

BBMRI.nl & Health-RI Expert Meeting: How to make health(care) data available for research?

REPORT EXPERT MEETING #2

The second virtual [BBMRI.nl](#) and [Health-RI](#) expert meeting on practical approaches to make health(care) data available for research was held on the 12th of November 2020. This initiative is a collaborative effort of BBMRI.nl and Health-RI to implement a data-driven health research infrastructure for optimal access to knowledge, tools, facilities, health data and samples. This data-driven health research infrastructure is fundamental in realising our ultimate goal of a learning healthcare system that enables sustainable and affordable personalized medicine and health. But how should this infrastructure be shaped? Which organisations are currently providing this service? And how do they provide access to their health(care) data?

The goal of this expert meeting series is to discuss the various practical approaches, and their strengths and weaknesses, to make different types of health(care) data available for scientific (re)use. During this meeting, the processes to find, access, request, share and link data within the CAPACITY study of the Durrer Center were presented by Wanda Hermans-Van Ast and Erik van Iperen. The data requesting, -sharing and -analyse procedures within Netherlands Comprehensive Cancer Organisation (Integraal Kankercentrum Nederland; IKNL) were presented by Vincent Ho and Peter Prinsen. This meeting was attended by ~30 experts from various organisations and universities.

CAPACITY study from the Durrer Center

[Durrer Center](#) is a facility of the Netherlands Heart Institute which aims to support researchers within the cardiovascular research field, focussing on biobanking and data management. The [CAPACITY-study](#) is a registry of patients with COVID-19 and aims to collect data regarding the cardiovascular history, diagnostic information and occurrence of cardiovascular complications in COVID-19 patients. By collecting this information in a standardized manner, CAPACITY can aid in providing more insight in (1) the incidence of cardiovascular complications in patients with COVID-19, and (2) the vulnerability and clinical course of COVID-19 in patients with an underlying cardiovascular disease. This registry is based on the case record form (CRF) of the ISARIC (international severe acute respiratory and emerging infection consortium) and WHO (World Health organisation). This CRF has been extended to meet the requirements for cardiovascular research, including the history of cardiovascular disease and prevalence of cardiovascular complications.

During the first pandemic spread of COVID-19 (April till September), more than 7000 patients have been included deriving from 75 centres in 13 countries. The registry was established in a short period. Consent for data collection was provided via an opt-out procedure. All hospitals in the Netherlands have been contacted to collaborate with the registry. This complex procedure is a collaborative effort of the Dutch Cardiovascular Alliance (DCVA), including Dutch Network for Cardiovascular Research (WCN), Dutch Heart Registration (NHR),

Netherlands Society of Cardiology (NVVC), Harteraad, the Dutch Heart Foundation, and the Durrer Center.

Each participating centre needed to accredit the CAPACITY research protocol and appoint a local coordinator. Following, a data transfer agreement (DTA) was signed by the legal representative of each hospital. Importantly, in the legal framework data sharing is organised centrally. The coordinating centre is UMC Utrecht, meaning that UMC Utrecht has the legal remit to share data deriving from the various hospitals for scientific use but only after approval of the data access committee (DAC). Each data access request is reviewed by the DAC. The DAC is formed by 10 experts from the partner institutes that established the study (Dutch Cardiovascular Alliance), including data coordinators, database experts, epidemiologists, and researchers.

Data entry was done via Research Electronic Data Capture (REDCap). A graphical representation of the data delivery procedure is depicted in figure 1. Data quality is monitored for completeness upon entry through an automatic script. This script identifies missing data and notifies the corresponding centre with the question to correct accordingly. All documentation involved in the data collection, -entry, -processing, and -access procedures are available on the [CAPACITY website](#).

