RESEARCH SUPPORT IN THE NETHERLANDS

CURRENT SITUATION AT UG AND UMCG

SURF

Many institutions are involved in research support: they want to support their researchers in their access to high-quality ICT infrastructure in order to process, analyse, visualise, manage and store research data. Each institution does this in its own way. SURF presents a series on how institutions organise their research support and their agenda for the future. This series is based on SURF's Support4research innovation project. This project aims to ensure that the services provided by SURF to researchers better meet the needs of the institutions and researchers. This episode focuses on the University of Groningen (UG) and the University Medical Center Groningen (UMCG). **If you want to share your approach to research support, please let us know by sending an email to support4research@surf.nl.**

SUPPORT FOR RESEARCH AT UG AND UMCG

UG / UMCG

Number of UG employees
5.900 FTEs (including UMCG Education & Research)

Number of UMCG employees
9.064 FTEs

Scientific staff
3.000 FTEs

Number of professors
400 (female: 100)

Number of doctoral candidates 2.000

Number of students
30.000

Scientific publications per year 6.000

Key research areas
- Energy
- Healthy Ageing
- Sustainable Society

Introduction

The University of Groningen (UG) has a rich academic tradition dating back to 1614. This research university sets itself apart internationally due to the strong link between its education and research, focused on three key areas: Energy, Healthy Ageing and Sustainable Society. Not only is the education and research of direct social relevance, the researchers also perform large-scale work with partners from the business community, social organisations and the government.

The university has three focal points in the area of digitalisation: Big Data, Smart Industry and Facets of Cultures – Digital Humanities. The Center for Information Technology (CIT) provides all IT facilities for the university, including for the management and storage of research data. The CIT namely has expertise in relation to very large-scale data and sensitive data (such as personal data and company-sensitive data).

The University Medical Center Groningen (UMCG) is the second largest hospital in the Netherlands and the largest employer in North Netherlands. More than 10,000 employees work daily towards the common goal: building the future of health. UMCG's core tasks are healthcare, education and research. In the area of research, UMCG works closely with UG, and in particular the university's Center for Information Technology (CIT). The research addresses new techniques and treatments, new medicines and new forms of care, focusing on helping people lead healthier, more active and longer lives, in other words Healthy Ageing.

Vision for research support

Up-to-date data management, training, IT support and infrastructure form the basis for excellence in research, education, social impact and talent development. UG and UMCG occupy a leading position in the area of High Performance Computing (HPC), visualisation, geo-analytics and the virtual (research) workspace. Besides providing facilities, significant expertise has been acquired in areas that include data science, data management, privacy and security. Centralised and decentralised expertise and facilities co-exist and collaborate to form a live and interactive network that comes together in the Data Federation Hub.

The underlying idea is to support researchers during the entire research data life cycle. Researchers need help in finding the correct datasets, processing this data, large-scale storage capacity and computing power. Legislative amendments increase the need for support.

UG and UMCG attach importance to a transparent research environment. We encourage this by applying the principles of Open Science, as set out in the UG Strategic Plan. UG and UMCG also encourage researchers to apply the FAIR-principes (Findable, Accessible, Interoperable and Reusable) to dealing with and managing research data.

Professional support entails brainstorming with the researchers, translating complex questions into concrete solutions and responding to the latest developments. This is why we invest in innovation and a user-oriented service here at IT support. UG and UMCG participate, for example, with different private partners in the testbed projects of the North Netherlands Cooperation Agency (Samenwerkingsverband Noord Nederland), where new and innovative techniques, such as the *dHealth*, *5G* and *Mining Big Data* testbeds are being developed.

Cooperation is vital for the optimal support of researchers. Short lines of communication are needed to link the different expertise units and facilities. It is important that the support is findable and user-oriented, and that the researcher is referred to the correct counter for assistance. The Data Federation Hub plays a major role in linking these different initiatives and promoting cooperation.

