats. This includes study specific zip files that follow detailed specifications on the embedded directory structures, side-loading description files and the naming of DICOM tags and dates embedded in the data. The Export tool also supports more generic export formats such as NIFTI-format files for volumetric data.

NoAssign

Fiona's NoAssign application can be used to pseudonymize data without adding them to research PACS. Studies need to be forwarded to fiona.ihelse.net but will remain in quarantine there (for up to 7 days). If NoAssign is used during this time period the user may select a study from the list and either "download" the study as a pseudonymized zip file or forward the pseudonymized study to other clinical systems like "CDRobot" or "clinical PACS".

Review

Any automated de-identification requires frequent review to ensure that the process is working as expected. In order to support this work by the research project without requiring technical expertise we provide the Review web-application that lists all remaining tags in the data after the anonymization.

Attach

Image data not already in clinical systems can be uploaded in the Attach application. This includes DICOM files from USB/CD/DVD as well as whole-slide images files for pathology. After uploading them using Attach they will appear in the list of examinations on FIONA and can be either forwarded to research PACS using Assign, or exported again using NoAssign.

Processing

Processing of data is a step that links software into the research PACS. Data is forwarded from the project to the software which in turn sends result data back to the FIONA system. Those dataset are automatically forwarded to the research project and appear side-by-side with the original data. This type of integration requires a setup of the software and the setup of a send destination. Examples for such software endpoints are workstations from BrainLab and TeraRecon as well as data processing systems like CerCare and NeuroQuant.

System architecture

The research PACS component is designed to run side-by-side with the clinical system.

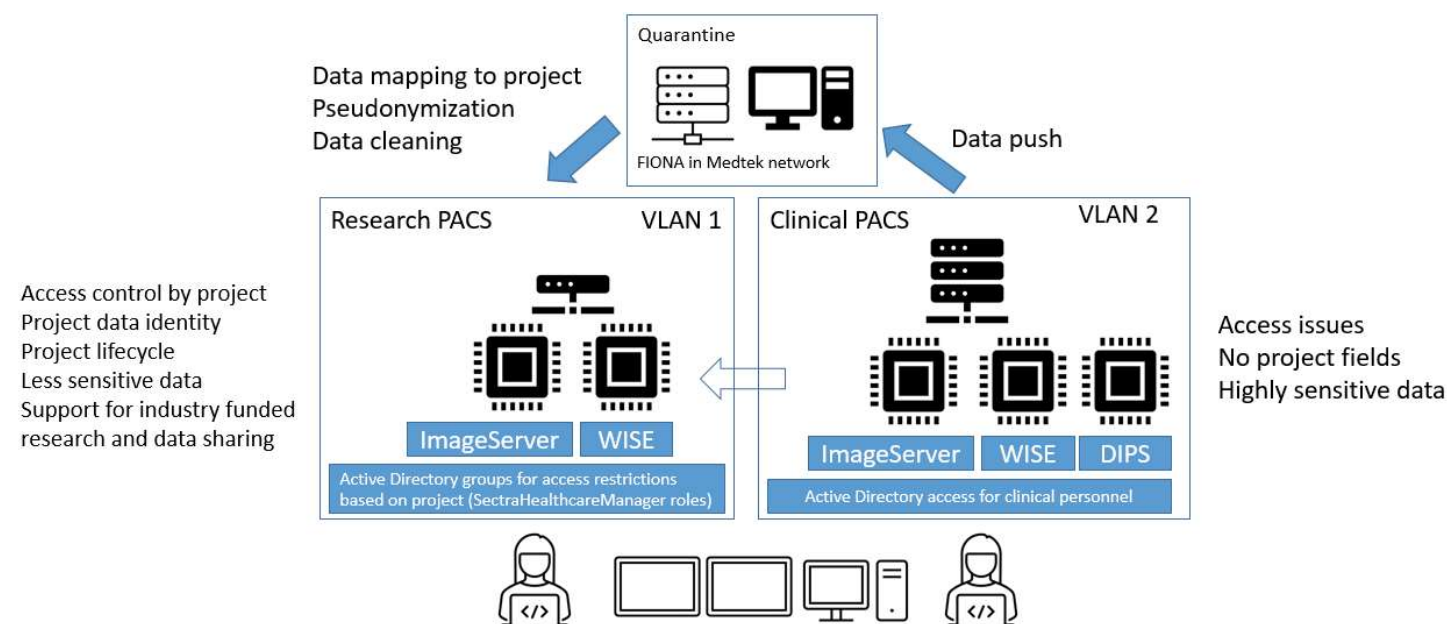


Fig. 1: Research PACS integration into a hospital environment.