Programme proposal

Data service for Human Subject Research

October 20, 2016
HSR Programme Proposal

Appendix 1. Open Science context .............................................................................. 39
Appendix 2. Legal context .......................................................................................... 40
Appendix 3. Vocabulary ............................................................................................... 41
  Functions .................................................................................................................. 41
  Definitions .................................................................................................................. 41
  Abbreviations ............................................................................................................ 42
Appendix 4. Related (inter)national initiatives ............................................................ 43
Appendix 5. Contributions ............................................................................................ 45
  Programme Team ...................................................................................................... 45
  Programme Board .................................................................................................... 45
  Advisors ..................................................................................................................... 46
  Special Interest Group .............................................................................................. 47
Appendix 6. Budget ....................................................................................................... 48
**HSR Programme Proposal**

**Summary**

Excellent research in the Big Data era challenges existing data handling routines, IT structures and responsibilities for data. The flood of data that has become available to researchers and the accompanying technological demands has increased substantially in recent years. In combination with the current visions on how to manage scientific data efficiently, securely and with integrity (e.g. Open Science, FAIR data) and the developments in legislation, ask for an accessible, high-quality and robust IT infrastructure. These demands particularly apply to data from research involving human subjects, due to the ethical and legal aspects of these data.

The currently available services and tools in the University of Groningen (UG) and the University Medical Center Groningen (UMCG) are not optimally connected and are not easily accessible for a wide range of researchers. These shortcomings may interfere with the aim for excellence and outstanding high-quality research, hampers the institute-wide implementation of proper research data management and Open Science/Open Data principles, delays compliance of the institutes to the current legislation, and could therefore pose a reputational or financial risk for the institutes.

A 2-year programme is proposed to set-up a service and IT infrastructure for UG/UMCG researchers. The programme aims to set up a service that facilitates research on human subjects by providing the researcher an up-to-date IT infrastructure that supports the collection, processing, storage and (re)use of research data, while ensuring data security and properly protecting the privacy of participants. The IT infrastructure will enable innovative research and is in sync with the current vision on research data management and FAIR data. In addition, the programme will support the responsibility of the Boards of the University and of the UMCG to foster compliance with The Netherlands Code of Conduct for Academic Practice on academic integrity and the Data Protection Law.

The programme is a joint activity of UG and UMCG, headed by a Programme Board that consists of senior researchers from the faculties with human subject research projects, the university IT center, the University Library and the Office of the University. The programme is executed by multidisciplinary teams from the participating institutes to stimulate exchange of knowledge, experience and tools. The infrastructure and expertise developed in this programme will build on the existing excellence and tools, create new solutions and add missing links. The programme will be financed by the ICT strategy fund and the UMCG Research IT budget.

At the end of the programme, the easily accessible service and robust IT infrastructure will be sustainably embedded within the UG/UMCG. It will consolidate the leading position of the UG/UMCG in large scale data collections and research IT excellence.
HSR Programme Proposal

1. Introduction

Excellent research in the Big Data era challenges existing data handling routines, IT structures and responsibilities for data. The flood of data that has become available to researchers and the accompanying technological demands has increased substantially in recent years. In combination with the current visions on how to manage scientific data efficiently, securely and with integrity (e.g. Open Science, Open Data, FAIR data initiatives) and the developments in legislation, ask for an accessible, high-quality and robust IT infrastructure. This IT infrastructure needs to be complemented with expertise on information management, privacy protection, data security, and other ethical and legal aspects. These demands particularly apply to data from research involving human subjects, due to the ethical and legal aspects of these data.

The University of Groningen (UG), UMCG and partners have a longstanding experience in the setup and management of complex large data collections. However, the currently available services, IT structures and expertise that need to deliver the required IT and expertise are not optimally connected and are not easily accessible for a wide range of UG/UMCG researchers. Also, crucial building blocks are lacking or require further development, like a research data register, a data archive, an infrastructure to share data safely with external researchers, and tools to make research data findable and re-usable.

These shortcomings may interfere with the aim of the UG for excellence and outstanding high-quality research, may hamper the institute-wide implementation of privacy by design, proper research data management and Open Science principles. It addition, it may delay compliance of the institutes to the current and shortly coming legislation, and could pose a reputational or financial risk for the institutes due to privacy breaches or when the institutes are unable to comply with requirements from scientific journals and funding bodies to provide details on data management.

The required IT structures and expertise for research involving human subject data have become too complex and too expensive to realise for most individual researchers. Therefore, a joint approach and combined efforts are vital. This has also been acknowledged by the UG in their strategic plan and in the ICT plans of the UG and UMCG. This proposal presents a programme to efficiently build on the existing IT building blocks and expertise within UG/UMCG and create an innovative, easily accessible IT infrastructure for human subject research in order to support innovative and transparent research.

2. Aim

The aim of the programme is to facilitate research on human subjects by providing the researcher service and an up-to-date IT infrastructure. The service and IT infrastructure will be in sync with the current vision on research data management, follow the FAIR principles of Open Data and support Open Science (Appendix 1). The IT infrastructure will enable innovative research and will be prepared for complex large scale sensitive data like imaging data, real-time streaming data, and highly detailed data for personalized health (omics data).

In addition, the programme will support the responsibility of the Board of the University to foster compliance of research activities within the institute with The Netherlands Code of Conduct for Academic Practice on academic integrity, ethical standards and the Data Protection Law (Appendix 2).

**Definition Human Subject Research**

Human subject research means systematic, scientific investigation that can be either interventional or observational and involves living individuals as research subjects. Human subject research can be either medical research or non-medical (e.g., social, spatial, business) research.
HSR Programme Proposal

3. Outcome
The programme delivers a sustainable and easy accessible service and an up-to-date and robust IT infrastructure connected with relevant expertise for research involving human subjects performed by researchers at all faculties of UG/UMCG to enable innovative research. The service will be designed to support the researcher in the complete research data life-cycle (Figure 1): from the design of the study, the re-use existing data and/or collect new data, through the research phase to ending of the study with long term archiving and preparation for re-use (one-stop shopping principle). The service will facilitate Open Science and re-use of research data, while ensuring data availability, security and integrity and properly protecting the privacy of participants. The infrastructure will consist of several, well-connected building blocks that can be used in combination or separately and are aligned with the Research Data Management policies of the faculties.

Users
The aim is to create a solid basis for excellent research in Groningen. All scientific disciplines and researchers performing human subject research are able to use the service and IT infrastructure, irrespective of the size, type and or duration of the research. The expertise and tools will also become available for other sensitive research data, for example IP-protected data. Making the service available to researchers outside Groningen and for partners in the private sector will follow during a later stage.

The programme recognizes that researchers are a diverse group of people operating in many different contexts. A flexible framework will be setup, to support the research appropriate to their circumstances and needs without having a ‘one size fits all’ approach.

Figure 1. Data Lifecycle with surrounding infrastructure and foundation
HSR Programme Proposal

Quality

The infrastructure will be compliant to the applicable regulations and laws that are in force (including the upcoming European General Data Protection Regulation) and apply security standards developed with and reviewed by peers and an external party. Also, the infrastructure will seek approval from external certifying bodies (for instance by obtaining a Data Seal of Approval) and comply with the data management demands of funding bodies. To facilitate research and enable smooth collaboration between UG, UMCG and external researchers without compromising security of the data, security managers of UG and UMCG and the legal officers of both institutes are actively involved in the programme. The functional quality will be evaluated by users and follow a Plan-Do-Check-Act approach. Within the UMCG the programme will follow the ISO9001 quality system. To ensure the developed service and infrastructure fit with the requirements of researchers, researchers from different disciplines are involved from the start of the programme.

Accessibility

The IT infrastructure and expertise are made available through a service portal (embedded in the UG/UMCG), that makes it easily accessible for all researchers working with data from human subjects.

The service portal will:

1) assist the researcher with an initial holistic assessment of the data (data already collected or to be collected) on amongst others privacy, security, and access aspects and help the researcher to formulate the needs for data management and IT. This assessment will lead to a research data management plan;

2) serve the researcher the proper data management tools for all parts of the data cycle according to a one-stop-shop principle and guide the researcher towards the right IT partners and other services in case of complicated requests;

3) document the research project and the scientific data collection (e.g. via register and persistent identifier);

4) perform (peer) audits and deliver services to the researcher during the complete research phase.

In addition, the service portal will:

5) create visibility for the Groningen partners, consolidate the Groningen position as an expertise hub for human subject research data and research IT infrastructures;

6) become a platform for knowledge exchange on different aspects of managing human subject data, increase familiarity with data management plans, develop training for researchers and support staff;

Integration

At the end of the programme, the service will be embedded within the UG/UMCG and closely connected to relevant programmes, networks and facilities including those at the individual faculties. The service will consist of a small, dedicated team of persons with expertise on data management, privacy, security and IT architecture with complemented competences on service and PR for first-line service. Data consultants (definition in Appendix 3) who provide the service will be positioned part-time in the faculties and institutes to ensure an effective collaboration with researchers and service provision.

Likewise, the new IT components will be imbedded in the different existing IT and data facilities of the UG/UMCG, to ensure proper use, maintenance and optimal connection to the other tools and structures of the different facilities and services. Defining a sustainable governance structure and proper embedding in the institutes for the service and IT infrastructure will be part of the programme (see deliverable WP5.4).
**HSR Programme Proposal**

The IT structures and expertise developed in this programme will build on the existing excellence and structures that are already present within the UG/UMCG or within (inter)national collaborations (e.g. BBMRI-ERIC, ELIXIR), create new solutions, add missing links and integrate existing knowledge and tools.

On a national and an international level, the infrastructure will be connected to relevant data infrastructures (e.g. HEALTH-Ri10, data4lifesciences11) and expert centres (e.g. LCRDM, RDA, EUDAT) to enable (inter)national collaboration, knowledge exchange and data sharing (Appendix 4).

4. Programme description

**Programme positioning**

The 2-year programme is a joint activity of UG and UMCG, headed by a Programme Board and executed by a Programme Team, that will be coordinated and supervised by a Steering Committee and a Programme Manager (Figure 2). The program will be hosted at CIT (pensvoerder). The activities of the programme are organized in Work Packages (WP), which are multidisciplinary teams from the participating institutes, to stimulate exchange of knowledge, experience and operational methods. The members of the Program Board, the Steering Committee and the Programme Manager will be appointed by the Board of the University ( CvB) after consultation of the Board of the UMCG (RvB).