Examples of current and future international cooperation include the large astronomical projects such as MUSE, MICADO, LOFAR, Euclid and KiDS. As the Euclid Netherlands Science Data Centre is housed in the CIT, the CIT plays a major role in developing the *Euclid Data Processing System* and the *Euclid Distributed Storage System*. The technology that is developed for Euclid will also support the SNN testbed project Mining Big Data. The data volumes of all these projects are enormous: 1 petabyte of data is already being stored for MUSE and KiDS at several locations, while Euclid will produce more than 50 petabytes of data.

Organisation of research support

Different units of UG and UMCG collaborate closely in Groningen. As the spider at the centre of the web, the Data Federation Hub plays a major role in this regard. Support for researchers entails that wherever they turn for help, they are always assisted further (no wrong door policy) and that all front-line counters have the same information or know where the required information can be found. The available IT services form part of the large-scale national scientific infrastructure of the Netherlands Organisation for Scientific Research (NWO) and can be found under the name CIT data warehouse. UG and UMCG also work with different national and international partners.

Units that provide research support:

- **Center for Information Technology (CIT)** The CIT is an ICT expertise centre that analyses large data sets and provides support in research data management and processing. The CIT has two focus areas: Big Data and sensitive data.
- **Research and Innovation Support (RIS)** RIS provides services in the area of HPC, visualisation (3D/AR/VR), the Geodienst, data science and research data management/repositories.
- **Research Data Office (RDO)** The RDO plays a role in front-line support and research data management plans, as well as supporting researchers in selecting the correct IT solution.
- **UMCG – Service Desk Clinical Research Office (CRO)** The CRO is a designated service portal for clinical researchers. The CRO has strong links to the RDO.
- **UMCG – Genomic Coordination Center (GCC)** The GCC provides support for the entire research data cycle and focuses on bioinformatics, genomics and biobanking.
- **UMCG – Research Data Support (RDS)** Research Data Support supports researchers in designing and conducting high-quality human research.

Current infrastructure and associated services for researchers

CIT provides a large part of the infrastructure for UG researchers. Data takes centre stage in this regard: big/complex/privacy-sensitive data is the essence of much scientific research. In addition to secure and stable storage facilities, the CIT provides additional facilities for researchers. The facilities are housed in four computing centres, with redundant network links for both internal and external connectivity. Researchers also use external data services, including DataverseNL, DANS-Easy, BBMRI-NL and CTMM/TRaIT.

	COMPUTING SERVICES
Number of clusters within the institution	± 14 clusters
Total scope of computing power	5640 CPU cores (Peregrine HPC cluster PG), 200 nodes (PG), 30 TB RAM (PG), 600TB storage
Third-party clusters within the institution	Genomic Coordination Center (GCC) clusters, Molecular Dynamics Group, Solid Matters Cluster, DAWN GPU cluster, LOFAR (Cobalt correlator/CEP4/CEP3) ASTRON (Aartfaac/Dragnet) OpenStack, VMware-rugcloud
Acquisition of computing services	SURFsara
	STORAGE
Central bulk storage capacity	• 3 PB Lustre; Data handling • 0.6 PB Lustre; HPC cluster • iRODS storage services • 0.5 PB incl. dedupe data user storage Windows workspaces • 4 PB UG-Cloud storage
Central archive capacity	• 2.5 PB Tape backup double copy (2018) • 1 => PB disk backup with dedupe double copy (2018) • 3.7 = > PB Tape Archive double copy (2018) • 2 => PB Research archive (dcache - iRODS) (2018-2019)
	NETWORK
Routed capacity (external)	2 x 10Gbit/s Campus IP 2 x 10Gbit/s HPC IP
MSP capacity	2 x 100Gbit/s MSP 1 x 10Gbit/s MSP
	AUTHENTICATION & AUTHORISATION INFRASTRUCTURE
Connected to SURFconext Federation	Yes
Other AAI suppliers	No
	DATA CENTRES
Landleven Data Centre (UG)	Tier 2 power 450 kW 1 x 100G SURFnet network link 2 x 100G Smitsborg interconnect
Smitsborg Data Centre (UG)	Tier 2 power 300 kW 1 x 100G SURFnet network link 2 x 100G Landleven interconnect
Eemspoort UMCG Data Centre (UG part)	Tier 3 power 150 kW 2 x 40G network link
DUO Data Centre (UG - HPC)	Tier 1 power 180 kW 2 x 100G network link
New-build data centre (UG)	Tier 3 power 720 kW planning 2018-2019
	EXPERTISE
	Data consultants, data scientists, scientific programmers, visualisation, Geodienst HPC experts, IT lawyers

