Project Requirements

We prepared a questionnaire to assist you in setting up your project. Please note that only projects willing to use electronic data capture will be considered. The electronic data capture system is provided as a service of Helse Vest RHF, and therefore, at least one user must be a member of any of the hospitals in the Western Norway health region. The majority of the provisions in this document are in place to comply with requirements for collecting sensitive human subject data.

> Notice: Be aware of the institutional guidelines for projects. At Helse Bergen you are expected to follow the routines outlined here (link to EK).

Who is entering data?

- [ ] Data is entered by hospital workers such as nurses, doctors and students with a computer and account from the health region.
- [ ] Data is entered by study participants onsite, during a visit, in an health-care institution with support from study personnel.
- [ ] Data is entered by participants from home. This might include patient reported outcomes.

What features would you want to use during data collection?

- [ ] Scoring based on ICD-10, SNOMED-CT, MDDB, RxNorm or others.
- [ ] Randomization of participants into independent groups receiving indivualized treatment.
- [ ] Electronic consent and assent during screening and enrollment with signature fields.
- [ ] Calculated fields with values that are not entered but computed by the system and are based on previously entered 'raw' data fields.
- [ ] Fields or instruments that can be empty dependent on entered values (branching logic, missing data by design). This could be questionnaires asking about details in for specific diseases or treatments.
- [ ] Data collection events and forms that are filled out more than once.
- [ ] Reports for all or parts of the entered data for dashboards or statistical analysis.

- [ ] Are you collecting directly identifying fields from your study participants? This might include Norwegian Identification Numbers, names, dates of birth etc..
- [ ] Are you collecting identifying information from non-study participants? This might include names of personel filling out forms or counter-signatures for electronic consent or information about participants family members.
- [ ] Are you cooperating with other national or international project partners? Do they need access to parts of the data?
- [ ] Do you collect data centrally for a multi-center study? Other centers would need access to their data and participant cohorts.

Try to answered the questions above, you will see which of the existing risk assessments are relevant for your project. Referencing these risk assessments should simplify the process and allow you to start collecting data sooner.

A technical solution for data collection together with the way you are using that solution in your project needs to be formally risk assessed. IKT provides a structured service for this. You can start the process by requesting a meeting at: http://itil.ihelse.net/ROS, or at their newer site: https://helsevest.sharepoint.com/sites/HVI-ROS/. Fill out the form and you will be assigned a risk assessment number, an entry in the ROS website and a date for a first online planning meeting.

What happens during a risk assessment?

During a first planning ROS meeting you will have a chance to briefly describe your project (5min, ~4 powerpoint slides). You next suggest participants with relevant expertise for the risk assessment. This might include the owner of the system, other project members, representatives from industry partners and IT experts that will support the implementation, network and routing. The members of the first ROS meeting will assess if further input from (i) the data protection officer of the institution and the (ii) information security manager is required before the risk assessment can continue with a second meeting. After the first meeting all members of the risk assessment can suggest additional risks relevant to the solution. Here are two example of relevant risks: "Project user looses access to the technical solution, password lost or user work no longer for the institution" or, "Data is intercepted by a third party during remote access". Risks are entered on the ROS website.

The second ROS meeting (2h) will discuss each of the risks in turn. You will be able to explain the issue and all relevant procedures and technical tools that mitigate that risk. The process asks the team to define (a) How likely is this risk to happen? and (b) How severe is this risk? Both questions are assigned a score and the sum-score will define a risk category (low - green, intermediate - yellow, high risk - red). This process is repeated with all risks. Additional procedures or changes to the proposed solution will be discussed briefly.

The result of a risk assessment is a ROS document. The document contains the consensus scores for each risk. The project and their responsible leadership will be able to use the result of the risk analysis to decide if the risks related to the project are acceptable.

Data Protection Impact Assessments

Check if your project benefits from a DPIA, basically if your project will likely result in a high risk. You may use the document here to start such an assessment.

Transfer Impact Assessments

A TIA clarifies our organizational risks for transferring data from EU residents to countries without adequate protection under GDPR.

Guidance

The sections below relate to existing risk assessments. It is expected that all projects collecting human subject data perform a *coverage* risk analysis (see link above). When you apply, you may point to an existing workflow covered by such a risk analysis to simplify the process. Hopefully you can follow another projects procedures on how to use our Helse Vest infrastructure and no detailed risk assessment is required.