The WP leaders and team members will be (or stay) employed at one of the participating institutes at UG or UMCG, and positioned both at the institute and the programme. The complete Programme Team will be positioned at one location for at least 2 days a week, to ensure an efficient collaboration and proper alignment of the deliverables. The other days, the members of the Programme Team may be positioned at the institute of employment or other faculty or facility to ensure effective embedding with the different facilities and faculties.

**Programme governance**

**Programme Board**

The Programme Board will guide the programme, is responsible for the execution and finances of the programme and takes care of proper alignment with other local, national and international projects, initiatives and stakeholders. In addition, the Board is responsible for the governance of the to-be-developed service within the UG and UMCG. The Board consists of senior researchers and/or research policy makers from the faculties with human subject research projects, the University Library, and the Office of the University (see Appendix 5 for an overview of the current members). At the start of the program, the Program Board will be chaired by the Director Research Data & Biobanking of the UMCG. The chair reports to the Board of the UG and Board of the UMCG.

**Steering committee**

The execution of the programme and operational activities are supervised by a data/IT expert from UG and from the UMCG. At the start of the programme the technical director of CIT (UG) and the head of the Genomic Coordination Center (GCC, UMCG) will hold these positions. Together with the Programme Manager they from the Steering Committee. The Steering Committee reports to the Programme Board.
**HSR Programme Proposal**

**Programme Manager**
The Programme Manager is responsible for the overall coordination of the programme and for the timely realisation of the deliverables within the planned budget. The Programme Manager supervises the WP leaders and the staff of the programme. The Programme Manager is the central day-to-day contact person of the programme and is responsible for good connection with other UG/UMCG facilities, projects, and stakeholders on an operational level, to ensure optimal alignment and exchange with existing initiatives and to ensure optimal dissemination of the tools and expertise.

**WP leaders and team**
Every WP is led by a WP leader, who is responsible for the coordination and execution of the WP. The WP leader is the prime contact person for the WP. The WP leader is responsible for timely progress of the deliverables, for alignment of the own WP with the other WP’s and supervising the WP team. In every WP team, a task leader is responsible for the realisation of the specific deliverable within the planned time and budget. The task leader is responsible for alignment with the other deliverables within the WP and reports to the WP leader.

![Diagram of Programme Governance](image)

**Special Interest Groups**
The members of the Special Interest Groups give functional input to the development of the IT infrastructure and the expertise, by providing use cases and functional demands (see Appendix 5 for an overview of the current members, which will be expanded during the programme). They give input on usability aspects and are available for testing IT components and assessment of the developed knowledge and work processes. In addition, they are the linking pin between the programme and their departments of facility on the level of the user.

At start, the SIG will be mainly formed by researchers and data managers involved in the pilot projects. During the programme, more members will be included and multiple SIGs can be formed. Actual involvement of the SIG members on different deliverables will be specified in collaboration with the members during the programme and may differ for the different WP’s.
HSR Programme Proposal

Pilot projects

During the design of the programme, four existing UG/UMCG projects joined the initiative and delivered input on the functional design of the infrastructure and service. The formulated requirements were discussed with other members of the SIG, partners and facilities, like RDO, UB/CMB and CIT. Involving these pilot projects ensures connection with researchers and provide concrete requirements for the specific fields. The first pilot projects are: Energysense (FWN), Enact (Medical Imaging, UMCG), KiVa (Social Sciences, BSS), and Personalised Medicine (Oncology, UMCG). During the programme many other projects, irrespectively of the size of the project, will be connected and involved. Interested researchers can contact the programme manager for this during the programme. Specific development requests will be discussed with the Programme Board. Actual use of the infrastructure may give raise to exploitations costs for the specific projects (for instance data archiving costs).

Scope

The programme will focus primarily on research data deriving from research involving human subjects. Business and administrative data and data from e.g. employees are outside the scope of the programme. Also, data from non-human subject research will not have priority, but may be included later. The primary users are researchers from UG/UMCG. Making the service and infrastructure available to researchers and clients outside Groningen or in the private sector will follow in another stage of the programme.

Innovation

The infrastructure and service developed in this programme will start with the existing tools, structures and expertise that are already present within the UG/UMCG. Some research data requirements ask for larger adjustment or for developing new infrastructures or tools. In particular this applies to streaming data technologies. In contrast to handling batch data, streaming data requires a different IT structure. This type of technology and its value for Human Subject Research will be explored further together with the department of Computer Sciences (UG). New technologies and tools will be developed with users and will be carefully piloted and tested before being implemented.

Working method

The working method will be agile where possible, by creating functioning building blocks throughout the 2-year programme. In that way, the expertise and IT structures can be tested by users and evaluated along the way. It also allows the programme to keep on track with the developing legislation, methodologies, insights and IT innovations. The set-up of the infrastructure and service will be scalable. At the start of the programme, the programme will focus on the four pilot projects and on development. During the programme more projects will join and the activities of part of the Programme Team will shift towards delivering services. This means that team members with functions in the data services (e.g. data consultants, data scientists, functional application manager, legal advisors) first have an important role in the development of procedures and the infrastructure and in a later phase of the programme in delivery of service.
Special attention will be given to align working methods of different groups. This will be one of the tasks of the WP leaders. Technologies and tools used in development will be shared over the WP’s and between the participating institutes and facilities. Also, differences in work culture will be considered. From the start, the creation of a shared identity and goal within the programme will be given attention and will be supported.

**Work packages**

A number of work packages (WP) have been defined (Figure 3). The definition of the WP’s follows the data flows when performing research and the general data lifecycle. Already at initiation of the research, the complete data life cycle will receive attention and be part of the intake (WP1). WP4 will deliver the underlying IT environment that enables WP1, 2 and 3. A separate WP is defined for the programme coordination and sustainable set-up of the actual services (WP5). The details of the WP’s and the deliverables are described in the separate chapters.

![Work packages diagram](image)

**Compliance to FAIR data**

To improve and facilitate the re-use of scientific data a concise set of principles have been defined that may act as a guideline: the FAIR Data Principles³. In the FAIR data approach, data should be: Findable, Accessible, Interoperable and Reusable. The service and IT infrastructure developed by the programme aims to enable these principles and support the researcher in making their research data FAIR. Table 1 describes how the programme will enforce these principles.
HSR Programme Proposal

Table 1. FAIR principles and corresponding programme output (between brackets the workpackage). A detailed overview of the deliverables of the programme can be found in Chapter 13 to 17.

<table>
<thead>
<tr>
<th>Principle</th>
<th>Description</th>
<th>Programme output</th>
</tr>
</thead>
</table>
| Findable  | Easy to find by both humans and computer systems and based on mandatory description of the metadata that allow the discovery of interesting datasets | ● Register (WP1) and online data catalogue, linked with persistent data identifier, meta-data and publications (WP3)  
● A service to consult on available data and data access policies (WP1)  
● Data capture and integration environment (WP2) |
| Accessible| Stored for long term such that they can be easily accessed and/or downloaded with well-defined license and access conditions, whether at the level of metadata, or at the level of the actual data content | ● Workspace solution for online data access from inside and outside UG/UMCG with differentiating access rights and authentication (WP4)  
● Technological infrastructure with connection to data archive, HPC facilities and ‘data highways’ (WP4) |
| Interoperable | Ready to be combined with other datasets by humans as well as computer systems | ● Integration and TTP function (WP2)  
● Data harmonisation and linkage possibilities (WP3) |
| Reusable  | Ready to be used for future research and to be processed further using computational methods. | ● Service portal with expertise and support on data access policies and access and IP agreements (WP1)  
● Data archive for long-term storage (WP4), searchable via online data catalogue and request workflow and data warehouse (WP3) |

5. Partners

Internal stakeholders

Within the UG and UMCG, collaboration is being established with relevant facilities, functions and aligning initiatives, like the Research Data Office (RDO), the Clinical Research Office (CRO), The Data Portal Project at the faculty of Economics and Business, BiKE (UMCG), the legal office of UMCG and UG (ABJZ), the Functionaris Gegevensbescherming (FG) of UMCG and UG, the Security Managers of UMCG and the UG, the University Library (UB), the Central Medical Library and the department Research & Valorisation of the UG. The Programme Manager informs and connects these stakeholders. The programme will stay open for other, yet unidentified, collaborations or connections. One of the goals of the programme is to connect existing expertise in the field of data management and support. Therefore the programme will actively strive to create connections and bring together expert groups and people in that field. Connection with the Boards of faculties, UMCG and university will be the responsibility of the Programme Board. Connection with researchers is organised via the SIG.
**HSR Programme Proposal**

**IT facilities**

IT development and support will be delivered by the existing IT facilities of the UG/UMCG, namely Research and Innovation Support (RIS) department of the CIT, the Genomics Coordination Centre (GCC) at UMCG, the Trial Coordination Centre (TCC) at UMCG, the IT service of the University Library and the IT instrumentation facility at the department of Behavioural and Social Sciences. Also, collaboration is sought with other expert groups, like the Distributed Systems and Data Science & Systems Complexity (DSSC) at the Faculty of Mathematics and Natural Sciences. These expert groups and facilities are connected through active participation in the different WP’s and deliverables of the programme. The programme manager will have regular meetings with the manager of the relevant IT departments and meet with the security managers and FG’s to ensure proper alignment and monitor collaborations.

**Research Data Office (RDO)**

In 2013 the board of the UG initiated the project “Inrichting RUG Research Data Office (IRRDO)”, as part of the three-pronged approach motivated by the recommendation from KNAW’s Schuyt committee on ‘Responsible research data management and the prevention of scientific misconduct’. End 2015, the end report of the IRRDO project was published and RDO received funding to continue the activities. The RDO has become the first gateway for UG researchers with questions on research data management, by starting an informative portal (www.rug.nl/research), and developing a network and setup collaboration of existing services within the UG. RDO delivers support with research data management plans and with publishing data via DataverseNL. The RDO created awareness university wide on research data management and put the topic on the agenda.