Virtual Research Workspace

The Research Workspace is a virtual Windows ICT environment that the user can partly set up themselves. This environment provides access to all research ICT services of UG and UMCG. These services are available 24/7 and are equipped, where necessary, with extra memory, extra powerful processors or graphic-intensive support. The Research Workspace is very suitable if the standard university workspace is not adequate and if the Windows applications are not suitable for HPC use. This environment is also suitable in case of stringent dataset security requirements or if additional privacy measures are needed.
The Research Workspace portfolio consists of the following workspaces that a researcher can use:

1. **Shared**: all users share all components (memories/graphics card/processor).
2. **Dedicated Start**: components are assigned to a user, i.e. there is no need to share memory, etc.
3. **Dedicated Middle**: the same as Dedicated Start, but with extra memory or computing power.
4. **Dedicated Large**: to be customised by the client/researcher.

All workspaces are delivered with a standard basic software package and standard data storage. The university is responsible for management (OS and applications). The provision of all other services is by agreement.

High Performance Computing

Researchers whose work includes large numbers or complex calculations can approach the university's Peregrine Linux computing cluster for assistance. Hundreds of researchers from a broad range of disciplines – from linguistics to molecular dynamics, and from robotics to genomics – use the computing cluster. The CIT has a second national HPC centre in the Netherlands (besides SURFsara). The Genomics Coordination Center (GCC) also provides HPC facilities for human research within UMCG. The CIT further accommodates and manages three clusters for the LOFAR radio telescope.

Reality Center

The CIT Reality Center has very advanced virtual reality (VR) facilities and provides scientists with hypermodern facilities in the area of 3D visualisation. By providing this as a central service, the Reality Center has been able to acquire optimal expertise, making it possible to reuse knowledge and software. Researchers can contact the Reality Center for data visualisation, simulations, 3D design, courses and workshops. The Reality Center makes it possible to obtain new information from complex data, explain complex concepts, avoid construction faults, save time in the design process and advance decision-making processes.

Network infrastructure

The university has a state-of-the-art network that is managed by the CIT. The CIT works continually on innovating and improving the network. The CIT also addresses improved network security and increasing network capacity and virtualisation (software-defined networking). Measures such as network zoning also ensure optimal security.

Big data handling

Significant expertise has been acquired over the last few years in data handling, data sharing and data reuse, particularly in the areas of astronomy, energy and the life sciences. Examples of such large research infrastructure and projects include LifeLines, Energysense, GLIMPS, LOFAR and Euclid. This Big Data expertise is documented in the existing research data support organisations, such as RIS and GCC, so that it can continue to be provided to researchers.

Data science

Proper and smart data analysis is required in order to use the increasing quantity of data for research. A data scientist can help to retrieve information from bulk data. Data scientists work with experts in multidisciplinary teams such as the Center for Data Science and Systems Complexity (DSSC) and the Genomic Coordination Center (GCC). Various courses are also on offer, ranging from data warehousing, programming in Python and R, and using HPC and visualisation facilities, to machine and deep learning. The CBS Academic Data Center is also based in the CIT. The Data Federation Hub provides support in bringing together different forms of expertise and reusing best practices in the area of data science.

