

# DCC als landingsplaats van eScience-producten

Wageningen, 4 februari 2020  
Netherlands eScience Center

*Joris van Eijnatten, directeur  
Tom Bakker, beleidsmedewerker*





# **Wat is het eScience Center?**

## **Wat kunnen we doen voor DCC's?**

Wageningen, 4 februari 2020  
Netherlands eScience Center

*Joris van Eijnatten, directeur*  
*Tom Bakker, beleidsmedewerker*



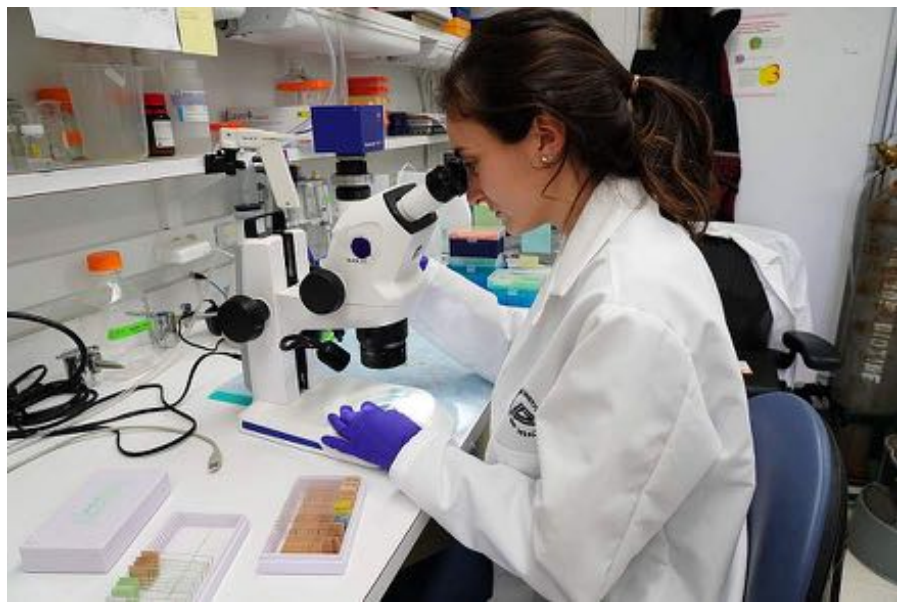




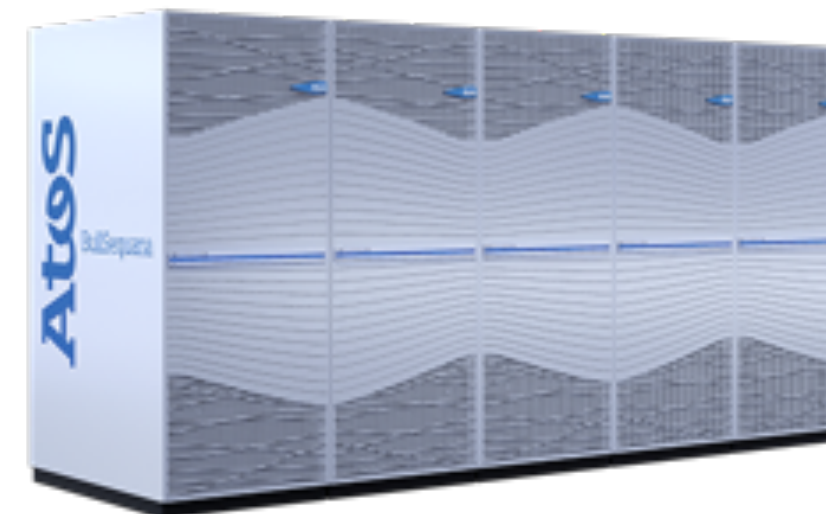
Photography: Elodie Burrillon

# Enabling digitally enhanced research

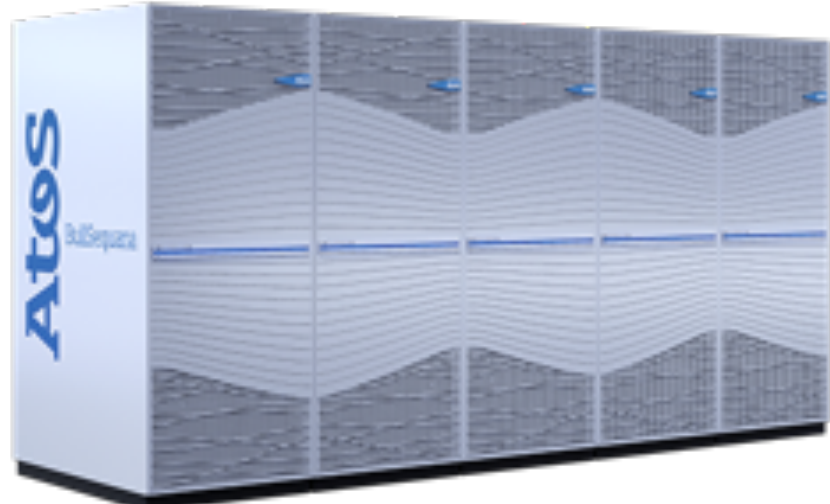