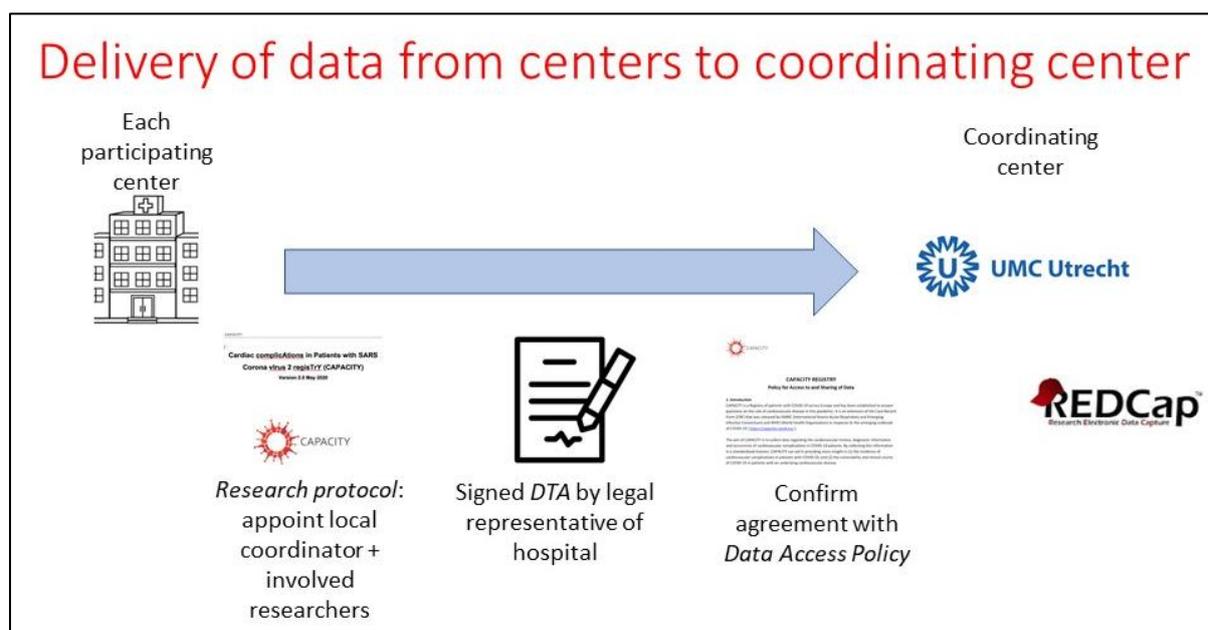


Figure 1. The data delivery procedure of the CAPACITY-study

In short, the procedure how to apply for data access is being described here:

- First check the overview of applications to avoid overlap with an existing project.
- Start your application by filling in the application form via <https://podium.bbmri.nl>. By sending this form the request will go to info@capacity-covid.eu
- To gain access, this form including your research questions, required data and/or samples, financing, and timeline of the research must be submitted

- Make your data selection using [this specific CAPACITY REDCap codebook](#). Please select your variables with corresponding variable names and labels. This file must be submitted as well alongside the application form using <https://podium.bbmri.nl>.
- The submitted application will be reviewed within one week by the CAPACITY Data Access Committee.
- Once an application is approved, CAPACITY will prepare a Data Transfer Agreement (DTA) for you to sign.
- Data will be released after signing the DTA.
- The data will be released in a digital research environment (the CAPACITY workspace).
- For further information regarding data access [see this file](#).

In more detail, the [data request portal Podium](#), developed within BBMRI.nl, is being used to manage data requests and make its progression through the review process transparent for the researcher. For more information upon the possibilities and use of Podium, please read the [user guide](#) or request a Podium demo at your organisation by contacting Erik van Iperen (e.p.vaniperen@amsterdamumc.nl) or Sofie Hansen (sofie.hansen@lygature.org). Registration as a Podium user can be done via [this link](#) or by contacting servicedesk@health-ri.nl. Organization can also register in Podium by contacting servicedesk@health-ri.nl. Account registration is being checked for validity and safety via the Health-RI ServiceDesk. Upon approval, the researcher can create a data request by filling in the necessary details such as the title, request type, organisation, etc. It is possible to request data from various data- or biobanks simultaneously within one application. Furthermore, linked requests are also possible within Podium. The research proposal needs to be specified by the researcher in order to allow the DAC to review the scientific validity of the application. Principal investigator details need to be specified. It is required that the [CAPACITY codebook](#), specifying the variables and data which are being requested by the researcher, is uploaded when submitting the application.

Once performed, submitted applications undergo two step validation. First, the coordinator validates the completeness and validity of the request. If necessary, revision of the application can be requested via Podium which results in automatic notification of the data requestor. Second, each individual member of the DAC will review the application following the acceptance by the coordinator. The review comments of each DAC member are automatically uploaded and communicated towards the coordinator. After review of the individual DAC members, the submitted request and the corresponding review comments are than being discussed during one of the biweekly DAC meetings, resulting in an approval, rejection (data not available), or revision.

The data delivery phase is initiated by the coordinator upon approval. The researcher will receive an automatic email notification that their data will be released in a digital research environment. An additional DTA is required to be signed between the requestor and the CAPACITY data manager (UMC Utrecht as coordinating centre) to allow for data delivery. The corresponding dataset is prepared in a digital research environment (DRE) developed by [ANDREA](#). A separate workspace is being developed per project. All authorized researchers will be granted access to this specific workspace via a two-factor authentication. The workspace contains the necessary statistical programs (SPSS, R, and STATA) to allow

analysis and documentation of the data without the data leaving this virtual environment. ANDREA monitors usage of the DRE, informing about the persons, time, and location of the data handling.

