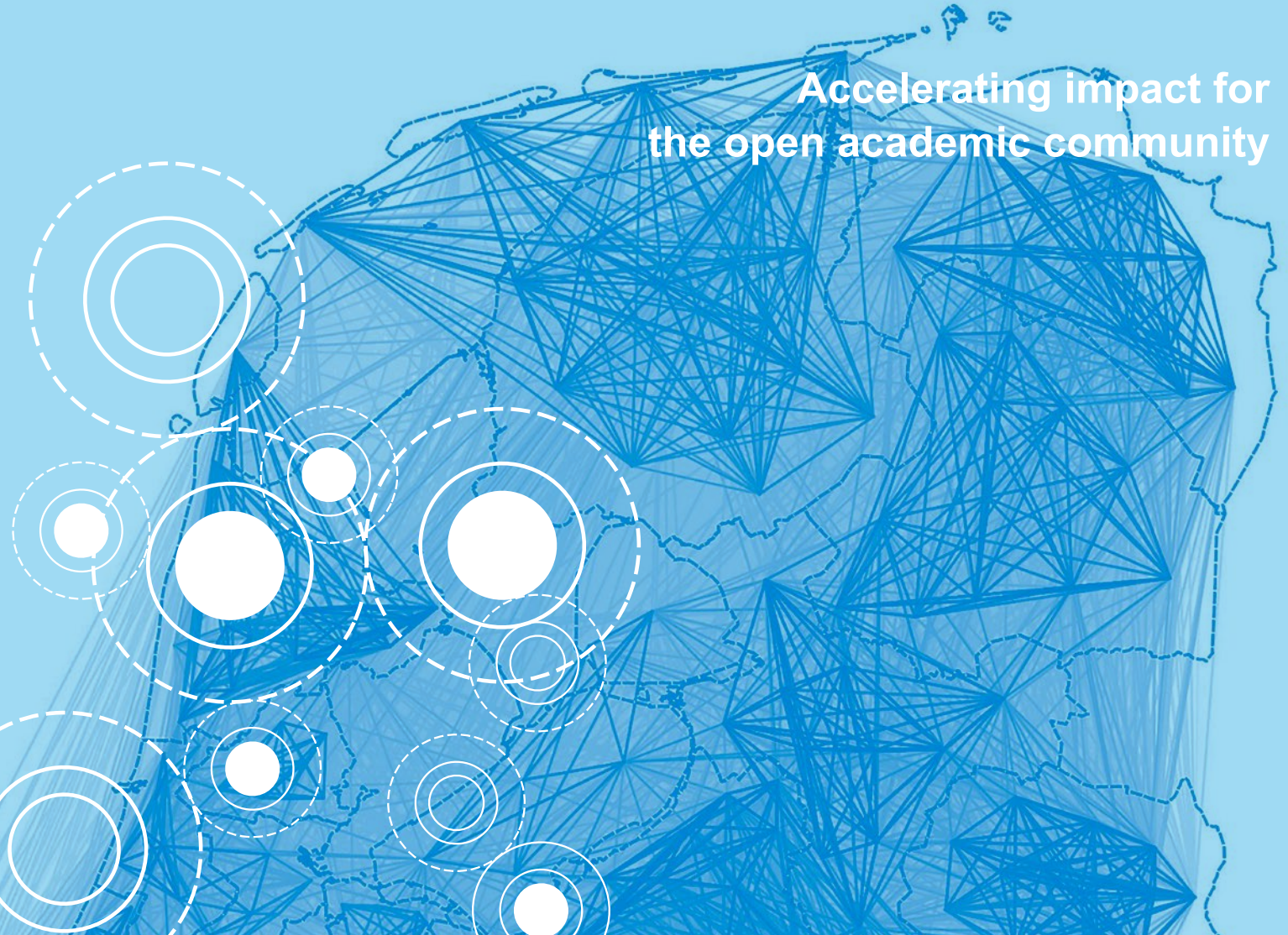university of groningen

center for
information technology

CIT RESEARCH ACTIVITIES REPORT 2021

Accelerating impact for the open academic community

Contents

PREFACE

Working together towards innovative teaching and research with societal impact

These days, one cannot imagine research without innovative digital research methods and data-driven research. The Digital Competence Centre (DCC) is a close partnership between the University of Groningen (UG) University Library (UB) and the Center for Information Technology's (CIT) Research Domain. It is our ambition to put our expertise to optimal use in the support that we provide, whether delivering storage infrastructure, processing, sharing or the re-use of data, offering High Performance Computing (HPC) facilities and support, or creating Extended Reality (XR) content and data science support.

What you may not know is that with our expertise, we can also support prospective research projects at an early stage by helping to write grant applications. In the final phase of a project we can contribute to publications that have been realized with our experts in the field of data science, XR, HPC and FAIR research data management.

In addition, we also want to deploy our unique expertise in the field of computing, XR, and data science to create innovative teaching and training offerings.

This report contains an overview of our activities over the past year. We are proud of our contribution to so many interesting teaching and research projects, and we hope that this report will inspire you to contact us.