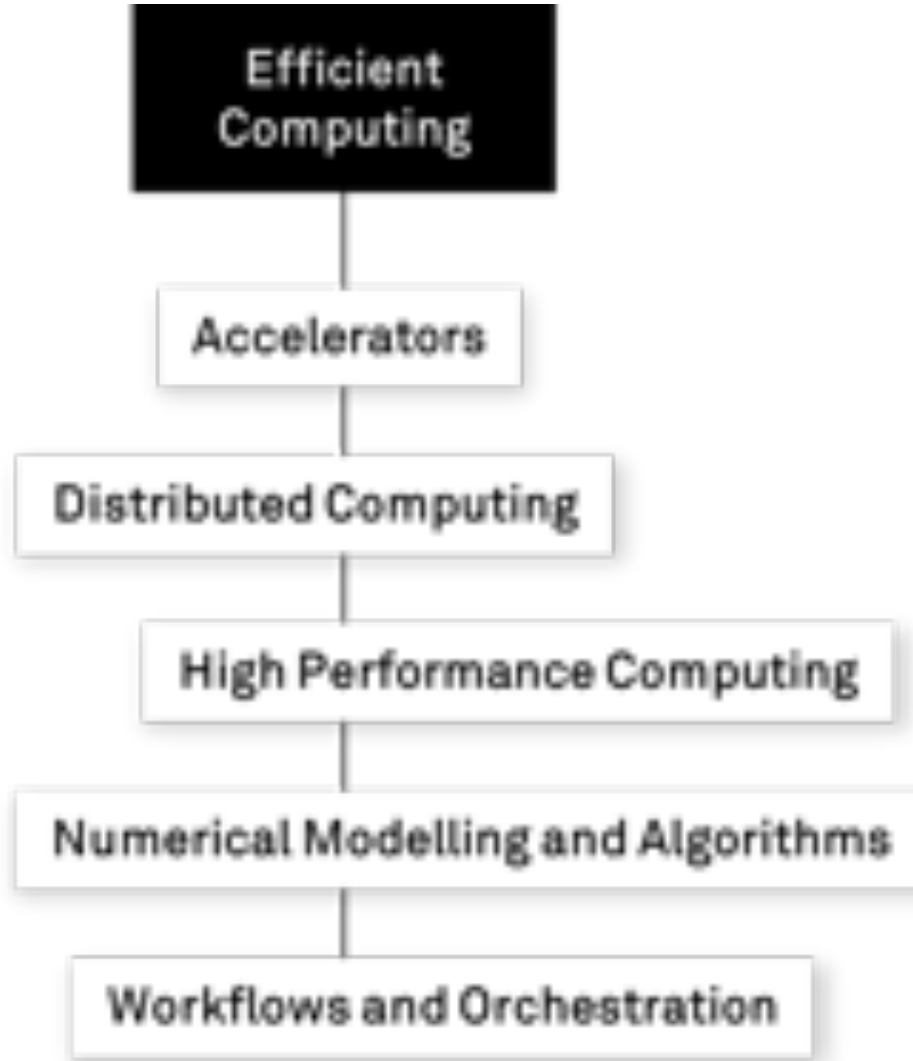
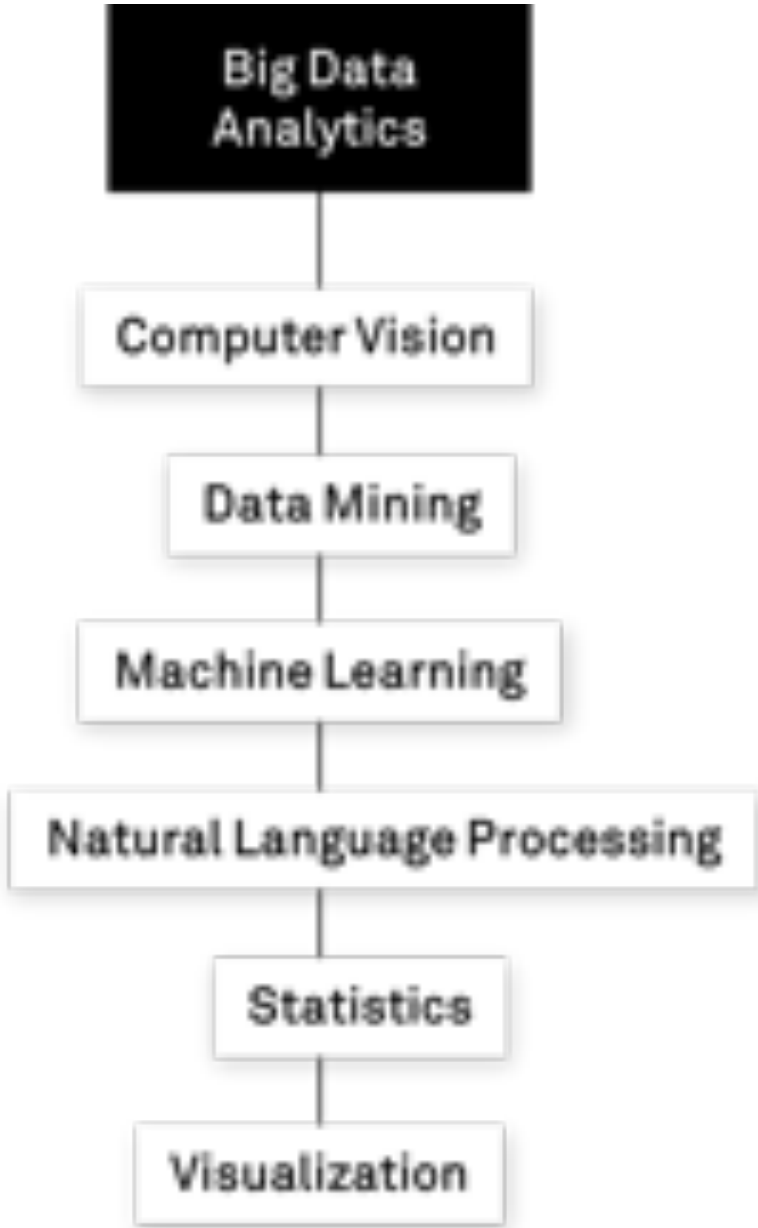
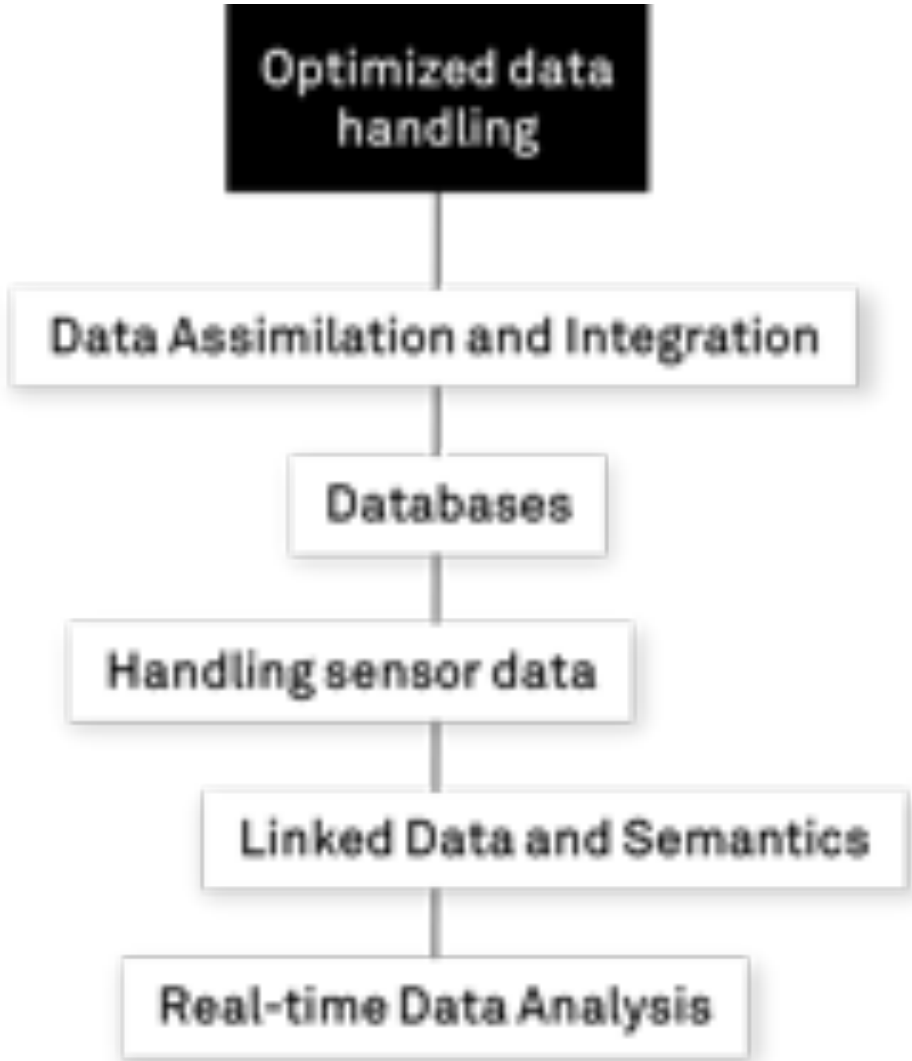
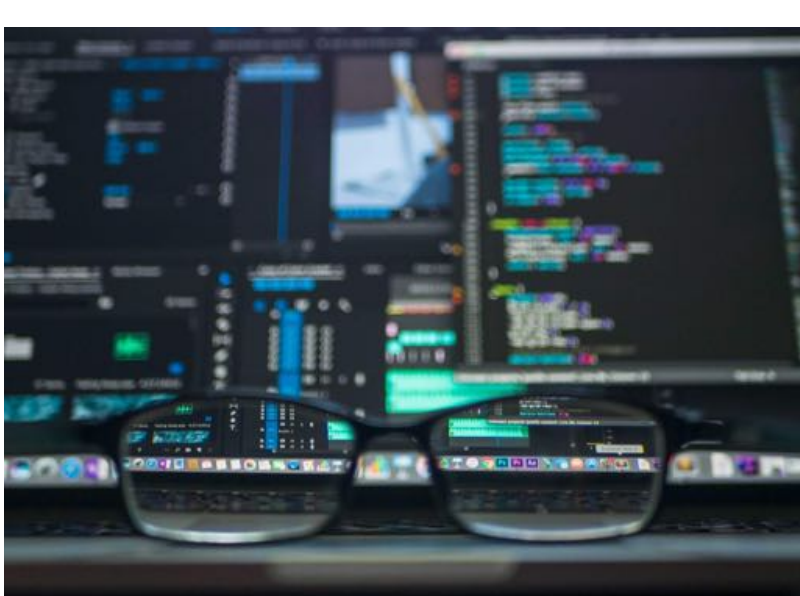