The RDO has been involved in the development of the programme proposal for Human Subject Research from the start and opportunities for collaboration have been identified at different aspects of the programme. The Human Subject Research programme will implement a number of the needs identified by the RDO, such as an easy accessible robust data infrastructure for sensitive data, a register for studies collecting personal data, a network of data managers and training for data managers on privacy and security aspects. During the development of these deliverables, the RDO will be consulted and collaborate as one of the prime stakeholders. The details on the collaboration and embedding the services and tools of the Human Subject Research programme will be worked out during the programme together.

The RDO will remain the central point of contact for research data management for UG researchers. In case of research involving human subjects, the RDO will offer tools or expertise developed in the programme or refer to the service of the Human Subject Research, depending on the type of question and the timing. The data consultant that participate in the Human Subject Research programme will collaborate with the RDO teams in delivering service for researchers and warrant continuous exchange of knowledge and experiences on research data management. Specific consultancy on research data management plans, meta-data, publication management, persistent identifiers, the RDO will be the prime partner for the Human Subject Research programme. Local or national policy development on Research Data Management and collaboration at the national and international level (VSNU, Landelijk Coordinatiepunt Research Data Management, Research Data Alliance) is part of the task of the RDO.

**Clinical Research Office (CRO)**

The Clinical Research Office is the service portal at UMCG for researchers in clinical research, which includes research on human subjects, or on data or samples of human subjects like clinical trials, biobanks, research on biomaterials, genetic research.
**HSR Programme Proposal**

The CRO develops and manages the UMCG Research Register and the UMCG Toolbox for clinical research with standardised tools, a quality system for clinical research including procedures and protocols, for instance for data management and storage to support research. Also, the CRO organises training for researchers within the UMCG and organises compliance visits. The aim of the CRO is to professionalise clinical research at UMCG and increase the quality and safety of the research by delivering support.

The CRO has a large network of experts and facilities that can be referred to, including IT facilities, but does not deliver the IT services itself. Partners involved in the development of the CRO tools are actively involved in the programme for Human Subject Research from the start. Together with the CRO opportunities for collaboration have been identified at different aspects of the programme. For instance, the Human Subject Research programme and CRO will collaborate in further development of the research register to make the tool available for non-clinical research. Also, on the implementation of legal and privacy aspects and development of tools for a Privacy Impact Assessment the Human Subject Research programme and CRO will join forces.

The CRO and Human Subject Research programme will align activities during the programme where useful and collaborate in engaging clinical researchers. The CRO will remain the central point of contact for research support for the UMCG researcher. In case of request for IT support for research involving human subjects, the CRO will refer to the service of the Human Subject Research.

**National and international partners**

National and international outreach and collaboration will be sought with SURF, the national coordination point for research data management (LCRDM), research data repositories (e.g. DataverseNL, DANS), BBMRI-NL/BBMRI-ERIC, HEALTH-RI\textsuperscript{11}, data4lifesciences\textsuperscript{10}, the eScience Center, DTL, the National Data infrastructure for Social Sciences (NDSW), and other (inter)national initiatives and organisation that are active in this field (e.g. EUDAT, KNAW) (see also Appendix 4).

**Participants**

Nowadays, participants in research (the ‘human’ in Human Subject Research) request more insight in their data that is being collected and used for research. Ways to translate this into practical tools (like a secured portal for participants) are explored in the programme. Also, the suitability of a (virtual) consultant or advisory group will be explored. For this the programme will collaborate with existing initiatives on data donation and data ownership like the Dutch Patient Federation (NCPF), Tippiq and the Social advisory board of BBMRI-NL.
HSR Programme Proposal

6. Rationale and return on investment

Alignment with strategic goals

The programme contributes to the central aim of the UG to perform excellent research that leads to scientific breakthroughs and societal innovation, by providing the researcher the up-to-date IT infrastructure that is indispensable to support disciplinary and cross-disciplinary research. Furthermore, the programme will support the researcher to align with the demands from editors of scientific journals and funding organisation on data quality and good data stewardship. It will help the researcher to implement and practise the principles of Open Sciences and Open Data and stimulate cross-disciplinary and transparent research.

Like the other faculties working with human data, the UMCG faces the challenges with human data, in particular patient data. The recently approved “Research IT policy” of the UMCG is aligned with the current Human Subject Research proposal. In addition, the UMCG has to cope with patient data for research which has additional requirements. This expertise will be available for the Human Subject Research programme.

Rationale for the programme

Currently, the collection, processing, storage and analysis of research data in the field of human subject research is merely done by researchers, in particular postdocs and PhD students. Only large studies, like TRAILS, Energysense, Kiva, have the possibility and resources to include specialists on IT and data management in their teams and to use a professional IT infrastructure. The programme aims to make this expertise and IT components available to a large range of researchers, including small data collections and create new collaborations.

The current programme will benefit the UG/UMCG and help the institutes to:

- increase the opportunity for collaboration between researchers, institutes, and private parties, by delivering a robust and safe IT infrastructure that can handle and link different data sources. Tools for identity and access management makes it possible to selectively give access;
- make the UG/UMCG reliable partners for data driven projects and increase the availability of new or large amounts of data and funding possibilities for researchers;
- become more time-efficient and increase research output, by offering researchers support and an up-to-date data environment;
- find new ways for researchers to use research data, to create insights and to develop new research questions;
- comply with the new regulations of funding bodies and scientific journals with regard to data management and assists researchers in these demands;
- comply with requirements with regard to Open Sciences, Open Data and re-use of research data;
- comply with legislation and prevent the occurrence of privacy breaches (“data lekken”), which may cause reputational and financial risks;
- prevent inaccurate handling of data and shortcomings in data archiving, which may lead to reputational damage and may harm public trust in the institutes and science in general.
Alternative options

The proposed approach is to develop this IT infrastructure and service together with the faculties, institutes and facilities within the UG and UMCG jointly. This way, existing expertise and tools can be reused and exchanged and new can be created in collaboration. In addition, a joint service improves good visibility and can deliver a high service level, because of the large scale approach and economy of scale.

An alternative approach could be to leave the initiative to arrange the IT infrastructure with the individual researcher or the faculties. However, an individual researcher or research project does not have the time and resources to setup an integrated and safe IT infrastructure and to allocate the expertise to successfully adhere to today's requirements with regard to data management. This hampers the researcher to use innovative IT solutions and data analytics, that could push forward new insights. Moreover, individual researchers will not be able to comply with the increased demands on data from journal editors and funding bodies, resulting in less (high ranked) publications and less research grants for the UG/UMCG.

Quality and performance indicators

At the end of the programme the deliverables described in the WP’s will be produced and be available for researchers. The progress of the deliverables will be monitored by the Programme Board. However, the general success of the programme will depend on the actual use of the service and IT infrastructure by researchers. The programme will be regarded as successful when the following milestones are met:

- The four pilot project are offered all the necessary tools and services to make their data FAIR. The researchers feel the service has matches with their demands.
- At least 10 other research projects from the different participating faculties have consulted the service and use(d) the services.
- The research register contains all the relevant information and documentation on these 14 research projects and the register is incorporated in the research process, so all new studies that contain personal data are reported and included in the register from May 1 2018.
- An online catalogue is available for UMCG and UG research data and contains the characteristics of minimal 20 studies with human subject data.
- In every participating faculty or facility, at least one contact person for Human Subject Research is established.
- The service desk in well-known by researchers. A survey among UMCG and UR researchers shows that 75% of respondents knows about the service and tools of the Human Subject Research programme.
- All persons with a data support function connected to the programme are trained on relevant aspects of data handling and data management (e.g privacy, integrity, functionalities).
- There is a visible and regularly updated product portfolio with tools, expertise and training from UMCG and UG for researchers.
- The service is sustainable embedded within the institutes with a suitable and effective governance, business model and financial resources.
- The responsible parties have insight in the required information (for instance on amount and volume of use of the service and the IT infrastructure, satisfaction of users, recurrent customers, amount of additional funding).
## Risk analysis

The following risks are defined in Table 2 and for every risk a counter measure is described. These risks will be regularly monitored during the programme by the programme manager and reported to the Programme Board. This risk monitoring is part of the Plan-Do-Check-Act cycle. In case the counter measure appears not to be sufficient, the risk situation remains and hampers the success and progress of the programme, the Programme Board will propose an adequate intervention and can decide to adjust the programme with regard to deliverables, planning or implementation. This will be communicated with the boards of the UG and UMCG.

### Table 2. Risks and counter measures

<table>
<thead>
<tr>
<th>Nr</th>
<th>Cause</th>
<th>Consequence</th>
<th>Type of risk</th>
<th>Counter measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Researchers are not consulted and are not optimally connected in the development of the functional design of the services, infrastructure and tools.</td>
<td>The developed services and tools do not fit with the needs of researchers. The service and infrastructure are not used by researchers.</td>
<td>The service and infrastructure will not be sustainable.</td>
<td>Researchers are involved in every step in the development of the tools and services from the start of the programme via pilot projects, use cases and the SIG group. The development and work process is iterative. Researchers have position at crucial point in the programme governance.</td>
</tr>
<tr>
<td>2</td>
<td>Faculty and university management, policymakers and other stakeholders in UG/UMCG are not consulted and are not optimally connected in the development of the services, infrastructure and tools.</td>
<td>The developed services and tools do not fit with the needs and processes of the management of the UG/UMCG. The service and infrastructure are not supported by the faculties and boards of the UG/UMCG.</td>
<td>The service and infrastructure will not be sustainable.</td>
<td>The programme will collaborate with policy makers and stakeholders of UG and UMCG on data management topics. The faculties and the Office of the University collaborate in the programme and have position in the programme governance.</td>
</tr>
<tr>
<td>3</td>
<td>The service and infrastructure are too expensive for researchers and faculties.</td>
<td>The service and infrastructure are not used by researchers.</td>
<td>The service and infrastructure will not be sustainable.</td>
<td>WP5 will develop a sustainable business model together with researchers, the financial departments of UG/UMCG and other stakeholders. The service will start small and scalable. The programme board will take care of sustainable governance and embedding in the institutes.</td>
</tr>
<tr>
<td>4</td>
<td>Differences in work culture, habits and interest at the different facilities in UG/UMCG interfere with successful collaboration during development of the infrastructure.</td>
<td>The developed tools and services are not well connected and do not work. The ‘silo’s of services’ remains. The service and infrastructure are not used by researchers.</td>
<td>Delay in delivery of programme output. Inefficient use of resources.</td>
<td>In the programme organisation close collaboration and exchange is established by creating multidisciplinary teams with employees of the different facilities. The programme organisation is located on one physical location for part of the week.</td>
</tr>
</tbody>
</table>
HSR Programme Proposal