*Marijke Verheij*
Research Domain Manager

REVIEW OF 2021 ACTIVITIES

General

Digital Competence Centre

In March 2021, the UB and the CIT jointly launched the UG's Digital Competence Centre (DCC), a central research support desk for research data and IT. The DCC provides accessible support in the field of FAIR (findable, accessible, interoperable, and reusable) research data management (RDMS), privacy, data protection, and associated IT solutions. It is also committed to developing RDM learning pathways for various target groups, offering clear generic information, and organizing a University-wide research support community. To this end, as a centre, the DCC is committed to connecting and professionalizing both centralized and decentralized (faculty level) data stewardship.

The new DCC has had a busy year. It has processed approximately 500 support requests, covering all UG faculties, with an average customer satisfaction rating of 8.6. DCC data stewards have spoken to nearly 200 PhD students during introduction days, organized 50 training sessions, network interviews, and events (for example, in collaboration with the Open Science Community, the UMCG DCC, and the Young Academy), supported nearly 80 data publications in Dataverse, and helped the UB Pure Team to make 2,550 datasets findable and accessible.
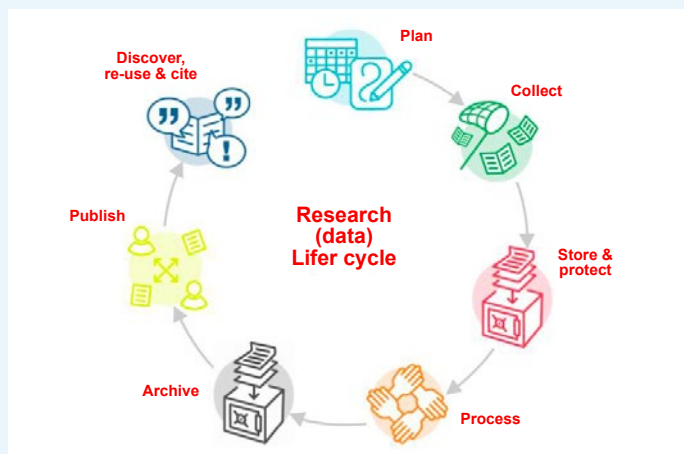
Virtual Research Environment

In recent years, the number of technological possibilities as well as the requirements for conducting research have risen substantially. In addition, the policies in the areas of privacy and integrity have been tightened considerably. Partly due to these developments, the research landscape has become increasingly more complex for both the organization and the researcher. In order to safeguard controlled access to IT/data facilities and the safe sharing of data, the Virtual Research Environment (VRE) will be extended and, if all goes to plan, made available to all researchers in 2022.

In this online environment, researchers will be able to collaborate with each other. The layout of a VRE can differ per research and is determined based on the needs and requirements of the research. There are versions of the VRE in which researchers are able to collaborate with wikis or forums, but there are also versions in which it is possible to collaborate on complex simulations that require a lot of computing power, for example. The VRE is developed step by step and extended with extra functionalities. The Virtual Research Workspace is the VRE's first component. More components will follow later, such as the Research Data Management System.

Virtual Research Workspace

The Virtual Research Workspace (VRW) is a virtual desktop environment where researchers can safely collaborate and perform data analyses. Within the VRW, collaboration is possible with researchers from the UG and other SURF institutions. The VRW is an environment in which researchers

have full control over their data and results, such that the risk of a data leak is minimized. The VRW has been available for a few years now, and it is still evolving. The most recent development is the construction of a self-service portal that includes the VRW as a standard offering.

In early 2022, a number of pilots were conducted with local devices to see whether sensitive video and other data could be transferred directly to the VRW from a local device. The VRW environment now also includes a central storage space (repository) for programming languages R and Python. An increasing number of research groups are applying for a VRW, both within the UG and the UMCG and beyond. Based on a pilot with Human Movement Sciences students, we are currently exploring options for allowing students to make use of the VRW for their research projects.


More information about the VRW

Research Data Management System

Since 2018, the UG and the UMCG have worked together to develop a Research Data Management System (RDMS). The CIT has built this RDMS, an environment in which researchers can safely store, process, publish, and archive their data, as well as automatically enrich it with metadata. These data can also be linked to various sources, such as HPC, allowing researchers to perform complex calculations. What's more, the option of sharing data with other researchers within and beyond the University promotes collaboration and the efficient use of research data.

During the pilot phase, the CIT multidisciplinary project team worked together with researchers on various customized use cases. Examples include EUCLID, UMCG XNAT/Genetic, Publishing to Pure, and GELIFES. As of writing, more than 600 researchers are using the RDMS. Over the past period, the project team has worked on preparing the ground so that the RDMS project can be delivered by mid-2022 as a self-service with support from the DCC, or as a customized use case.