netherlands  
**eScience center**









What are DCCs?

- ✓ generic term: infrastructure, data, software, research
- ✓ where are they?

eScience Center's new strategy?

- ✓ Spring 2020

What do we need to make this work?

- ✓ peer to peer communication: Research Software Engineers (RSEs)
- ✓ softwareness
- ✓ mutual commitment

- Onderzoeksprojecten
- Workshops & trainingen
- Online resources
- Netwerken & communities

Samenwerking met DCC's  
voor netwerk-effecten?





Science center Funding Expertise Projects People About

# eScience Projects

Current Completed All Discipline

**ESWACE2**  
For future accurate climate and weather predictions

**IS-ENES3**  
Providing the infrastructure to better understand and project climate variability and change

**TADPOLE-SHARS**  
Optimizing TADPOLE's algorithms for Feature and Evaluation

**DTL Semantic Analysis of radiology Reports utilizing Lexicon**  
Uncovering large volumes of knowledge locked in natural text

**Digital twins: monitoring ship's state in real-time**  
Advanced data science to assist the design of cleaner, safer and smarter ships

**European Climate Prediction system**

**ePODSUM**  
Early Prediction of Dyslexia in Infants Using Machine Learning

**Understanding visually grounded spoken language via multi-tasking**  
An alternative approach for intelligent systems to understand human speech

**Monitoring tropical forest recovery capacity using RADAR Sentinel**

**MODARC**  
Modeling tree level and mortality

**PROCESS**  
Providing Computing solutions for

**Enhance Your Research Alliance (EYRA) Benchmark Platform**

[www.esciencecenter.nl/projects](http://www.esciencecenter.nl/projects)





### 1. Skills & open science training

- Version control, unit testing
- Reproducibility, Jupyter, myBinder
- Software licences
- Train-the-trainer



### 2. Technology workshops & SIGs

- Machine learning & AI
- Containers & Virtualization
- GPU, HPC, cloud computing

### 3. eScience Center software & tools

- McFly (machine learning for time series data)
- Xenon (middleware for computing resources)
- Case Law (visualisatietool Rechtspraak.nl)





F A I R

12345

ENDORSE

FIVE RECOMMENDATIONS FOR FAIR SOFTWARE

LET'S GO! →

ENDORSE



WHAT'S FAIR

ABOUT US

fair-software.nl

What can you do to be more FAIR?



#1

USE A PUBLICLY ACCESSIBLE REPOSITORY WITH VERSION CONTROL

GitHub, Bitbucket, GitLab



#2

ADD A LICENSE

tidrlegal.com, choosealicense.com



#3

REGISTER YOUR CODE IN A COMMUNITY REGISTRY

bit.ly/awesome-registries



#4

ENABLE CITATION OF THE SOFTWARE

CodeMeta, The Citation File Format



#5

USE SOFTWARE QUALITY CHECKLIST

Basically, any checklist will do.