<table>
<thead>
<tr>
<th>Nr</th>
<th>Cause</th>
<th>Consequence</th>
<th>Type of risk</th>
<th>Counter measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Shortage of the required human resource capacity. Vacancies cannot be filled, or persons are needed for other projects.</td>
<td>The deliverables cannot be met according to planning.</td>
<td>Delay in delivery of programme output.</td>
<td>The persons connected to the programme work in dedicated teams. The programme will connect with the IT and service facilities of UG/UMCG, to be able to request for additional human resources. The Programme Manager will monitor progress of the programme closely and contact the line manager in case of conflicting activities.</td>
</tr>
<tr>
<td>6</td>
<td>An overload of interest and requests for the services of the programme.</td>
<td>The service cannot meet the request from researchers. Development resources are required for operational tasks.</td>
<td>Delay in delivery of programme output.</td>
<td>The extra demands for service will create additional income, which will enable the infrastructure to scale up. The requests will be monitored closely to react on time and adjust capacity. National outreach will start when the Groningen customers can be served successfully.</td>
</tr>
<tr>
<td>7</td>
<td>The deliverables of the programme are dependent and not well aligned. The programme is dependent and not well aligned with other projects and services in UG/UMCG.</td>
<td>The deliverables cannot be met according to planning.</td>
<td>Delay in delivery of programme output.</td>
<td>The dependencies within the programme and with other projects in UG/UMCG are identified and discussed with those involved. The WP leaders and Programme Manager monitor closely the progress of the deliverables and alignment with the dependent factors.</td>
</tr>
<tr>
<td>8</td>
<td>The pilot projects and connected researchers do not have time to adequately deliver functional input or test tools.</td>
<td>The tools cannot be developed or the developed services and tools do not fit with the needs of the researcher.</td>
<td>Delay in delivery of programme output.</td>
<td>The pilot projects are existing projects and have already started. The expected workload has been discussed with the pilot projects. In case of substantial delay, new project will be sought within the SIG network.</td>
</tr>
</tbody>
</table>

7. **Timeline and planning**

The programme will run for 2 years. At start of the programme, a separate project implementation plan will be developed for the deliverables of the WP’s. The deliverables of WP1 will be produced first and become available for researchers within the first year of the programme (with priority to deliverable WP1.1, WP1.2 and WP1.4) (see also Table 3). Also, the portfolio of existing tools and services (WP5.1) and the training for data consultants and data managers (WP5.5) will be delivered in the first phase of the programme. In general, in the first 6 months of the project the current IT infrastructure will be prepared and the recruitments of new tools (e.g. the data archiving tool) will be investigated, evaluated with users (starting with the pilot projects), and tested.
HSR Programme Proposal

Table 3. Overall planning and timing of realisation of deliverables (a detailed description of the deliverables can be found in Chapter 13 to 17).

<table>
<thead>
<tr>
<th>Deliverable</th>
<th>Year 1</th>
<th>Year 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Q1-2</td>
<td>Q3-4</td>
</tr>
<tr>
<td>WP1</td>
<td>WP1.1</td>
<td>WP1.3</td>
</tr>
<tr>
<td></td>
<td>WP1.2</td>
<td>WP1.6</td>
</tr>
<tr>
<td></td>
<td>WP1.4</td>
<td>WP1.8</td>
</tr>
<tr>
<td>WP2</td>
<td>WP2.1</td>
<td>WP2.2</td>
</tr>
<tr>
<td></td>
<td>WP2.3</td>
<td>WP2.7</td>
</tr>
<tr>
<td>WP3</td>
<td>WP3.1; WP3.2</td>
<td>Go/no go</td>
</tr>
<tr>
<td></td>
<td>WP3.16</td>
<td>WP3.13</td>
</tr>
<tr>
<td></td>
<td>WP3.19</td>
<td>WP3.17</td>
</tr>
<tr>
<td>WP4</td>
<td>WP4.5</td>
<td>WP4.1</td>
</tr>
<tr>
<td></td>
<td>WP4.8</td>
<td>WP4.2</td>
</tr>
<tr>
<td></td>
<td>WP4.9</td>
<td>WP4.3</td>
</tr>
<tr>
<td>WP5</td>
<td>WP5.1</td>
<td>WP5.5</td>
</tr>
<tr>
<td></td>
<td>WP5.2</td>
<td>WP5.6</td>
</tr>
<tr>
<td></td>
<td>WP5.3</td>
<td></td>
</tr>
</tbody>
</table>

8. Finances and resources

Investment

The programme will be funded by the UG ICT strategy fund (ICT meerjarenplan) matched with funds from the UMCG Research IT budget and completed with in kind contribution for the non-IT activities from UMCG, ABJZ, CIT and the University Library (Table 4). The current proposal covers the development costs of the programme: 527k in year 1 and 583k in year 2. These costs cover the majority of the personnel costs and resources that are needed for the deliverables of the programme as described in the five work packages. The programme will use the available technical infrastructure within UG and UMCG, so this is not included in the budget. For details see Budget in Appendix 6.

Operations costs

During the programme the infrastructure will be used by the pilot projects and the first users as a test version. The technical infrastructure needs to be maintained and adequately staffed. These exploitation costs, 83k in year 1 and 83k in year 2, are covered by the programme.

The long-term exploitation costs (after the 2 years programme) depend on the tools and service that will be developed, on the number of users of the infrastructure and the size and complexity of the studies. These exploitation costs will be monitored closely during the programme. To keep the service portal and IT infrastructure operational, maintained and managed after 2 years, an annual budget of 200k (83k plus coordination costs and extension of data consultancy tasks) is estimated. This estimation does not include the use and service costs of the technical infrastructure for individual research projects (e.g. costs for use of applications, storages, network). These will be covered by individual research projects, based on actual use and with an adequate service level.

It is crucial to create a sustainable service and infrastructure for the institutes and researchers. For that reason long-term funding of the programme and development of a cost model are part of the programme (deliverable WP5.4).
Funding

The current proposal submitted to the ICT strategy fund concerns Recommendation 4: “Data in mensgebonden onderzoek.”

In addition, part of the budget from Recommendation 15: “Research data (management)” will be, in time, requested together with the University Library, to deliver the output that are mandatory for the UG to become compliant to the Data Protection Law (details in Appendix 2). Within the Human Subject Research programme IT tools and support will be developed, implemented and managed together with RDO and ABJZ to enable researchers from UG and UMCG to meet the requirements of data management. This includes amongst others the set-up of a register for studies with personal data (WP1.4 and WP2.3), develop expertise and a methodology for a Privacy Impact Assessment (PIA) (WP1.2), a data classification methodology and tool using the principle of Privacy by Design and FAIR data (WP1.1 and WP2.2) and privacy enhancing technologies (WP2.4).

The development of the IT infrastructure for Human Subject Research and successful design and implementation of the deliverables of WP4 heavily depend on other IT innovations included in the strategic ICT plan. This concerns in particular Recommendation 3 “Management en archivering van onderzoeksdata”, Recommendation 8 “Virtual Research Environment (VRE) from the section ‘ICT and Research’ and Recommendation 31 Identity and Access Management (IAM) from the section ICT and infrastructure. ‘Human subject research’ or the aims of Human Subject Research programme are explicitly mentioned in these Recommendations. The deliverables defined in the Human Subject Research programme are maximally aligned with the deliverables in these Recommendations.

The complete implementation of Recommendations 3, 8 and 31 is however beyond the scope of the current proposal and asks for involvement of other stakeholders. To optimize collaboration and alignment with these Recommendations and not delay the implementation of the Human Subject Research programme, the programme includes three part-time developers for these topics (funded by Recommendation 4). These persons will be appointed at the CIT departments that will have a major role in execution these Recommendations, can serve as forerunners and prepare for the larger project. This way the expertise and tools developed in the Human Subject Research programme are preserved and can be easily reused and developed further.

<table>
<thead>
<tr>
<th>Source</th>
<th>Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICT strategy fund UG</td>
<td>1016K (from Recommendation 4)</td>
</tr>
<tr>
<td>ICT strategy fund UG</td>
<td>260K (from Recommendation 15, for deliverables WP1.1, 1.2, 1.4, 1.5, 2.2, 2.3, 2.4) [to be requested in time]</td>
</tr>
<tr>
<td>UMCG Research IT</td>
<td>6.9 fte in year 1 and 5.7 fte in year 2 (in total 1094K)</td>
</tr>
<tr>
<td>ABJIZ</td>
<td>In kind legal expertise and support (0.4 fte + 1.0 fte 2 years) 25K training</td>
</tr>
<tr>
<td>University Library</td>
<td>In kind information management expertise and support (fte’s to be determined)</td>
</tr>
<tr>
<td>Central Medical Library, UMCG</td>
<td>In kind information management expertise and support (fte’s to be determined)</td>
</tr>
</tbody>
</table>
HSR Programme Proposal

Alignment to UG and UMCG ICT strategy

The programme can start by using the current available technical infrastructure (the spare infrastructure at the collaborating IT facilities and the infrastructure already available for the pilot projects). However, to be successful and sustainable, the Human Subject Research programme depends on a number of technological innovations and investments included in the UG ICT strategy.

In particular this concerns the following recommendations:

- Recommendation 2: Data handling
- Recommendation 3: Management en archivering van onderzoeksdagta
- Recommendation 6: Nationale en internationale e-infrastructuur
- Recommendation 8: Virtual Research Environment (VRE)
- Recommendation 15: Research Data (Management)
- Recommendation 31: Identity and Access Management (IAM)
- Recommendation 28: Datacenters en datacenter netwerk
- Recommendation 32-37: Security

The Human Subject Research programme therefore intends to work closely together with the projects that deliver these recommendations and align with their planning. The share of the newly developed technical infrastructure from these recommendations will be based on the actual use and capacity needed by the Human Subject Research programme. Since the programme involves a number of scientific disciplines and engaged researchers, user input and use cases for these other recommendations will also be arranged via the Human Subject Research programme.