B3CARE 'Overview'
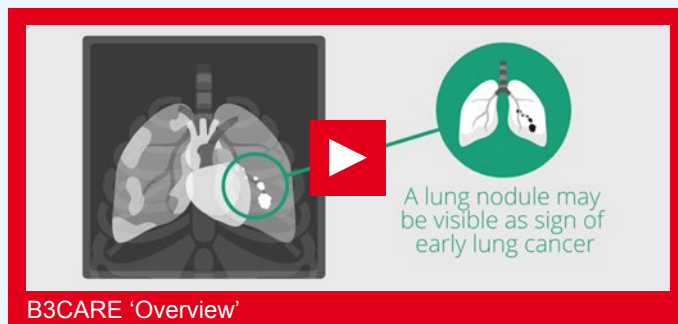
B3CARE

B3CARE is a large-scale project by the UMCG (Radiology, Radiotherapy, and the Data Science Center in Health [DASH]) that focuses on the computed tomography (CT) evaluation of three major diseases: lung cancer, pulmonary emphysema, and cardiovascular diseases, collectively known as the Big 3. This project aims to integrate the CT biomarkers of the Big 3 into a single screening test, to learn to identify and diagnose them at an earlier stage, and to validate and implement automatic artificial intelligence (AI) software in assessing the relevant biomarkers. The B3CARE project is managed by a multidisciplinary team. Within this team, the CIT is responsible for making sure that the infrastructure is suitable for assessing the more than 45,000 scans, and delivering the complete infrastructure required to do so, including the VRW. The infrastructure is now nearly ready to be delivered as a service.

2021 Calls for proposals

In 2021, the CIT issued two calls for proposals. In early 2021, the CIT called on UG and UMCG lecturers and researchers to submit proposals for research and teaching projects that make use of virtual and augmented reality (XR) and 3D data. Three proposals were accepted: Lightning Imaging with LOFAR, Brothers of the Common Life 3D, and Building Bounce Back. In the second half of 2021, the CIT Visualization, Data Science, and Geodienst teams issued a joint call for proposals for innovative UG projects on visualization and geo and data science with the potential of making an important contribution to science. The accepted proposals were announced in early 2022.

🔗 *News reports March 25, 2021*
'2021 Grant for visualization projects'

🔗 *News reports January 20, 2022*
Fourteen proposals accepted in the 2021 Visualization, Geodienst and Data Science Call for proposals

Visualization

Lightning Imaging with LOFAR
Prof. Olaf Scholten (Faculty of Science and Engineering) and Laurens Voerman (Visualization)
Together with Prof. Olaf Scholten, our visualization team developed new 3D visualization techniques for visualizing lightning in 3D with the LOFAR telescope. Until recently, the development of lightning over time could only be explored in 2D. With the help of these new 3D techniques, researchers can now walk inside the lightning flash and study it from different perspectives.
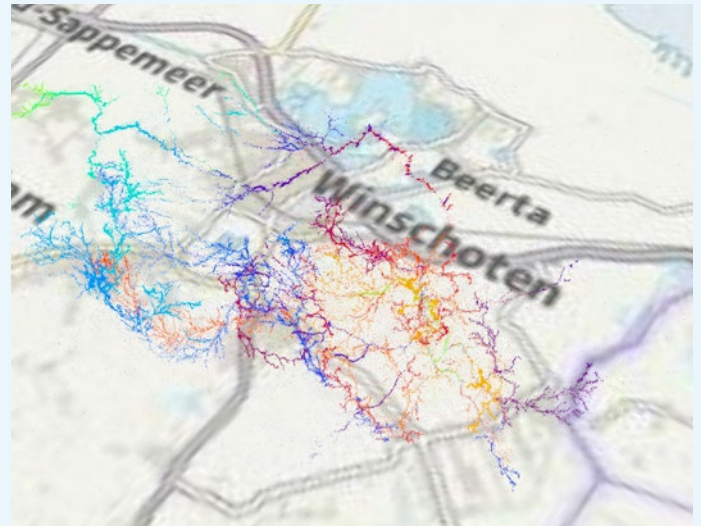
Brothers of the Common Life 3D: Reconstruction of the interior space of the Heer Florenshuis
Dr Margriet Hoogvliet (Faculty of Arts), Bart Campman and Pjotr Svetachov (Visualization)
The future of history in 3D visualizations is now possible thanks to the successful partnership on the Brothers of the Common Life 3D project. Our project team worked on creating an inter-active digital reconstruction of two fifteenth-century interior spaces of the Heer Florenshuis in Deventer, with 3D models of the original books, stationery, and furniture. Digital brothers are ready to answer the visitors' existential questions, based on

their original sermon texts. In this virtual environment, visitors can walk through the house, leaf through books, and talk with one of the brothers.

Building Bounce Back (3B)
Dr Catheleine van Driel (Faculty of Medical Sciences/UMCG), Bart Campman and Pjotr Svetachov (Visualization)

Dr Catheleine van Driel is the founder of the Bounce Back Lab, where research is conducted into various ways to strengthen people's stress resilience. Our visualization team contributed by developing a virtual stress resilience training programme and creating a VR simulation for a virtual job interview with a stress-related scenario. Users can use this simulation to practise stress management techniques and learn to control their stress response.