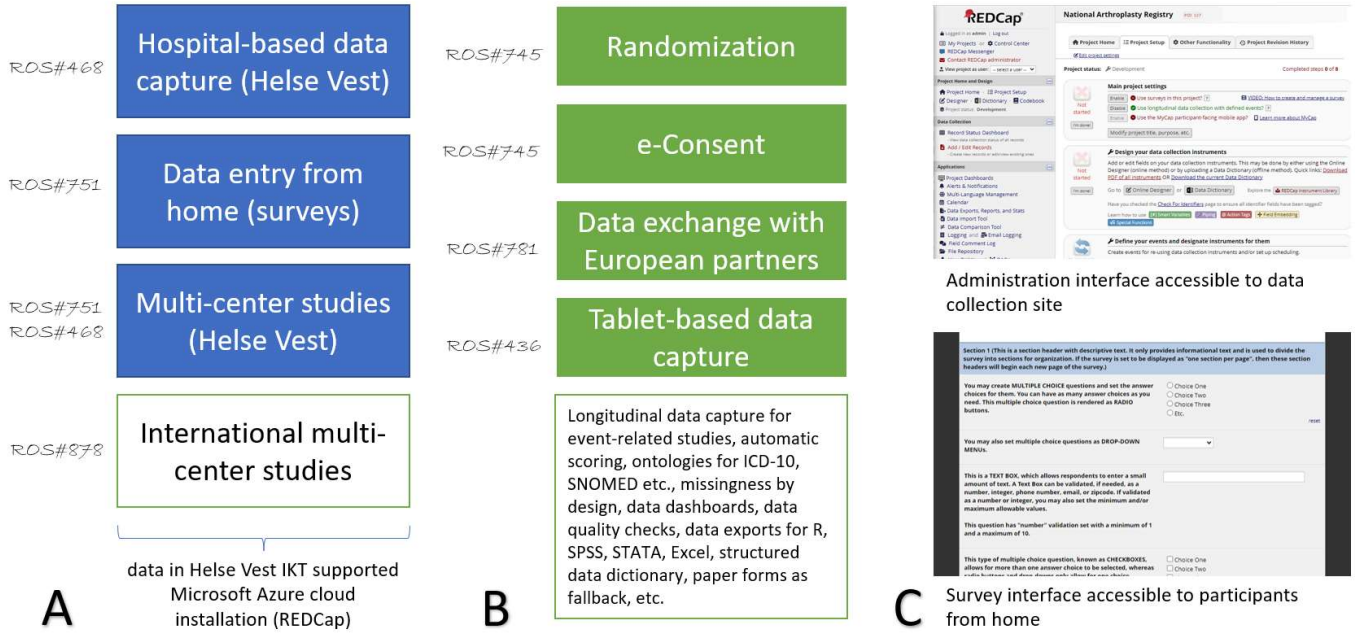

Fig1: Overview of supported Helse Vest study types, features and administration and participants interfaces.

In Fig1 above we display the types of studies supported by our Helse Vest infrastructure (column **A**). An abbreviated list of features accessible to all projects is displayed in column **B**. "ROS#.." annotations for study type and feature indicate existing risk assessments covering this functionality. Column **C, top** shows the administration interface used by study sites performing data collection, **C, bottom** shows an example survey interface that that can be filled out by study participants from home.

Overview of relevant risk analysis documents

A risk analysis can either be done for a limited feature set in our infrastructure or, as a coverage risk analysis for a project using one or more workflows. Each of the workflows may be based on an existing risk assessments. This *network* of risk analysis and the connected workflows may guide a project in establishing a good solution for electronic data capture.