ENDORISING ORGANIZATIONS



DANS



eScience center



KB nationale bibliotheek



NWO



DTL



elixir



## Online resources

- version control
- branching model
- coding style
- testing
- code quality tools
- releases
- licenses
- packaging
- citation

<https://guide.esciencecenter.nl>  
<https://the-turing-way.netlify.com>

The image shows a screenshot of a web browser displaying the 'Netherlands eScience Center Guide' at <https://guide.esciencecenter.nl>. The browser window has a single tab titled 'Introduction - Netherlands eScience Center Guide - Mozilla Firefox'. The address bar shows the URL. The page has a search bar at the top left and a navigation menu on the left side. The main content area is titled 'Guide' and contains a paragraph: 'This is a guide to software development and projects at the Netherlands eScience Center. It both serves as a source of information for exactly how we work at the eScience Center, and as a basis for discussions and reaching consensus on this topic.'

Below the main content, there is a detailed view of 'The Turing Way' section. It includes a list of 12 topics: 1. Introduction, 2. Reproducibility, 3. Open Research, 4. Version Control, 5. Collaborating on GitHub/GitLab, 6. Credit for reproducible research, 7. Research Data Management, 8. Reproducible Environments, 9. Testing, 10. Reviewing, 11. Continuous Integration, and 12. Reproducible Research with Make.

To the right of this list is a section titled 'Welcome to the Turing Way' with the following text: 'The Turing Way is a lightly opinionated guide to reproducible data science. Our goal is to provide all the information that researchers need at the start of their projects to ensure that they are easy to reproduce at the end. This also means making sure PhD students, postdocs, PIs, and funding teams know which parts of the "responsibility of reproducibility" they can affect, and what they should do to nudge data science to being more efficient, effective, and understandable.'

Below the text is a colorful illustration of a community. It shows a group of people interacting in a park-like setting with a large tree, a small house, and a sign that says 'OUR COMMUNITY'. Speech bubbles indicate various activities like 'SOLVING PROBLEMS TOGETHER', 'DISCUSSING RESEARCH', 'WELCOMING NEW PEOPLE', and 'SHARING KNOWLEDGE'. The illustration is signed 'S. B. 2014'.



## Special Interest Groups

- Recurring meetings on technology areas
- Sharing knowledge and new technological developments
- Invited speakers from industry and academia
- Internal workshops, hackathons, following tutorials and MOOCs, etc.