After completion of the data analysis and conduction of the research, researcher will present the findings of the research conducted with the requested data. This will be reviewed by the DAC to ensure compliance with the initially submitted research proposal. Publications are also reviewed on authorship and acknowledgements. Currently, 26 requests have been submitted of which 5 are under review, 15 are approved, 2 are rejected, and 4 revised.

Netherlands Comprehensive Cancer Organisation

Netherlands Comprehensive Cancer Organisation (Integraal Kankercentrum Nederland; [IKNL](#)) hosts the Netherlands Cancer Registry ([NCR](#)). NCR was established to monitor the incidence and prevalence of cancer on request of the Ministry of Health, Welfare and Sport (Ministerie van Volksgezondheid, Welzijn en Sport). The NCR reached nationwide coverage in 1989 while regional data has been collected since 1955. It is estimated that currently 90-95% of all new cancer cases are captured in the NCR. More than 115.000 new cases of cancer are being diagnosed every year in the Netherlands. Data managers of IKNL operate in almost all hospitals in the Netherlands to collect the data. The NCR contains more than 2 million cases in the database.

The NCR receives an automatic notification when the diagnosis of a patient with cancer is confirmed in a pathology laboratory. Other cancer cases are identified later through the LBZ (Landelijke Basisregistratie Ziekenhuiszorg). After a certain amount of time, typically a few months to a year, depending on the tumor type, data managers add information such as patient characteristics, treatment details, and other relevant information, which they extract from hospital data. This first moment of data collection by the data managers is performed after completion of the primary treatment. Increased capacity of data managers would allow for more routine monitoring and data collection. Information concerning the vital status of the patient is gathered via data linkage to the [Personal Records Database](#) (Basisregistratie Personen), formerly known as gemeentelijke basisregistratie voor persoonsgegevens (GBA). It is challenging to collect data on the follow-up of the treatment, because of the work involved. This is partly achieved on a project basis in which additional data collection timepoints are implemented for specific tumor types. These are not part of the “normal” registry but are individual projects with a specifically defined scope, duration, and aim. The NCR includes the minimal data set defined by the WHO/IARC: patient demographics (birth date, sex, postal code, vital status), and basic tumor characteristics (topography, morphology, stage). Additional data includes additional diagnostic information and treatment details (type, date, etc.). Depending on the tumor type, additional data can be requested concerning molecular markers, complications, etc. Requestors are informed about the details of the available data sets (such as the variables and data collection period) on the [IKNL website](#).

Data is collected and shared through opt-out consent based on legitimate interest (“gerechtvaardigd belang”) and a special exemption in the GDPR: special category data. Agreements have been established with participating hospitals to define the circumstances

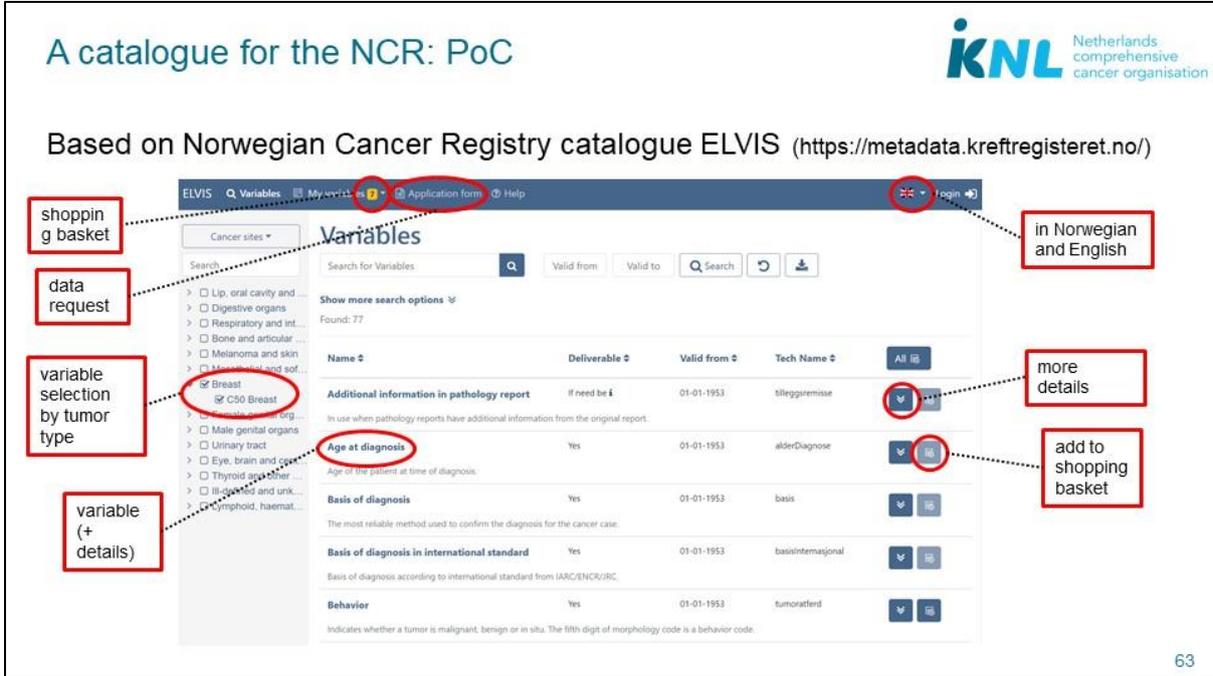
and requirements for ethical and safe data collection and access. Each data request is reviewed by the supervisory board of the NCR. This board evaluates the privacy protection on a patient-, doctor-, and institute-(hospital) level. The supervisory board members include a lawyer, patient representatives, medical practitioners, and hospital representatives. Furthermore, scientific committees have been appointed for every tumour group to guarantee the necessary scientific expertise to review every data request on scientific validity.