Knowledge sharing in the XR Hub
The XR Hub brought together lecturers and researchers from all UG and UMCG faculties to share their knowledge and experience of developing and using XR applications. All new options and techniques, as well as recent visualization projects and partnerships, were highlighted during the four XR Hub meetings held this year.

Return of the European bison to the Netherlands

As part of her research project, Lisa Sánchez Aguilar (PhD student at GELIFES) investigates the return of the European bison to the Netherlands. GELIFES researchers use virtual reality to explore new options for organizing nature reserves. Based on 2D illustrations, the visualization team created a 3D visual space in which the European bison can be observed walking all the way from the Białowieża forest in Poland to the Dutch dunes.

Virtual Practicals

For the Virtual Practicals (ViP) project, our team developed a 3D-VR application in which students can safely practice complex laboratory skills in a virtual, yet realistic, environment.

The sky is the limit. In a virtual environment, almost anything is possible. From learning to plead your case in a virtual court of law to exploring an archaeological mound near Rome in 3D.

Niek Eisink (Lecturer in X-Lab Chemistry at the UG) is one of the lecturers supervising students in chemistry and organic chemistry practicals in a VR environment.

'Some students find the combination of working with breakable glass containers and dangerous chemicals frightening. Instead of focusing on their learning objectives, they spend much of the time worrying about something going wrong. Thanks to the VR environment, we can introduce students to the lab and glasswork in a calm and playful way. The greatest advantage is that they can safely make mistakes. The combination of VR simulation and supporting instruction videos ensures that students are well-prepared and feel calmer in the actual practical.'

🔗 *News article*
'Students learn to plead their case virtually with PleitVRij'

🔗 *News article 04 March, 2021*
'The untold story of Crustumerium: 3D exploration of an archaeological mound near Rome'
▷

Compute

Delivery of the HPC Data Center building

A year after construction began in September 2021, the new HPC Data Center building was delivered. In December, it was announced that the energy-efficient building would be named the Coenraad Bron Center. The Board of the University chose this name from the many submissions in the competition for a name for the new data centre. In the coming period, the Coen Bron Center will be further prepared for housing new and existing systems. In the course of 2022, the first systems with high-quality computing power will be made ready for use by UG researchers.

'This new building allows us to meet the University's growing need for high-grade computing power and options for reliable data storage and processing. It allows the CIT to maintain its position as an expert in research data. This kind of exceptionally strong infrastructure is also crucial to be able to join international partnerships on large-scale ground-breaking research projects.'

Anke Breeuwsma,
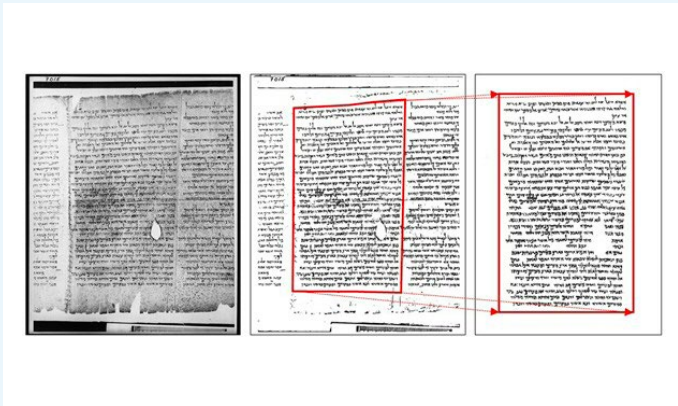Technical Director of the CIT

Opening 'Coenraad Bron Center'

News article 20 September, 2021
Name of UG's new data center refers to 'Source of knowledge'.

News article 16 December, 2021
'UG HPC Data Center on the Groningen Zernike Campus completed'

Dead Sea scrolls

Discovered some 70 years ago, the Dead Sea scrolls contain the oldest manuscripts of the Hebrew Bible and other, previously unknown, Jewish texts. The scrolls' authors did not sign their work, but based on their handwriting, researchers suspect that some manuscripts were produced by a single author. UG researchers used a combination of science and humanities to crack the writer's code and identify this author. For example, they used digital images to perform various types of computer analyses, down to the micro level of individual letters. To analyse these huge data sets, the researchers used Peregrine, the University's computing cluster.

> *News article 21 April 2021*
> 'Cracking the code of the Dead Sea scrolls'

European Environment for Scientific Software Installations – HPC software repositories

The European Environment for Scientific Software Installations (EESSI) is a partnership bringing together a number of European partners in the HPC community. As initiator and partner, the UG's HPC team is closely involved in this project, which aims to create joint research software installations that can be used on different platforms, such as personal workstations, cloud environments, and super computer

infrastructure. Thanks to this collaboration, the partners can make efficient use of each other's knowledge and networks. The CIT HPC team also co-authored a recently published research article on EESSI.

> *Research article EESSI*
> A cross-platform ready-to-use optimized scientific software stack

> *Scientific output of the Peregrine cluster*
> This page offers an overview of research articles produced on the basis of work conducted by the UG Peregrine computing cluster and other computing work supported by CIT staff.