Geodienst

The Geodienst is UG's spatial expertise centre that provides research support and innovative services in the area of geodata, geo-analysis and GIS software.
The Geodienst maintains an open data portal and provides customised courses.

Toolbox UMCG and RDMP (Research Data Management Plan) webtool

Toolbox clinical research includes a broad range of Standard Operating Procedures (SOPs), templates and best practices in support of human research. Researchers can use the toolbox as an aid to translate codes of conduct, guidelines and legislation into research practice. Researchers can also specify which facilities (internal and external) they use in their research projects. The RDMP webtool is moreover available for research data management. The RDMP webtool is set up specifically for each research institute and helps researchers to use the most suitable facilities for their research.

Training and education

Customised training and workshops are available for the use of applications and IT facilities. These include data analysis, HPC, data science, geo-software and tools and 3D visualisation. Both UMCG and UG also provide different training courses and workshops on research data management and privacy aspects of research data. A website with a summary of all available training is under development.

Agenda for the immediate future

It is important for the university that data is easily accessible, reliable and secure from several locations on a 24/7 basis. Secure collaboration on the basis of federated identity management (FIM) is highly desirable. The data must be accessible from the workspace, research workspace, computing cluster and visualisation systems. A national federated infrastructure with adequate broadband is needed for this purpose. Low-threshold access to a range of expertise is also important.

Data Federation Hub

In order to coordinate and make the existing facilities more visible, we are working on intensifying collaboration and pooling knowledge. The Data Federation Hub is a shared resource in the area of data management and data science for all research domains. The primary focus areas are training, FAIR data catalogues, knowledge exchange and data science.

The researcher centrally

The researcher is increasingly taking centre stage. To this end, it is firstly important that the researcher knows where to find support. We are also working towards a central point of contact. The Data Federation Hub further serves as a gateway to all available data expertise.

Virtual Research Environment (VRE)

Over the coming years, we will be creating a Virtual Research Environment (VRE). Support throughout the entire data life cycle is at the forefront of this initiative. We are also aiming to connect with other VREs in the Netherlands. A properly functioning VRE will also enable us to grant external parties access to the different internal systems.

Human Subject Research (HSR) Programme

In January 2017, UG and UMCG started the university-wide H*uman Subject Research (HSR) Programme* for data from human research. The purpose of the HSR is to provide up-to-date infrastructure and resources for storing and processing privacy-sensitive research data. Infrastructure that stimulates innovative research in line with current policy in the areas of research data management, Open Science and FAIR data.

National and international positioning

Over the coming period, we will be strengthening the national and international position of our IT services and actively participating in national and international e-infrastructure. We work together with national federated services, in collaboration with partners that include SURF, Netherlands eScience Center and the NWO. In the area of health, we also participate in various national and international initiatives, including BBMRI, ELIXIR, Health-RI and DTL. And we continue to collaborate internationally in the area of astronomy. Lastly, we actively participate from Groningen in the European Open Science Cloud initiative.

For contact with the Data Federation Hub, send an e-mail to: dfh@rug.nl

COLOPHON

Author
Wietske Degen, Rijksuniversiteit Groningen
Salome Scholtens, Universitair Medisch Centrum Groningen

Project management
Jan Bot, SURFsara

Editors
Erik van der Spek, Hendrikx van der Spek
Jocelyn Manderveld, SURFnet
Judith Pijnacker, SURFnet

Editor-in-chief
Judith Pijnacker, SURFnet

Translation
Xplanation

Design
Crasborn Communicatie Vormgevers
www.crasborn.nl

Date
January 2018

More information

SURF
Moreelsepark 48
3511 EP Utrecht
The Netherlands

Postbus 19035
3501 DA Utrecht
The Netherlands

+31 20 88 787 30 00
info@surf.nl
support4research@surf.nl
www.surf.nl/s4r

SURF is the collaborative ICT organisation for Dutch education and research. SURF offers students, lecturers and scientists in the Netherlands access to the best possible internet and ICT facilities.