1. [ROS468 REDCap COVID-19] REDCap for hospital based data capture
2. [ROS468 REDCap COVID-19] REDCap as a tool for electronic data capture
3. [ROS468 REDCap COVID-19] REDCap for COVID-19 project
4. [ROS478 Sectra DMA Forksning] Sectra PACS IDS7 installation for research
5. [ROS745 BMX-BAR] REDCap based randomization
6. [ROS745 BMX-BAR] REDCap based e-Consent
7. [ROS745 BMX-BAR] REDCap for BMX-BAR project
8. [ROS751 REDCap on Azure] REDCap for data entry from home
9. [ROS781 REDCap on Azure] REDCap for data exchange with European partner
10. [ROS751 REDCap on Azure] REDCap for REBECCA project
11. [ROS878 REDCap for multi-center studies] *Work in progress*
12. [ROS436 Bloodbank tablet] Tablet computer inside hospital

Data is entered by hospital personel

All our technical solutions are sufficient to capture data by employees of the hospitals. We suggest to use the more restrictive REDCap installation ([[REDCap 1 - Helse Bergen]]) for such projects that do not depend on data entry from participant homes. Note, that all study personnel needs to use a health region computer account.

Data is entered by study participants in the hospital

This case is similar to the hospital worker only projects. The difference is that parts of the forms would also be entered by study participants when they are visiting the health-care institution. This use-case is covered by the risk analysis [[ROS #468]] for surveys and risk analysis [[ROS #436]] for data entry using tablet computers. The data entry using tablet computers includes a workflow where study personel logs in to REDCap on the tablet, searches for the participant and event and selects the first data collection form in a survey queue. Study personnel logs out of REDCap and hands over the tablet to the participant with the first survey open. The participant can answer the first survey and additional surveys in the survey queue will be presented.

Data is entered by participants from home

For data entry from home the cloud based REDCap ([[REDCap 2 - Helse Vest RHF]]) needs to be used. This installation allows for access to data collection surveys using general internet access from home or from phones. Access to the REDCap user interface is only possible from hospital computers or using the hospital VPN. This use-case is covered by the risk assessment [[ROS #751 - REDCap on Azure]].

Cooperating with other national or international project partners

If you are planning to cooperate with other national or international partners (outside of the health region) you can have the role of either a data center (responsible for data capture in a study) or you the role of a data collection site responsible for local participants data capture.

Data center role: In order to support a national or international project as a data center you would need to be able to provide access to REDCap's user interface from outside the health region. This is currently only possible if you project partner has a least a "0%" position at one of our hospitals, a laptop computer from Helse Vest IKT and a chip-card for remote authentication. There is currently no project which is using our infrastructure in this way.

[Update]: The A4D and BMX-BAR projects are in the process of setting up multi-center access for hospitals from outside our health region (see ROS#878).

Data collection role: The REBECCA project is an example for a project that locally collects data (from home) but exchanges the data with another European center. The project provides access to the data using the REDCap API access. See the REBECCA project and risk assessment [[ROS781 REDCap on Azure]] for more information.

Establish a data processor agreement with all data processing partners based on for example the Standard databehandleravtale med veileder. There are alternatives to these individual agreements. For example consider a "joint controller agreement" between all partners that collect or access data. If data is transfered outside of the institution use data pseudonymization similar to the one performed in the research PACS (see EDPB recommendations on supplementary measures for data transfers, Case 2).

Directly identifying information

No directly identifying information should be stored in a REDCap *data project*. Data projects use pseudonymized participant identifiers such as "PROJECT_001". These fields are not considered anonymous if there is a coupling list pointing back from the id to the real person somewhere in the world (electronic or otherwise). Use a separate REDCap *ID project* to store directly identifying information about your participant. In the simplest case this would be the initials of the participant (only allowed in ID project).

The data and id projects allows project to implement specialized access permissions. A user with the role of "data-entry" would be allowed to access the data and the ID database for participants at their site. A statistician on the other hand only requires access to all data collection sites in the data project.

Data collection from non-study participants

Any person you collect data from needs be mentioned in the ethical approval for your project. This includes study personel names in consent forms as well as information about family members and neighbors of the participant in the study.

Example project COVID-19

This project resulted in the risk assessment "ROS468 REDCap COVID-19".

The COVID-19 study captured side-effects of the vaccination against COVID-19 from hospital workers like nurses. All study personel and participants had access to hospital accounts and received regular surveys from REDCap ([[REDCap 1 - Helse Bergen]]) on their institutional email accounts. The forms could be filled out by the participants if they were connected to the VPN (hospital laptop) or from workstations at the hospital.

Example project BMX-BAR

This project resulted in the risk assessment "ROS745 BMX-BAR".