The programme manager will have regular meetings with the Manager Research and Innovation Support, Manager Server and Infrastructure, Manager Workspace and Middleware Services and the Security Manager of the CIT to ensure proper alignment and monitor dependency.

Connection and alignment with the UMCG research IT strategy is ensured through the active involvement of all three members of the daily board of the UMCG Biobank/Research IT Platform (BRIP) in the Human Subject Research Programme.

9. Research and funding opportunities

Although own research activities are not within the scope of the Human Subject Research programme, the programme aims to identify funding possibilities and create new collaborations for innovation of the infrastructure (deliverable WP5.7). This is part of the sustainability approach (WP5). Applications to infrastructure funds (like NWO and H2020-INFRA) will be submitted from the Human Subject Research programme itself, whereas applications to research funds (both computer sciences as well as life sciences) will be submitted by scientists, supported by the Human Subject Research programme. Such applications will include data support which is handled by the Human Subject Research programme.

Details will be developed in the business model for the programme (deliverable WP5.4). For this a funding officer/business developer will be appointed in WP5. With these funding initiatives, the programme creates new research opportunities for UG and UMCG and keeps the infrastructure innovative.

Already during the preparatory phase of the programme new research opportunities have been identified. By the end of 2016 a PhD project will start that will support WP1: “Legal aspects of data collection by eHealth”. This is jointly supervised by Jeanne Mifsud Bonnici from the Faculty of Law and Ronald Stolk from the Medical Faculty. Concrete plans have been drafted for a proposal on streaming data technologies to the upcoming call within “commit2data” from NWO-EW between departments of Computer Sciences (FWN) and Genetics (UMCG). Moreover, discussions have started with the eScienceCenter and ZonMw to support infrastructure projects in health data, as part of Health-RI.
**HSR Programme Proposal**

10. **Programme control and reporting**

The steering committee and programme manager supervise the overall coordination and execution of the programme and report to the Programme Board on progress of deliverables, planning and budget every board meeting. Twice a year a progress report is submitted to the Programme Board and to the main funding bodies (Table 3). At half-time a go/no-go moment is introduced, at which the Programme Board decides based on the progress reports whether to continue with the programme or not. Changes in the direction of the programme, budget and planning will be submitted to the Board for feedback and approval. At start of the programme, the WP leaders will develop a Project Plan with detailed planning for the deliverables of the WP. In some cases, a Project Plan per deliverable will be developed. The Project Plans will be defined in collaboration with researchers and will be shaped according to the requirements expressed by the researchers. These project plans (based on the CIT Templates) will contain concrete SMART deliverables, planning, names of team members, stakeholders, etc and be reviewed and approved by the Programme Board.

The programme will work according the Plan-Do-Check-Act cycle and implement agile working method where possible, to ensure continuous evaluation of the service and IT components with users and stakeholders. Biannual external audits will be organised for security and privacy compliance. The outcomes of these audits will be shared with the Programme Board. The programme will be supported by the Project Controller of CIT.

11. **Communication strategy**

Internal communication (within UG and UMCG) will be aimed at informing researchers about the service, getting a clear picture about the needs of researcher, and aligning with stakeholders and other facilities. Questions that arise from the research data management plans of via the current institutes or services are important input for the communication strategy and content. Regular meetings and workshops will be organised for researchers and research support. A website will be set-up that will serve also a central point for dissemination of expertise and best practices.

Visibility of the Human Subject Research service outside UG and UMCG is important to attract new users, create collaborations and PR for UG and UMCG. The programme will collaborate with the communication officer of the CIT and the UG to develop and implement the communication strategy. For internal and external communication the programme will work closely together with the communication departments of the UG and UMCG.
12. References


13) KNAW (2013). Responsible research data management and the prevention of scientific misconduct

13. WP1: Preparation and design

Context and aim

The aim of this WP is to access the needs of the researcher and the recruitments of the research and the data, and translate these needs and recruitments to concrete solutions, measures and tools. Each researcher will start with an intake. This can start at any time during the research data cycle. In some cases the intake will take place before the research data collection starts, while in other cases, it concerns reuse of existing data or data archiving. For the intake an online data classification tool will be developed that is accessible for all researchers that translate requirements to concrete measures and IT solutions. The outcome of the intake will be recorded in a register, to document all (new) data collection, together with Privacy Impact Assessment and security reports or other documents connected to the data (e.g. Bewerkersovereenkomst). This WP will work closely with other existing initiatives within UG and UMCG which have a connection to this programme and connect with initiatives outside the organisations, like BBMRI-ERIC, to share expertise.

Connection with other WP’s

The information architecture is input for the infrastructure and tools developed in WP2 and WP3. The register and classification tools is hosted in WP4. The service that organises and executed the intake is set-up in WP5. WP5 valorises the tools and expertise developed in WP1.

Deliverables

WP1 will deliver the expertise and tools necessary to perform an holistic intake of the data request and identify the necessary IT tools, IT architecture and organisational and contractual measures needed. This intake results in a completed data management plan and a list describing the concrete requirements of the IT infrastructure for the study.

The required security expertise will be contributed by CIT and the participating institutes. In collaboration with CIT, a separate proposal will be submitted to the ICT strategic commission concerning security (ICT strategy fund, Recommendation 32 to 37).

<table>
<thead>
<tr>
<th>Nr</th>
<th>Deliverable</th>
</tr>
</thead>
<tbody>
<tr>
<td>WP1.1</td>
<td>Input for online data classification and information analysis tool, with decision tree to support the end-user (researchers, data managers) in their selection process to decide which data management application/tool is suitable for specific research purpose. The tool classifies the data collected (new data or existing data) on sensitivity and availability needs into specific categories, and guides towards the optimal concrete IT components, measures and protocols to guard confidentiality, security and integrity of the data. The outcome of the assessment can be documented. The tool will build on the expertise and tooling of the data classification tool from CRO toolbox, the RDMP tool and other relevant developments in this field (e.g data-tags) and is technically developed in WP2.</td>
</tr>
<tr>
<td>WP1.2</td>
<td>A protocol, template and method to perform an initial risk analysis (including a Privacy Impact Assessment and a security assessment). The risk analysis protocol and method are developed and tested using the existing IT components that are connected to the Human Subject Research data infrastructure and build on existing risk analysis expertise and tools. For example the PIA and security assessments from SURF and from Hanze) and Norea</td>
</tr>
</tbody>
</table>
The IT components connected to the Human Subject Research data infrastructure are labelled according the data classification from perspective of guidelines (law and other regulations that apply to research data). A differentiated list of measures and IT components to be able to translate the information flows and outcome of the data classification and risk analysis (e.g. PIA) to concrete IT solutions and measures (protocols, work processes, communications, contracts, etc). Missing links and improvement are identified. Input for the deliverable are the IT portfolio (output in WP5) and protocol developed in WP1.2.

WP1.4 Input for a register (functional design) for all reported research data collections, with allocation of a persistent identifier and documentation of the PIA and security reports study level. In line with Article 30: Register with handling and processes of personal data. The register can be accessed (specified by request and function) by different parties and stakeholders in the data management process, for instance ethical bodies, the FG, the faculty board. It is a living document with analysis of rights on re-use of data and aligned with the developing standards and regulations. The tool will build on existing tools and expertise in UG and UMCG, connect to existing business information (BI) tools and is technically developed in WP2.

WP1.5 Develop and implement a methodology to use a persistent identifier for research data integral in the Human Subject Research infrastructure along the complete data life cycle (from register, to data capture environment, data catalogue, data archive and publication database) and link it to international standards (e.g. DOI, EPIC). This will be done in close collaboration with the University Library and RDO.

WP1.6 A Library with Best Practices, Template documents and protocols for data management. For instance a Template Data Access Policy, an Authorisation policy, a Data Access Agreement, a Data Security Plan and a Data Processing Plan (bewerkersovereenkomst), etc. Template texts for grant applications and data management plans.

WP1.7 A Library with advice, guidelines (SOPS) and best practices for data capture solutions in case of new data collection, and on data harmonisation and linkage solutions in case of new and existing data collection and on data processing. Aligned with outcomes of the information analyses, and privacy and security assessment. On user request, standards will be re-used and made available, or newly developed across the data capture platforms. Input for the Library will be provided by WP2 and WP3.

WP1.8 A network with (internal and external) experts who can be consulted in case of complicated privacy and security issues and who can perform peer-PIA’s and security audits. This expertise is made visible in WP5 (consultancy) toolbox and offered to external parties.

Team

<table>
<thead>
<tr>
<th>Function title (see Appendix 3 for function description)</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>WP leader</td>
<td>Coordinates overall WP, reports to programme manager, aligns tasks and resources, monitors planning.</td>
</tr>
<tr>
<td>IT legal expert</td>
<td>Develop PIA and security assessment protocol and method. Develop template protocol and documents.</td>
</tr>
<tr>
<td>Information architect</td>
<td>Translate functional request from researcher and outcome of data classification to information and IT design.</td>
</tr>
<tr>
<td>Data consultant</td>
<td>Develop PIA and security assessment protocol and method and intake procedure in general. Perform data intake, coordinate PIA and security analysis. Deliver service for researcher. Built network of experts. We start with a data consultant with focus on legal aspects and for support for the legal IT specialist, a data consultant</td>
</tr>
</tbody>
</table>
**HSR Programme Proposal**

with library expertise and general data expertise. In the second year, scaling according to use and requests.