The data request procedure is described on the [IKNL website](#). Questions can be directed to gegevensaanvraag@iknl.nl. A data request is initiated by filling out a data request form, signed by both the applicant and IKNL. The number of filed requests for NCR data is around 500 per year. The majority of the requests are 'regular' requests. Requests also include, in smaller amounts, quick-information, linkage, and complex linkage requests. The complexity of the request impacts the lead time, varying from ~1 week for quick-info requests to ~14 weeks for complex linkage requests. Importantly, these lead times start from the moment of signing of the contract. Prior to this, the duration of the preparation of the contracts varies from ~1 week for quick-info requests to ~12 weeks for complex linkage requests. IKNL evaluates the experience of the data applicant. Feedback is generally positive, with customers citing a short duration of data delivery, a clear data request procedure, and excellent communication.

Although an effective and efficient data sharing procedure is in place, IKNL is aiming to further improve by 1) decreasing the lead times, 2) increasing the transparency of the data request process, and 3) providing more information on the contract preparation times as well as on the available data sets. For the latter, PDFs on the IKNL website are currently used to inform potential data requestors about the available data sets for each tumor type. IKNL is aiming to improve this by developing an online catalogue with an overview of variables and details (variables, definition, details, values, collection period, quality, etc.). Preferably, this would be integrated with the data request procedure. This is very challenging because data requests can be customised when requesting IKNL data, for example by requesting derived variables (for example hospital type or hospital region instead of hospital (note: hospital itself cannot itself be requested due to privacy)) or requesting groupings of values of a variable (for example age categories instead of age). Furthermore, variables can be used to select a cohort, for example based on a specific collection period or patient age. A proof of concept (PoC) catalogue will be launched for breast cancer in Q2 2021. Use of the PoC catalogue will be stimulated among data requestors that are interested in this specific cancer type to gain input on requirements from researcher-perspective and to identify limitations. For this PoC catalogue the Norwegian Cancer Registry catalogue ELVIS (<https://metadata.kreftregisteret.no/>) will be used, modified to fit the NCR. ELVIS describes the various tumor types, the available variables, and more details such as collection period, quality, values, etc. (see Figure 2). It also offers the possibility to add variables to a list and export this list. The development of this PoC catalogue is challenging since the information about the variables is not all centrally located in IKNL. If you have any questions or remarks, feel free to contact Vincent Ho (v.ho@iknl.nl) and Peter Prinsen (p.prinsen@iknl.nl).

A catalogue for the NCR: PoC

Based on Norwegian Cancer Registry catalogue ELVIS (<https://metadata.krefregisteret.no/>)



shopping basket

data request

variable selection by tumor type

variable (+ details)

in Norwegian and English

more details

add to shopping basket

63

Figure 2. Screen shot of the Norwegian Cancer Registry catalogue ELVIS. The PoC NCR catalogue will be based on this catalogue.

The next and final expert meeting will be held on Thursday the 26th of November. During this meeting, Trynke de Jong will present the processes to find, access, request, share and link data within Lifelines, while Pascal Suppers will share the experiences of DataHub. If you are interested to join one of the following expert meetings or do you have any questions or comments, feel free to contact Robin Verjans (robin.verjans@lygature.org).