The BMX-BAR project collects patient information in a randomized clinical trial using REDCap. There are two projects, one BMX-BAR for pseudonymized data and one project BMX-BAR-ID that implements e-Consent workflows as well as a link to personal information using patient initials. The coupling list with full names is not stored in REDCap ([[REDCap 1 - Helse Bergen]]).

The risk assessment [[ROS #745]] covers the use of REDCap for study participant randomization. Such randomizations should be unknown to the study personnel. Usually a statistician of the study will create a randomization table (see below) and upload it to the REDCap project. Once a participant has been screened and can be enrolled, a randomization value is pulled from the table and assigned to the participant. This can only be done once (press the Randomize button in REDCap) and is documented in REDCap (user, date).

General steps for randomization

1. Create a randomization table with the total number of participants and a value of "0" for group 1 and a value of "1" for group 2. This table should be created by a program such as R using code such as (two groups with 30 participants each):

   ```
   n = 30
   t = sample(append(rep(0, n),rep(1, n)))
   d = data.frame(randomization=as.numeric(t))
   write.csv(d, file="randomization_table.csv", row.names=FALSE)
   ```

   Update: Instead of the single large block of N=30*2 entries we suggest to use a block-based randomization with a smaller block size, i.e. N=4. After each block the randomization is guaranteed to be balanced:

   ```
   n = 30        # number of entries that are 0
   BlockSize = 4 # after BlockSize random value we are balanced
   NBlocks = ceiling((n*2)/BlockSize)
   t = list()
   for (b in seq(1, NBlocks)) {
     ar = append(rep(0, BlockSize/2), rep(1, BlockSize/2))
     t = append(t, sample(ar))
   }
   d = data.frame(randomization=as.numeric(t))
   write.csv(d, file="randomization_table.csv", row.names=FALSE)
   ```

2. Enable the randomization feature in REDCap and upload one version of the table for the Development modus and a second (independently generated) version of the table for Production mode.
3. Test the solution in Development mode and, at the start of the study, switch to Production mode - remove all previous data. REDCap uses two different randomization tables for development and production.

The BMX-BAR project also uses REDCap for the collection of e-Consent documentation.

Example project REBECCA

This project resulted in the risk assessments "ROS751 REDCap on Azure" and "ROS781 REDCap on Azure".

The REBECCA study is an international breast cancer study that uses REDCap in the cloud ([[REDCap 2 - Helse Vest RHF]]) for data capture. The European REBECCA project provided the REDCap data entry forms. The international group is using the REDCap API (application programming interface) to get access to the Helse Vest REDCap for a subset of the data collected in Norway. The connection from the European REBECCA consortium cloud is secured with a project and user specific REDCap token as well as a specific firewall opening based on the two IP addresses (European cloud computer and Helse Vest RHF cloud computer). API access is only provided if such a dedicated connection can be established.

ROS #468 - REDCap use of surveys

This risk assessment was performed by Helse Bergen to allow for the use of REDCap ([[REDCap 1 - Helse Bergen]]) in the COVID-19 project for electronic data capture. No personally identifying information was captured by the project but survey links send weekly to hospital workers required the use of email address fields in the study.

Email address fields are required in a data project if survey links need to be send by REDCap. Automatic survey distribution can be used if (i) the email field is marked as an "Identifier" in the REDCap designer application, (ii) the email field has the action tag "@PASSWORD" to prevent accidental viewing of the field in data entry forms and (iii) proper export settings (roles) prevent users from exporting this field as part of reports.

ROS #745 - REDCap based randomization

See the documentation for the BMX-BAR project on the use of REDCap randomization of participants.

ROS #745 - REDCap based e-Consent

Electronic informed consent or assent in a study is required before any other data collection can start. If a potential participant is interested in the study the first step is "Screening". After a successful screening the participant should be fully informed about the study. This is documented on a consent document. The language of this document is established as part of the application of the project at the

regional ethics board.

The approved consent text needs to be implemented in the ID project in REDCap. This is done using HTML features for colors, font sizes and highlights in the text. At the end of the consent document two verification sections are added. The participant verification section expects the participant to enter his/her full name as well as to provide a *screen signature*. Such a signature is not a full biometric type data field. REDCap signatures only store the manually drawn signature as a picture together with verification information like date and time. The second verification section requests the informant in the study to sign in the same way.

Project need to export a paper version of the filled-out e-consent document from REDCap as a PDF. The PDF document should be stored or printed and the paper version is stored in a secure location.