<table>
<thead>
<tr>
<th>Role</th>
<th>Responsibilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data manager</td>
<td>Model and load the data from the use cases and develop the data management templates and protocols. Assist users with use of the tools and infrastructure offered in the Human Subject Research programme</td>
</tr>
<tr>
<td>Data scientist</td>
<td>Support users to develop pipelines deploy their pipelines/analyses on the cluster</td>
</tr>
<tr>
<td>Security specialist</td>
<td>Advisor/reviewer (in kind contribution by CIT/UMCG)</td>
</tr>
<tr>
<td>Legal expert medical data</td>
<td>Advisor/reviewer</td>
</tr>
<tr>
<td>Medical information officer</td>
<td></td>
</tr>
</tbody>
</table>

**Partners and Stakeholders**

- Security Manager UG
- Security Manager UMCG
- Legal department UG
- Legal department UMCG
- FG UMCG
- FG UG
- Clinical Research Office UMCG
- Research Data Office
- University Library (Dienstenteam, Team O&O, Team Innovation and Development)
- Central Medical Library, UMCG
- Privacy werkorganisatie UMCG
- EPD information specialist, UMCG
**HSR Programme Proposal**

14. WP2: Collect, process and release

**Context and aim**

This WP takes care of the data capture process (creating data), the integration, validation and de-identification of data (processing data), it facilitates the data release process from one or more data source(s) and makes it ready for presentation in the catalogue. The collected data go through a process of de-identification, data quality monitoring and verification/validation steps. The processed and de-identified data are released and become available for data requests. The data processing and release procedure are automated and fully reproducible.

**Connection with other WP’s**

After the intake (WP1) the study protocol or blueprint needs to be translated to IT structures and tools. This requires for the connecting and further development of existing IT structures and tools. WP2 is the technical implementation of the requirements defined in WP1. WP2 prepares the data so it could be requested via the catalogue and data request workflow (WP3) where it enables researchers (and other end-users) to find and request research datasets. WP2 will depend on WP4 to provide suitable application servers, storage facilities, archiving and long-term facilities for preserving data. Deliverables of this WP (tools, procedures, methods, instructions) are deployed in the service portal (WP5).

**Deliverables**

*Data capture, data integration, data verification, data release*

Based on the data classification (WP1) the researcher or data manager selects an appropriate data management infrastructure from the product portfolio. Each tool in the portfolio (data capture, integration, verification, release process) consists of a description of functions, and the specific purpose for which it could be used. It also contains information about the available knowledge from UG and UMCG, training materials, workshops, contacts and criteria for getting access (i.e. pricing). It becomes possible to efficiently combine real-time streaming data (for instance energy data, Twitter readings) with batch data for processing, accessing and archiving. All data are properly stored, findable, and reproducible during these processes.

**Pseudonimisation (TTP)**

According to legal requirements, data from human subjects should be protected within the research datasets. Therefore all human identified information is removed from the dataset (anonymization) which makes the data not traceable to a unique person. Tracing of persons may be desirable if it appears that a participant requires an unknown risk, additional research is necessary (linking data) or to follow people in time (develop an endpoint such as illness, complications or death). For those situations the pseudonimization method is another option where identified information is replaced by a pseudo code by means of a trusted independent party (Trusted Third Party). Under very strict conditions, personal identifiers are backwards traceable.
<table>
<thead>
<tr>
<th>Nr</th>
<th>Deliverable</th>
</tr>
</thead>
<tbody>
<tr>
<td>WP2.1</td>
<td><strong>Data Management products:</strong> Deliver input for the product portfolio (WP5) containing already existing data management applications/tools (data capture, data processing) which are in-use at the UG/UMCG departments. The tooling shall first be inventoried, followed by identifying missing functionality and addressed at the right place (i.e. functional roadmaps).</td>
</tr>
<tr>
<td>WP2.2</td>
<td><strong>Data classification and information tool:</strong> Technical development of an online data classification and information tool including a decision tree, to support the end-user (researchers, data managers) in their selection process for a proper and adequate data management solution. The tools will align with UG and UMCG data classification protocols. WP1 is responsible for the functional requirements (WP1.1).</td>
</tr>
<tr>
<td>WP2.3</td>
<td><strong>Web based register:</strong> Technical development of a web based register (accessible via the internet) for all reported research data collections, with allocation of a persistent identifier, study documentation, PIA and security reports. WP1 is responsible for the functional requirements (WP1.4).</td>
</tr>
<tr>
<td>WP2.4</td>
<td><strong>Privacy enhancing technologies:</strong> Develop standards and IT solutions (e.g. data processing protocols, scripts, encryption, and measures) for privacy enhancing technologies. These solutions may be based on methods like k-anonymity and needs to be adjustable, so it is suitable for other data sources and situations.</td>
</tr>
<tr>
<td>WP2.5</td>
<td><strong>Data processing platform:</strong> Publish data capture data to end-user (researchers, data managers) via the catalogue (WP3) and research portal (WP3) by means of a data integration and harmonization solution (generic layer method). It includes functionality to configure the platform (data modelling, data mapping) and enable end-users to load historical versions of datasets. This data processing platform (all data domains) includes data integration functionality, de-identification steps (TTP), storage and release for streaming and batch data combined. It also enables the catalogue to present available data and allow to order datasets (subsets, counts, etc.)</td>
</tr>
<tr>
<td>WP2.6</td>
<td><strong>TTP service:</strong> Deployment of a pseudonimization (TTP) service (technique and process) and realize good price contracts with a supplier TTP. The service has a high degree of autonomy for researchers. Under very strict conditions, personal identifiers are backwards traceable. The milestone is in collaboration with the Mondriaan/UMCG initiative where linking data functionality by means of a TTP is developed and deployed.</td>
</tr>
<tr>
<td>WP2.7</td>
<td><strong>Streaming data:</strong> Explore and design a data capture environment to capture and process (real-time) streaming data in combination with batch data. Explore required tools and adjustment to existing tools.</td>
</tr>
<tr>
<td>WP2.8</td>
<td><strong>Documentation:</strong> Deliver input for the development of training materials, instruction manuals and procedures (SOPS) for the above-mentioned milestones.</td>
</tr>
</tbody>
</table>
**Team**

<table>
<thead>
<tr>
<th>Function title</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>WP leader</td>
<td>Coordinates overall WP, reports to programme manager, aligns tasks and resources, monitors planning.</td>
</tr>
<tr>
<td>Software/database developer</td>
<td>Make existing tooling and application ready for use cases. Implement configurable data workflow and deploy and develop interfacing with TTP service.</td>
</tr>
<tr>
<td>IT/technical architect</td>
<td>Model and design the overall technical architecture. Connect existing building blocks. Develop and ensure connections with WP1, WP3 and WP4 in close collaboration with information architect.</td>
</tr>
<tr>
<td>Expert on streaming data</td>
<td>Advise and develop technical infrastructure that can efficiently handle and combine streaming and batch data</td>
</tr>
<tr>
<td>Security specialist</td>
<td>Advice on development of new tools. Review and audit existing and new applications and tools (in kind contribution by CIT/UMCG).</td>
</tr>
</tbody>
</table>

**Partners and Stakeholders**

- Clinical Research Office UMCG
- Research Data Office
- University Library (Dienstenteam, Team O&O, Team Innovation and Development)
- Central Medical Library, UMCG
- Instrumentatiedienst GWM
- EPD information specialist, UMCG
- Pilot projects
- Security Manager UG
- Security Manager UMCG
- Collaborations with NFU D4LS, Mondriaan, PROFIT program UMCG
HSR Programme Proposal

15. WP3: Request, access and analyse

Context and aim
This WP enables researchers in following the FAIR data principles, by making the data findable, requestable, and promoting re-use. The WP delivers a warehouse to structure and code the data (pseudonymized data only), catalogues where we can make available all relevant human subject datasets (level 1-4), provide data request workflows for researchers to ask permission to a data access committee, and when provided access provide a study portal with tools to analyse and interrogate the data.

Connection with other WP’s
The data will be delivered via WP2 which takes care of data capture (creating data) and de-identification (processing data). WP3 will depend on WP4 to provide suitable application servers, storage, archiving for preserving data and high performance computing services to enable the technical access to the WP3 services. Authentication is also part of WP4. WP1 will provide functional support for the tools (first line), and WP4 will provide technical support for the infrastructure environment.

Deliverables

Analysing data - Research data warehouse
All de-identified data sets from WP2 will be loaded into a data warehouse. The data warehouse, based on for example Molgenis tooling, is connected to a ‘dataset creator’. This creates a dataset from the de-identified environment (WP2), based on characteristics specified by the researcher via the data catalogue. The data managers of the projects will be provided tools and templates to structure and code the data using international data stewardship standards in collaboration with the University Library, CRO and RDO. WP3 will ensure sharing of best practice templates across the research data providers.

Giving access - Catalogue
Meta-data and anonymous summaries of the data will be easily findable online and researchers will be able to view, search and browse the data that is available for research to evaluate whether the data are appropriate for their research. This makes the data ‘open’ for future use. The meta-data and the attributes (e.g. variable names) of the data collection are catalogued in combination with a persistent identifier of the dataset (e.g. EPIC, DOI). Who can use the data for what purposes is defined in a Data Access Policy and in the access procedures defined by the researcher at the initial intake.

Giving access - Request workflow
From the catalogue users can request (sub)sets which will start a data request workflow that supports interaction between researcher and the designated data access committee(s) for the requested datasets. Data access committees can configure what information must be provided by the researcher (e.g. research proposal, items needed, optionally associated samples needed). Optionally, users can use an ‘integrator’ to assist in harmonization of data from multiple studies for pooled analysis. The tailor-made dataset is made available in an environment with a lower security level. The requests or datasets for individual researchers are efficiently archived and can be restored or reproduced at all times.
HSR Programme Proposal

Re-using data / Analysing data - Research study portals

When research request is approved, data managers can give researcher access to the data subset using a research study portal. Here a subset of the data from the data warehouse is provided in the context of a study. In addition tools are provided to the researcher to analyse and interpret these data such as R statistics, scripts and pipelines (backed with the high performance computing services from WP4).

Researchers can here also add their own (reference) data. The results of a study can be again added to the warehouse so they can be reused in future studies. To enable big data analysis and interpretation support by data scientists is provided to assist researchers in use of this facility and pipeline development and installation of necessary software.