Peregrine compute cluster 2021 – Key figures

Number of active Peregrine users over the last years

Year	Number of active users
2019	824
2020	930
2021	1069

Number of active research groups over the last years

Year	Number of groups
2019	183
2020	259
2021	265

| Peregrine training	
Basic Peregrine course	10x
Advanced Peregrine course	2x
UMCG cluster course	2x
Total number of attendees:	>150

Walk-in sessions
27 individual sessions in 2021

Geodienst

Portable Antiquities of the Netherlands

Portable Antiquities of the Netherlands (PAN) is the most complete and advanced overview of hobby archaeological findings in the Netherlands. Our Geodienst team has supported PAN by designing and building a number of interrelated systems, databases, and websites. This year, PAN was further extended with functionalities for publishing and making maritime finds accessible on the website. This was done in collaboration with VU Amsterdam and the Cultural Heritage Agency of the Netherlands (Rijksdienst voor het Cultureel Erfgoed). The extra option of importing finds from the North Sea has increased the scientific value of the data.
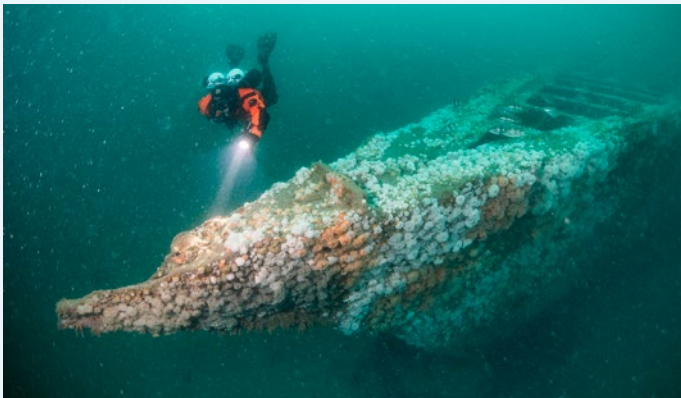
Wreck finds in PAN

Connected Contests – Ancient Athletes online

Sjoerd de Jong and Paul Haan (Geodienst), Cristian Marocico and Jonas Bulthuis (Data Science)

Connected Contests is a website with an intuitive database that has been available since 2017, where researchers can find information about athletics and festivals in Ancient Greece. In the third phase of this long-term project, the Geodienst and data science teams are working on extending the database with additional festival dates and integrating spatial search assignments and interface to allow for user input.

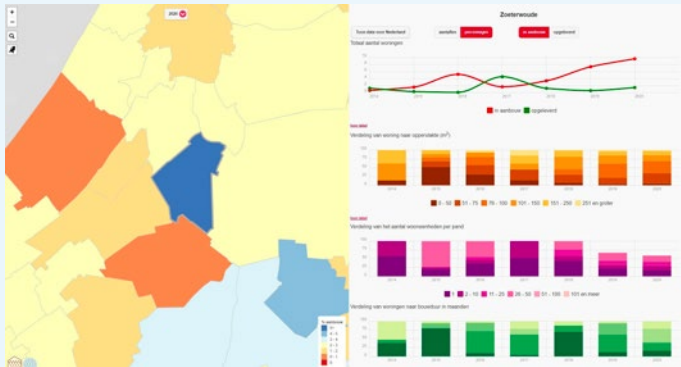

Database Connected Contests

Spatial Data Infrastructure

The Geodienst team is responsible for managing the UG Spatial Data Infrastructure (SDI). The SDI is a network of dozens of applications, web services, databases, and storage systems that allow researchers to work with large data sets that would otherwise be hard to use. This integration of geographical databases, applications, and knowledge creates a coherent system with access to thousands of datasets. This makes it easier to share research, as well as run analyses and produce visualizations. An example of the use of the SDI is the New Building Monitor.

New Building Monitor

Prof. Arno van der Vlist (Faculty of Spatial Sciences), Leon van der Meulen, Ronnie Lassche, and Sjoerd de Jong (Geodienst)
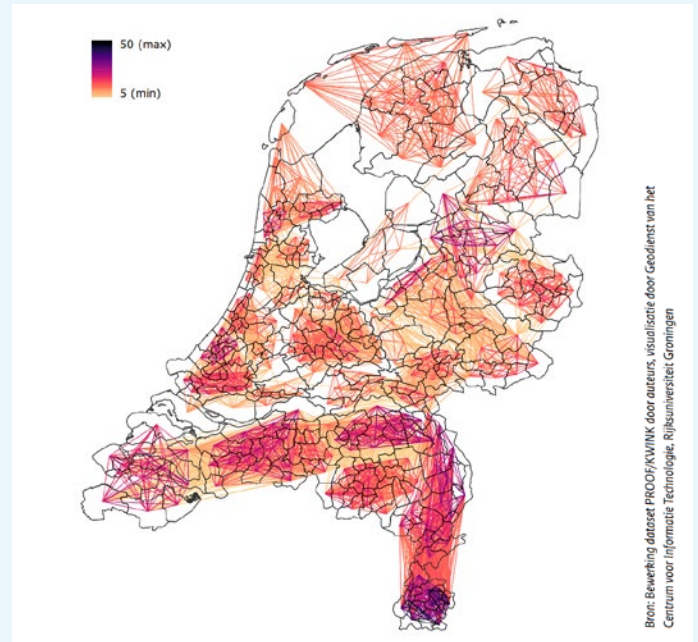
In response to research questions from the Faculty of Spatial Sciences and in collaboration with Merkator, the Geodienst team developed the New Building Monitor. This application consists of a colourful dashboard that provides a status overview of all new building projects at the national and municipal levels. The monitor provides basic information concerning the number, type, surface area, and construction duration of both delivered buildings and buildings under construction per year (since 2014). A unique feature is that the monitor now also includes corporate buildings.