ROS #436 Tablet access

This risk assessement performed by the blood bank of Helse Bergen established the workflows to perform data entry on tablet computers. Two option where evaluated, an iPad and a Samsung tablet version. The devices need to be purchased through Helse Vest IKT which prepare the tablets to run a single application. Options for the application are either a single web-page like REDCap or a dedicated tablet application for example for cognitive testing.

ROS #781 - external access to REDCap API

For any external access to REDCap we require a 2-factor authentication. This is only possible with the REDCap API if an IP-to-IP firewall exception is used.

ROS #751 - data entry from home

Access to survey instruments for participants from home is only supported in the [[REDCap 2 - Helse Vest RHF]] cloud installation. Participants receive email notifications that contain a link to the REDCap survey.

ROS #478 - image data pseudonymization for research

Image data requires specialized procedure for pseudonymization. This includes the pseudonymization of meta-data (DICOM) as well as the removal of burned in sensitive information in images.

ROS #878 - international multi-center study support

The Western Norway health region extends its support for hospital based multi-center studies. Partners from outside the health region can participate as data collection sites with access to the administration interface for a study and control over their own participant cohorts. All data is still collected centrally under a single study design. This risk assessment is work-in-process (WIP) and requires dedicated firewall rules to allow access to users from partner institutions.

Feature: Scoring

For electronic data capture solution all scoring should be done during data collection. REDCap provides access to many Ontologies for scoring that allow users to search during data entry in Ontology REDCap fields based on 2 to 3 characters. Users have to select from the presented list and REDCap will store both the numeric code of the Ontology entry as well as the textual description presented to the user. To export such information from REDCap perform two export steps. First 'export as data' to obtain the numeric codes in such fields. Next 'export as labels' to get the textual description.

The feature of Ontology fields is a basic feature of REDCap ([[ROS #468]]) and has no separate risk assessment.

Feature: Calculated fields

Data collection forms distinguish between 'raw data' or 'primary data' and 'derived data' or 'scores'. Whereas raw data is single item information entered during data collection the scores are calculated from the raw data. As a single example a sum of 10 raw data fields each represented as a checkbox might result in a sum-score between 0 (no checkbox selected) and 10 (all checkboxes are selected). REDCap should be used to compute such simple scores to prevent data entry errors.

A limitation of calculated fields in REDCap is that they should not depend on other calculated fields. Only raw data fields should appear in the calculation.

If the calculation needs to be changed and prior entered data already exists the project can use the "Data Quality" page in REDCap to recompute all calculated fields.

The feature of calculated fields is a basic feature of REDCap ([[ROS #468]]) and has no separate risk assessment.

Feature: Missingness by design

The visibility of a data collection field can depend on the value of previously entered fields. This feature is called 'branching logic' in REDCap. A simple example are followup questions in case a particular diagnosis has been indicated.

REDCap can provide branching login on the level of whole data collection instruments or on individual fields. Branching logic should not be used to implement data collection event logic. Instead the REDCap feature for longitudinal data capture should be used.

The feature of calculated fields is a basic feature of REDCap ([[ROS #468]]) and has no separate risk assessment.

Feature: Repeated data collection forms

The same data collection form might be filled out more than once. If the form belongs to different REDCap events (longitudinal project) the study personel can select the event for which data should be entered. This works well for studies with well-defined events. In cases that studies need to collect an unknown number of instances of a form a more complex "repeated form" feature can be used. Check first if you can instead assume a maximum number of events and implement them as such. Only in cases that there are no well-defined events should you attempt to use the repeated forms feature.

REDCap 1 License - Helse Bergen

A Research Electronic Data Capture installation based on the following Vanderbuilt license (Haukeland University Hospital):

> Contact: Hauke Bartsch [Hauke.Bartsch@helse-bergen.no]
> Date: 2020-03-19

REDCap 2 License - Helse Vest RHF

A Research Electronic Data Capture installation based on the following Vanderbuilt license (Helse Vest RHF):

> Contact: Hauke Bartsch [Hauke.Bartsch@helse-bergen.no]
> Date: 2022-08-30

Statistical analysis tips

We find the guidelines of the EU (ema) very useful. Here an example of the texts on how to adjust covariates for clinical trials.