<table>
<thead>
<tr>
<th>Nr</th>
<th>Deliverable</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Public catalogue service (‘finding data’)</strong></td>
</tr>
<tr>
<td>WP3.1</td>
<td>Develop a data collection catalogue to find data collection so UG/UMCG research data collection owners can make their data findable. This should contain type of data, meta-data and the organizations/contacts behind the collections, including permissions and citations conditions. It should enable automated links to ingest data updates from source warehouses (in WP2) and automated advertisement to national catalogues, including 'stream' based data collections. This tool builds existing expertise and available tooling. Input from WP1 and WP5 for user interface and UG/UMCG styling. Integration with UG AAI (identification and authentication) is essential (deliverable in WP4). Collaboration with University Library on FAIR data and metadata standards.</td>
</tr>
<tr>
<td>WP3.2</td>
<td>Further develop and populate a data collection catalogue with the pilot projects and known UG/UMCG collections (summary level only). The data consultants (WP1) bring in users and data owners. The data manager(s) in WP1 assist the researcher in loading the data (i.e. select data attributes, map them to the standard data model, execute transformation, define how data will be maintained in the future). Also manuals and training will be developed so the data owners themselves can enter these data.</td>
</tr>
<tr>
<td>WP3.3</td>
<td>Expand the catalogue to enable counting of available data collection. Some collections may be willing to share a limited set of data per individual, e.g. 'data on Alzheimer' or 'GWAS available' which enables counting which is very relevant when searching for particular data. This catalogue expansion will enable these searches and facilitate an anonymity checker to ensure counts do not yield individual level data.</td>
</tr>
<tr>
<td></td>
<td><strong>Request service (‘giving access’)</strong></td>
</tr>
<tr>
<td>WP3.4</td>
<td>Develop a request form and request workflow (including status reports) to enable users of the catalogue to define a data request by selecting data items they are interested in and to enable data managers to define data items per collection and create tailor-made datasets based on requests (following the data access policy defined by the data owner). Multiple data collections can co-exist in the same facility and connect this to the catalogue.</td>
</tr>
<tr>
<td></td>
<td><strong>Data access service (‘reusing data/analysing data’)</strong></td>
</tr>
<tr>
<td>WP3.5</td>
<td>Data warehouse for data collections and data loaded from WP2: configure system to provided source datasets as basis for data access requests (linked to WP2 to enable receiving data after de-identification and pre-processing from WP2). This creates a connection to the processing environment with interface to WP2. This is the basis for the 'catalogue publication' of study data sets.</td>
</tr>
</tbody>
</table>
WP3.6 Develop a reproducible study data creator on basis of data item selection in catalogues that enables the researcher or data manager to automatically extract data configured in a data request from the source data collection and either download it as raw files or push it into a study portal (in collaboration with WP2 and WP4). On top of MOLGENIS we create a plug-in that takes a data request (data items + inclusion/exclusion filters) as input and then generates query to extract the requested data. Linkage of difference data sources and release via TTP will be developed together with WP2. Data sets that are delivered are reproducible (e.g. by archiving a copy or by storing the query on a versioned source).

WP3.7 Create a web-based study portal where researchers can easily access all data they received for their study. This service includes data explorer to quickly select and filter data and APIs for analysis using R (statistics) and python (for scientific algorithms). Data can be ingested automatically (as basis for dataset creator, below); researchers can also post back their (intermediate) results as basis for collaboration with colleagues, publication and/or archiving; the portal can be accessed from within a workspace (deliverable of WP4) or securely over the Internet in case of less sensitive data.

WP3.8 Develop a study to catalogue publishing tool, that enables study owners can push their results back into the catalogue as new (derivate) datasets in connection with WP2. It will create a plugin that can extract from a dataset a metadata record that can be published in catalogue.

WP3.9 Develop recoding and study pooling tool to mapping and harmonise different. This enables researchers and data managers to rapidly code their data to a self-defined or standard coding system (e.g. from free text to an ontology). In addition it enables users to rapidly transform multiple data sets on one common standard schema as basis for pooling (union) of data from multiple studies. This will build on the SORTA and BiobankConnect tool and includes a user interface to make it multi-tenant and creation of manual and training materials.

WP3.10 Develop connection between online data catalogue, data portal and data storage tool (e.g.iRODS) and data archive to facilitate authorized users to access, upload, download of data .. Also they can stage data to the HPC environment for batch data analysis. This includes the development of a plugin that enables preview of data in iRODS; enable uploading of data into iRODS; enable staging of data from iRODS to HPC and back. This deliverable will be developed in close collaboration with WP4.

WP3.11 Some data collections are continuous data streams instead of data batches, or a combination of both. Explore and design infrastructural and tooling requirements to request and analyse (real-time) streaming data in combination with batch data. Explore required tools and adjustments to existing tools. Develop user interfaces to access these data from within the study portal and HPC environment to 'listen' to these streams and define API's (in collaboration with WP2 and WP4).

HPC environment (large scale analysis in trusted environment)

WP3.12 Develop, in collaboration with WP4 (deliverable WP4.7), a federated identity service for the login into the HPC environment (using SSH) so that users can login using their institute accounts (UG or UMCG). The federation means that instead of using a local user database for the HPC we connect to the same central database used for the other part of Human Subject Research. We will build on existing work from NFU D4LS, SURF and BBMRI. The required technological infrastructure will be delivered in WP4. This deliverable ensures connection with HPC.

WP3.13 Develop and implement a user interface (self service portal) for application admins (e.g data managers) to easily manage access to HPC and for allocation/monitoring of storage quota. Also, application admin can easily allocate TBs to certain users and monitor their usage. This asks for an interface on top of IDvault that is developed in WP4. The current user interface
**HSR Programme Proposal**

will be extended to link to the quota system.

**WP3.14** Develop and implement, in collaboration with WP4, a user interface (self service portal) for allocation of cpu/mem resource quota (for researchers) so users can get certainties about the capacity available. It delivers the functional input for deliverable WP4.7.

**WP3.15** Explore and test the scale out of the Human Subject Research infrastructure to the EGI cloud building on the infrastructure from WP4. Users with large analysis needs can scale out to the EGI cloud facility of UG currently being developed. This will build on expertise of current UMCG cluster and Elixer project and needs configuration of a virtual cluster conform Human Subject Research norms. The EGI cloud is part of the UG ICT strategy plan.

**Training and support**

**WP3.16** Deliver a central software deployment of user specific software (self-service) for the Human Subject Research infrastructure that is automated via EasyBuild (in collaboration with WP2 and WP4). It enables users to add their own easy build configurations, without the need for a central sysadmin to do this work. We have a method for users to upload their binaries.

**WP3.17** Implement existing data processing and data analysis pipelines for the pilot studies and help the pilot studies and new users to deploy their pipelines/analyses on the cluster. The support is integrated in the service portfolio offered to the researcher in WP1.

**WP3.18** Integrate authorisation and identification with IAM solution delivered in WP4, to ensure that data request and data access tools can connect to IT infrastructure of UG, UMCG and SURFCONNEXT and users can login with their institute accounts.

**WP3.19** Develop together with WP1 manuals and course materials for users, data managers and for connection developers to use and maintain pipelines, use of HPC system, use of the study portal and catalogue system. The training or e-learning courses are organised and offered by WP5.

**Team**

<table>
<thead>
<tr>
<th>Function title</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>WP leader</td>
<td>Coordinates overall WP, reports to programme manager, aligns tasks and resources, monitors planning.</td>
</tr>
<tr>
<td>Software/database developer</td>
<td>Make warehouse ready for use cases, configure catalogue for use cases, implement configurable data request workflow module and make research study portal multi-tenant/self-service to reduce management cost.</td>
</tr>
<tr>
<td>IT/ technical architect</td>
<td>Model and design the overall technical architecture. Connect existing building blocks. Develop and ensure connections with WP1, WP2 and WP4, HPC and external data infrastructures in close collaboration with information architect.</td>
</tr>
<tr>
<td>Expert on streaming data</td>
<td>Advise and develop technical infrastructure that can efficiently handle and combine streaming and batch data</td>
</tr>
<tr>
<td>Data scientist</td>
<td>Install packages on cluster for research, develop pipelines, user interfaces to manage pipelines and monitor quotas and use.</td>
</tr>
<tr>
<td>Security expert</td>
<td>Uplift the security level of the data request and catalogue tools to the same level as research workspace to enable use without workspace and perform audits on tools (in kind contribution by CIT/UMCG).</td>
</tr>
</tbody>
</table>
**HSR Programme Proposal**

**Partners and Stakeholders**

- Clinical Research Office UMCG
- Research Data Office
- University Library (Dienstenteam, Team O&O, Team Innovation and Development)
- Central Medical Library, UMCG
- Pilot projects
- Security Manager UG
- Security Manager UMCG
- Center for Information Technology (CIT, UG)
- Collaborations with NFU D4LS, BBMRI.NL, DTL/ELIXIR-NL
16. WP4: Virtual research environment and technical infrastructure

Context and aim

All tools and solutions developed in the WP1, WP2 and WP3 of the programme need a technological backbone. When a research dataset is made ready for release (in WP2) and the specifics of the dataset specified via the data catalogue are implemented (WP3), the research data are made available in an analysis environment (for instance the research workspace) that fits with the characteristics of the data.

In the analysis environment, a data selection and analysing tools are available for the researcher. The analysis environment connects to a HPC, data analytics (BI, AI) and visualisations tools. At all times the exact dataset the researcher was given access to can be restored. A data manager, PI and/or the researcher has insight in the status of the data access and use of the data. After the researcher has completed the research can be archived in a long-term storage facility in connection with the identifier of the original dataset, a data catalogue to enable research verification and reuse of the data.

The starting point of WP4 is the capacity and functionality of the current technological infrastructure and research workspaces. Part of the functionality and the technological infrastructure will be developed in collaboration with other UG ICT strategy projects, this is indicated in the table.

Connection with other WP's

The tools, software and pipelines of WP2 and WP3 run on a technical backbone delivered in WP4. WP4 creates the analysis environment to analyse the data created collected in WP2 and made available for research via WP3. WP1 sets the specifics of the analysis environment.

Deliverables

WP4 creates the virtual research environment and technical backbone of the infrastructure to perform the data intake of WP1, the data handling of WP2 and WP3 and enables the data analysis phase.