Impactstory data.overheid

New Building Monitor

Bron: Bewerking dataset PROOF/KWINK door auteurs, visualisatie door Geodienst van het Centrum voor Informatie Technologie, Rijksuniversiteit Groningen

Administrative Regional Ecosystems (Bestuurlijk Regionale Ecosystemen), administrative partnership

Prof. Caspar van den Berg and Sofie Dreef (PhD student) (Campus Fryslân), Sylvia de Boer (Geodienst)

The Dutch government system seems at first sight to be clear and well-organized: a central government, provincial governments, and municipalities, complemented by water authorities, each with their own tasks and powers. In practice, however, things are a lot more complex. In addition to these three formal administrative layers, there are many partnerships at the regional level, meaning above the municipal and below the provincial level. This network of administrative regional partnerships is referred to as the Administrative Regional Ecosystem (Bestuurlijk Regionale Ecosysteem, BRE). Campus Fryslân researchers explored what these BREs look like and carried out in-depth case studies on individual BREs. Geodata scientists supported the research team by geographically mapping the BREs. They developed a model that translates data into a matrix and then geo-refers these data. The data can then be visualized in an innovative manner. This project will continue in 2022.

Facts and figures Geodienst

2,200 data sets this year, 8,000 in total

More than 45 applications

40,000 logins on online platforms

20,000 hours of ArcGIS Pro-use

108,000 addresses geo-coded via ArcGIS Online
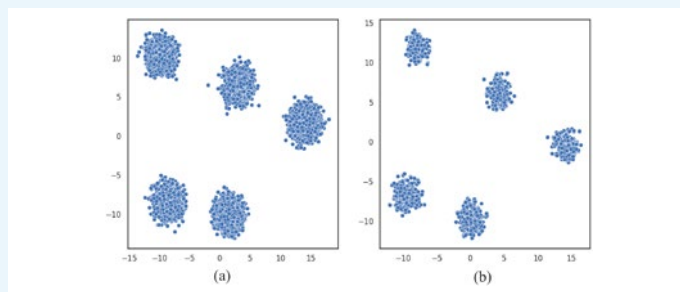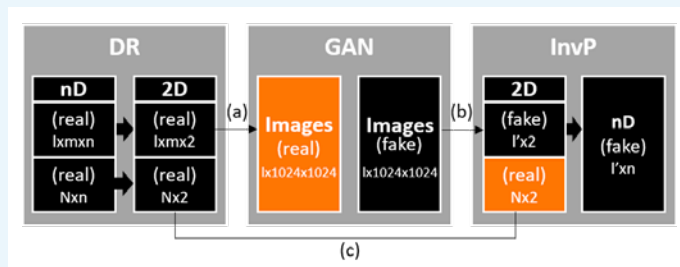
Data Science

Bacterial promotors

Anne de Jong and Jan Kok (Faculty of Science and Engineering) and Hung Chu, Cristian Marocico, Leslie Zwerwer, and Nicoletta Giudice (Data Science)

Understanding bacterial gene expression and regulation is a growing area of interest in healthcare, food processing, and other bio-industry sectors. Data from next-generation sequencing (RNA-seq) provide insight into the activity of all genes in an organism at a specific moment by revealing gene expression levels. Predicting functional genetic elements can help to improve and optimize processes, without the need for further experiments. With the help of deep learning models, the Data Science team developed an algorithm for predicting promotors in bacterial genome sequences. This kind of tool did not exist until now, and it is very much in demand, among other things because it makes it possible to both better understand and combat pathogenic (multi-antibiotics-resistant) bacteria, as well as to improve the use of 'healthy' bacteria used worldwide, for example in the food and probiotic sector.

Data synthesis pipeline

Youngjoo Kim (PhD student, Faculty of Science and Engineering), Hung Chu (Data Science), Frank Pijpers and Barteld Braaksma (Statistics Netherlands) and Jos Roerdink, Alex Telea, and Scott Trager (Faculty of Science and Engineering)

As general provisions around data protection grow more stringent, access to the analysis of complex microdata is becoming more of a challenge. Although aggregate data can provide users with high-level data representations, in-depth analysis is limited due to the lack of individual data. Furthermore, aggregate data are affected by the integrity of the original data, which cannot be fully verified without accessing them. In an effort to address these limitations, our team used Generative Adversarial Networks to create synthetic data from the original microdata. These synthetic data can then be used to perform analyses and develop models, without having to share the original microdata.