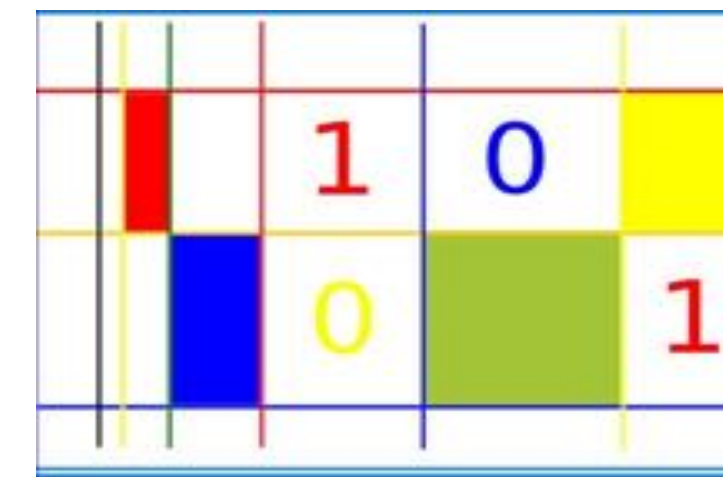
Natural Language Processing



Analytics & Data Mining



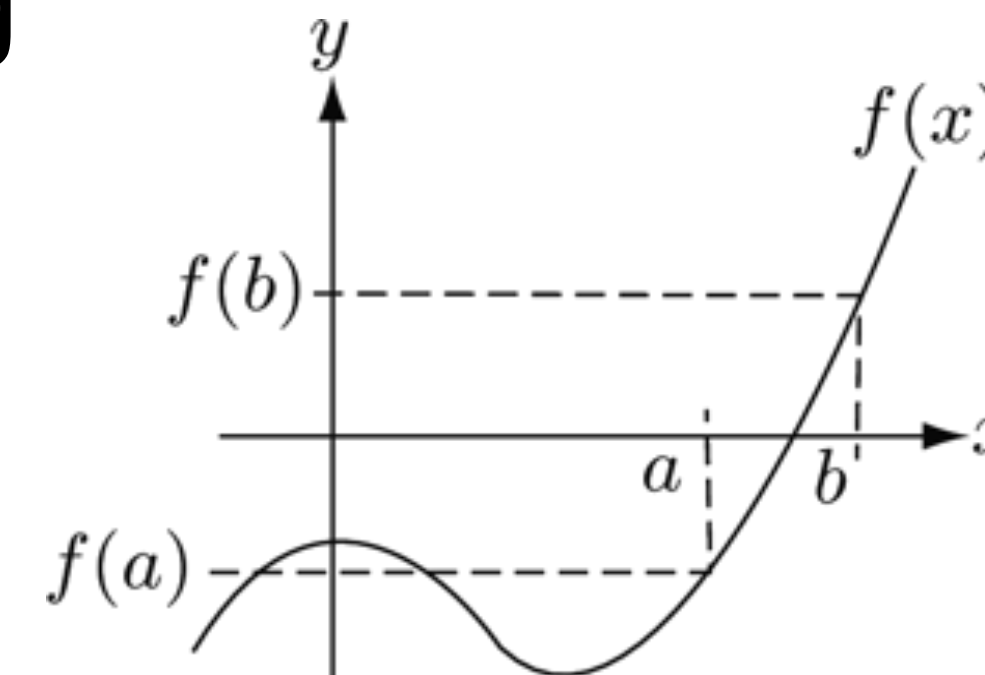
Data



Machine Learning  
& Computer Vision



Visualization



Numerical  
techniques



Workflows and Orchestration



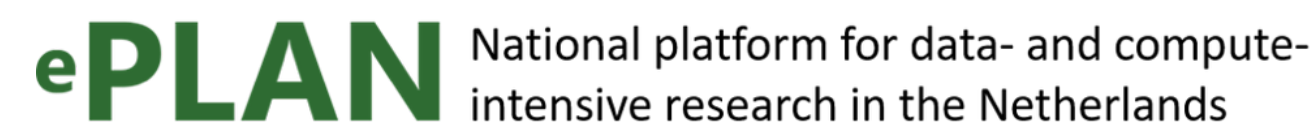
Efficient Computing





- Nationaal netwerk van Research Software Engineers (RSEs)
- Internationaal: Scandinavië, V.S., Duitsland, UK
- NL: 35+ instellingen, ~200 leden
- Activiteiten: meetups, programmers meetings, conferentie

*en verder..*





## Content Management System for research software

- Free, Open
- Customizable
- Rich metadata

## Aggregate data from many sources:

- GitHub, Zenodo, Zotero, CFF, Medium, YouTube, ...

## Expose this data using standard formats

- schema.org, CodeMeta, OAI-PMH, ...

## Provide citation information:

- BibTex, EndNote, RIS, ...



NARCIS



The screenshot displays the GGIR software entry on the Research Software Directory. The page includes a description of GGIR as a tool for converting raw wearable data into reports for physical activity and sleep research. It features a line graph showing mentions from 2015 to 2019, with a peak in 2018. The 'Cite this software' section provides options to copy the DOI (10.5281/zenodo.3474227) and download citation files in BibTeX, EndNote, or RIS formats. The 'What GGIR can do for you' section lists key features and usage statistics. The bottom of the page lists participating organizations, including Activinsights, Inserm, MRC, UCL, and the University of Leicester.



- Onderzoeksprojecten
- Workshops & trainingen
- Online resources
- Netwerken & communities

Samenwerking met DCC's  
voor netwerk-effecten