<table>
<thead>
<tr>
<th>Nr</th>
<th>Deliverable</th>
</tr>
</thead>
<tbody>
<tr>
<td>WP4.1</td>
<td>Design, test and implement solution for long term archiving of research and research data after and during the research is being conducted. Storage and archiving solutions for raw data, newly created data, scripts, etc (research archiving). Connection with publication databases (e.g. Pure) and study catalogue to ensure that the data are findable and traceable. The Data Access Policy, authentication and authorization policies are developed in WP1. Input for this deliverable is the running data handling and storage project: CIT project nr 418 FReDS (FAiR Research Data Services)</td>
</tr>
</tbody>
</table>

The required technical infrastructure and large scale deployment for researchers will be funded by Recommendation 3 from the UG ICT strategy fund: “Realization of a data management and archiving service for research data”. For this a separate proposal will be submitted to the ICT strategic commission. The programme will work closely together with the initiator(s) and project leader of this Recommendation, deliver functional input and make all the expertise and pilot work available for the Recommendation 3 project.

| WP4.2 | Develop new version of Research Workspaces for broader use. Increase available tooling and applications based on request of researchers. Allow for different types of workspaces with different levels of security. Create connection with HPC and data analysis pipelines. |
The required technical infrastructure and large scale deployment for researchers will be funded by Recommendation 8 from the UG ICT strategy fund: VRE Research Workspaces and Research Cloud. For this a separate proposal will be submitted to the ICT strategic commission. The programme will work closely together with the initiator(s) and project leader of this Recommendation, deliver functional input and make all the expertise and pilot work available for the Recommendation 8 project.

**WP4.3** Develop a web interface that facilitates the connection to required functionality in the data archiving environment and make self service by a researcher or data manager possible.

The required technical infrastructure and large scale deployment for researchers will be funded by Recommendation 3 from the UG ICT strategy fund: “Realization of a data management and archiving service for research data”. For this a separate proposal will be submitted to the ICT strategic commission. The programme will work closely together with the initiator(s) and project leader of this Recommendation, deliver functional input and make all the expertise and pilot work available for the Recommendation 3 project.

**WP4.4** Prepare the infrastructure (storage, workspaces) for future on-demand use, create a request buffer on a storage and workspace level.

The required technical infrastructure and large scale deployment for researchers will be funded by Recommendation 8 from the UG ICT strategy fund: “VRE Research Workspaces and Research Cloud”. For this a separate proposal will be submitted to the ICT strategic commission. The programme will work closely together with the initiator(s) and project leader of this Recommendation, deliver functional input and make all the expertise and pilot work available for the Recommendation 8 project.

**WP4.5** Explore the needs and possibilities to enable inter-institutional dataflows (e.g. between researcher of UG and UMCG). This includes both the technical measures and policies solutions. Explore possibilities to connect different data infrastructures and align privacy and security policies of external partners (e.g. federative cloud solution) to facilitate collaboration between researchers of different faculties and institutes. The security manager of UG and UMCG and the Managers of the relevant CIT departments will be actively involved in this deliverable.

The required security expertise will be contributed by CIT and the participating institutes. In collaboration with CIT, a separate proposal will be submitted to the ICT strategic commission concerning security (ICT strategy fund, Recommendation 32 to 37).

**WP4.6** Design and develop authentication and authorization solutions for data access from UMCG, UG, SURF/Research Community/external via a self service portal for access to HPC, Workspace Environment, data archive, etc.

The required technical infrastructure and large scale deployment for researchers will be funded by Recommendation 31 from the UG ICT strategy fund: “Identity and Access Management”. For this a separate proposal will be submitted to the ICT strategic commission. The programme will work closely together with the initiator(s) and project leader of this Recommendation, deliver functional input and make all the expertise and pilot work available for the Recommendation 31 project.
HSR Programme Proposal

WP4.7 Design and implement a robust and secure data IT infrastructure for data handling, data transport and data storage based on the functional requirements and recommendations from the risk analyses and PIA from WP1. Input for this deliverable is the running data handling and storage project: CIT project nr 418 FReDS (FAIR Research Data Services)

The required technical infrastructure and large scale deployment for researchers will be funded by recommendation 2 and recommendation 29 from the UG ICT strategy fund: Infrastructure for data handling.

WP4.8 Deliver technical production environment, technical application management and technical support for tools that are offered via the Human Subject Research service. For instance installation, technical application management, migration. WP2, WP3 supply detailed product and use description.

WP4.9 Deliver server capacity (common application server) for data capture, data handling and data integration tools in WP2 and WP3. SPECs are supplied by WP2 and WP3.

Team

<table>
<thead>
<tr>
<th>Function title</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>WP leader</td>
<td>Coordinates overall WP, reports to programme manager, aligns tasks and resources, monitors planning.</td>
</tr>
<tr>
<td>Software/database developer</td>
<td>Develop the infrastructure and tools for IAM (identification and authentication management), servers, network and research workspaces. Connect with tools from UMCG and UB (Molgenis, data capture tools, Pure, etc). Develop connection to HPC.</td>
</tr>
<tr>
<td>IT (technical) architect</td>
<td>Model and design the overall technical architecture. connect existing building blocks. Develop and ensure connections with WP1, WP3 and WP4 in close collaboration with information architect.</td>
</tr>
<tr>
<td>Security expert</td>
<td>Review and align security policies of internal and external parties, (UMCG-UG, UMCG-LUMC, etc) (in kind contribution by CIT/UMCG).</td>
</tr>
<tr>
<td>Expert in streaming data</td>
<td>Advise and develop technical infrastructure that can efficiently handle and combine streaming and batch data</td>
</tr>
<tr>
<td>Technical application and database manager</td>
<td>Technical management of data archive (taping), storage, servers, network, workspaces and HPC management.</td>
</tr>
</tbody>
</table>

Partners and Stakeholders

- Security Manager UG
- Security Manager UMCG
- Clinical Research Office UMCG
- University Library (Dienstenteam, Team O&O, Team Innovation and Development)
- Central Medical Library, UMCG
- Research Data Office
- EPD information specialist
- CIT project nr 418 FReDS (FAIR Research Data Services)
HSR Programme Proposal

17. WP5: Sustainable service and outreach

Context and aim

The goal of WP5 is to sustainably set-up the service portal for researchers. This means sustainable in the sense of streamlined governance, finances, organisation structure, equipped personal, PR an internal and external network. This WP will also assure proper documentation of the expertise and tools, create a portfolio and keep it updated. The WP will create visibility and awareness and make the service easily accessible and create connections between professionals working in data management to facilitate knowledge exchange and training. WP5 is supervised by the Programme Manager and supports the overall coordination of the programme.

Connection with other WP’s

The input of the product portfolio for the comes from WP1, 2, 3 and 4. WP5 makes them findable and visible and creates a sustainable situation.

Deliverables

This WP delivers a sustainable and governed service and infrastructure that is well-embedded in the UG and UMCG, easy accessible for researchers in all faculties. Also the service and infrastructure has good visibility on a national level and has created new research opportunities.

<table>
<thead>
<tr>
<th>Nr</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>WP5.1</td>
<td>An up-to-date product portfolio that matches the needs of researchers is made available through a catalogue (includes amongst others templates, SOPs, policies, best practices and background documentation) and is updated regularly. WP5 collects the information and products that are available at the different facilities and services in UG/UMCG and makes them visible and findable. This input comes amongst others from with WP1, 2, 3, 4.</td>
</tr>
<tr>
<td>WP5.2</td>
<td>Communication and PR materials (logo and create concept and template ppt, etc) and communication channels (website, contact address, social media) to create visibility and support information and service function. The required communication staff will be contributed by CIT and the participating institutes. The programme contains budget for communication materials.</td>
</tr>
<tr>
<td>WP5.3</td>
<td>Develop PR and marketing plan to increase (inter)national visibility. Monitor of researcher satisfaction and service. Identify valorisation options for the service and IT infrastructure and explore new possibilities for collaborations between researchers and private parties facilitated by the infrastructure. Consolidate best practises and knowledge.</td>
</tr>
<tr>
<td>WP5.4</td>
<td>A sustainable and robust governance and business model and function/staffing structure for the service and IT infrastructure, including a plan for proper and effective technical and functional management, to be able to maintain the service and IT infrastructure and create a financial sustainable service. This includes an estimation of future required capacity (e.g. storage capacity, back-up system) and future required functionalities (e.g tools for pop-up data, data harvesting). Collaborate with HR level on newly created functions and roles.</td>
</tr>
</tbody>
</table>
**HSR Programme Proposal**

**WP5.5** Training for programme members, data consultants and data support staff in privacy, security and scientific data management (in collaboration with RDO, ABJZ and existing trainings like, Hanze module on privacy, IT academy, etc). This will build on expertise and training material that is already available at RDO and ABJZ.

**WP5.6** Active network of UG/UMCG employees with information and data management tasks in the different faculties and facilities (e.g. demand managers, research support staff, funding officers) is formed. The network is regularly updated and informed and is connected via amongst others biannual information meeting, communication channels. Good connection with the wide range of research disciplines and research practices via these staff members is established.

**WP5.7** Identify IT research funds and submit at least 2 research proposal together with researcher.

**Team**

<table>
<thead>
<tr>
<th>Function title</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>Programme Manager</td>
<td>Overall supervision and coordination of the programme</td>
</tr>
<tr>
<td>Business developer</td>
<td>Develop PR and valorisation plan, identify funding and valorisation possibilities, support grant writing, represent service and programme.</td>
</tr>
<tr>
<td>Communication officer</td>
<td>Develop and implement communication strategy. Develop website, PR materials, set-up and monitor social media accounts; in kind contribution by CIT</td>
</tr>
</tbody>
</table>

**Resources**

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communication and PR</td>
<td>Development of creative concept, logo, PR materials, website. Website/email/social media maintenance + hosting. A CRM/contact management system.</td>
</tr>
<tr>
<td>Training</td>
<td>Training for functions and members of programme members. Information meeting for users of the Human Subject Research programme</td>
</tr>
<tr>
<td>Management and coordination</td>
<td>Costs for coordination of the programme, like meetings and travel costs. Document management system for documentation of the procedures and SOPs.</td>
</tr>
<tr>
<td>Consultancy, audits and security tests</td>
<td>Consultancy on for instance security, marketing, valorisation, IP issues</td>
</tr>
</tbody>
</table>

**Partners and Stakeholders**

- SBGG / Legal department UG / UMCG (to assist with IP and governance)
- Clinical Research Office UMCG
- Research Data Office
- University Library (Dienstenteam, Team O&O, Team Innovation and Development)
- Central Medical Library, UMCG
- BiKE, UMCG
- Research & Valorisation UG
- Communication offices UG and UMCG