Data-driven infection management at the intensive care unit

Christian Luz and Sinha Bhanu (UMCG), Leslie Zwerwer and Nicoletta Giudice (Data Science)

Infection-related consultations at intensive care units (ICU) are an important cornerstone in the care for critically ill patients with established or suspected infections. Timing is essential, but until now consultations have been largely event-triggered and reactive. In this project, we supported the research team by using a proactive approach to predict infection-related consultations, using machine-learning models and routine electronic health records (EHR). This empirical work resulted in a paper that is currently under review by NPJ Digital Medicine.
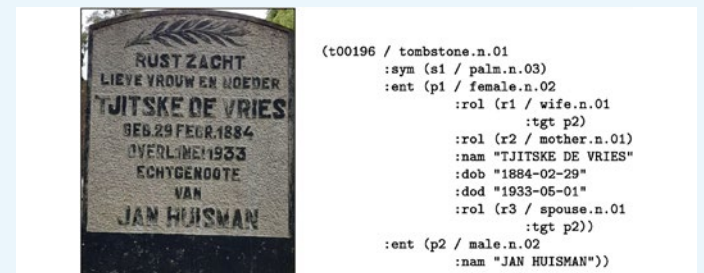
Automatically reading tombstones

Johan Bos (Faculty of Arts) and Cristian Marocico, Emin Tatar, and Yasmin Mzayek (Data Science)

Tomb inscriptions are of great interest to researchers from various humanities disciplines, and there is an urgent need to digitize tombstones, since many disappear or become illegible over time. Transcribing and interpreting inscriptions is time-consuming and requires skill and domain-based knowledge, so there are clear benefits to be gained from automating the process. We are currently working on developing a pipeline consisting of several deep-learning algorithms that can detect different components of a tomb inscription, subject them to a process of optical character recognition, extract the texts, and finally create a standard representation of the inscription in Penman notation.

FUTURE DEVELOPMENTS

The multidimensional research issues and complex societal problems that UG aims to help solve require an interdisciplinary approach. Our expectation is that Open Science and FAIR data will make an important contribution to promoting collaboration within and beyond the University. All of this requires the support of a modern and reliable IT infrastructure, tools, and applications. We expect the Digital Competence Centre (DCC), launched earlier this year, to play an important role in the coming year in providing optimal facilities for the storage and reuse of data in accordance with the FAIR principle.

The DCC is responsible for the FAIR Data & Software work package within the Open Science programme. This programme was commissioned by the Board of the University with the goal of stimulating Open Science within the University and fully integrating it into the research and teaching culture of the UG within two years.

Digital Competence Centre

This year, the DCC will invest in further professionalizing its existing services and developing new services. Below is an overview of our key activities and ambitions for 2022:

- Further designing the DCC website and knowledge bank for clear and accessible information provision for researchers and support staff, with information on DCC services and data-based themes such as FAIR data management, privacy, data security, Open Science, and research-related IT.
- Contributing to the realization of the UG's strategic objectives by stimulating Open Science and FAIR data. In this context, the DCC is responsible for the FAIR Data & Software pillar of the Open Science programme.
- Developing a training programme around Research Data Management according to the FAIR principles, focused on various target groups.
- Investing in community building by creating a network of local data stewards, in collaboration with faculties and institutes. This also includes organizing regular meet-ups and events for researchers and support staff.

- Professionalizing DCC services around the protection of sensitive and personal data (services catalogue, training programmes, and fine-tuning procedures and measures).
- Creating services around a Research Workspace that integrates (within a single dashboard) research administration, data storage and management, as well as access to analysis software in a safe workplace.

2022–2026 ICT Innovation Fund

With the institution of the ICT Innovation Fund, the UG continues to invest in digital innovations that are crucial to keep the University at the top of worldwide rankings. As from 1 January 2022, the CIT is inviting all UG staff members to submit a proposal for the ICT Innovation Fund. The UG has established the 2022–2026 ICT Innovation Fund to stimulate innovative ICT ideas that may grow into innovative projects for realizing the University's strategic objectives. These can be ideas in the field of teaching, research, or business operations. The ICT Innovation Fund offers support in both exploring and implementing new technological solutions. A new aspect of the ICT Innovation Fund is the Innovation Lab, a platform for generating and exchanging new ideas and projects through interaction between people with different backgrounds and roles.

ICT Innovation Fund

STAY IN TOUCH

Contact details
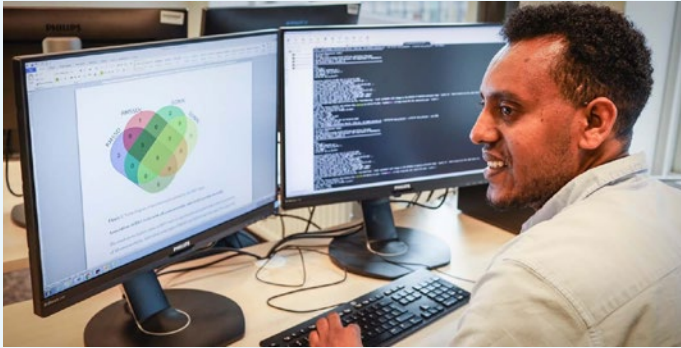
Digital Competence Centre

For any questions about research data management, the Geodienst, data science, XR, 3D, or Geographic Information Systems (GIS), please contact the Digital Competence Centre (DCC).

🔗	rug.nl/dcc

✉	dcc@rug.nl

CIT Research

All available services and products are listed in Iris, our product and service catalogue.

👁	rug.nl/iris

🔗	rug.nl/cit/research

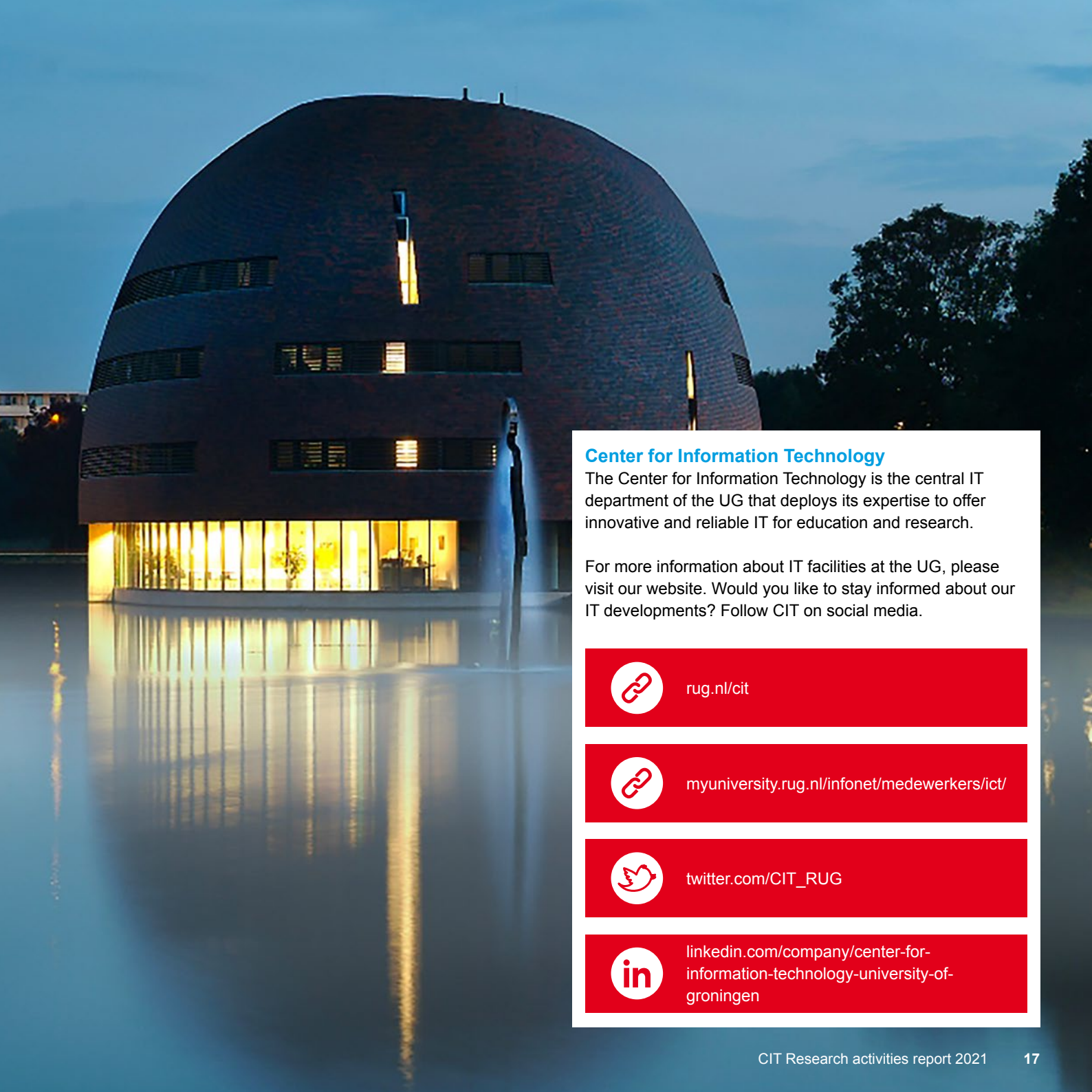|	
🐦	twitter.com/CITResearch

Center for Information Technology

The Center for Information Technology is the central IT department of the UG that deploys its expertise to offer innovative and reliable IT for education and research.

For more information about IT facilities at the UG, please visit our website. Would you like to stay informed about our IT developments? Follow CIT on social media.

rug.nl/cit

myuniversity.rug.nl/infonet/medewerkers/ict/

twitter.com/CIT_RUG

linkedin.com/company/center-for-information-technology-university-of-groningen