

Making Research Software a First-Class Citizen in Research

Anton Akhmerov, Maria J. Cruz, Niels Drost, Cees Hof,
Tomas Knapen, Mateusz Kuzak, Carlos Martinez-Ortiz,
Yasemin Turkyilmaz-van der Velden

DUTCH
TECHCENTRE
FOR LIFE SCIENCES



eScience center



 TU Delft

VU  VRIJE
UNIVERSITEIT
AMSTERDAM

Data Archiving and Networked Services

DANS



Who we are



Maria J. Cruz

Community Manager RDM
Library VU Amsterdam
m.j.marquesdebarroscruz@vu.nl



Yasemin Türkyilmaz-van der Velden

Data Steward
TU Delft Faculty of 3mE
y.turkyilmaz-vandervelden@tudelft.nl



Carlos Martinez-Ortiz

eScience research engineer
Netherlands eScience Center
c.martinez@esciencecenter.nl

Who we are



Anton Akhmerov
Associate Professor
TU Delft Faculty of Applied Sciences
A.R.Akhmerov@tudelft.nl



Mateusz Kuzak
Scientific Community Manager
Dutch Techcentre for Life Sciences
mateusz.kuzak@dtls.nl



Tomas Knappen
Assistant Professor
VU Faculty of Behavioural and Movement Sciences
t.knappen@vu.nl

Who we are



Niels Drost

Senior eScience Research Engineer
Netherlands eScience
N.Drost@esciencecenter.nl



Cees Hof

Project Acquisition Manager
DANS-KNAW
cees.hof@dans.knaw.nl

Making Research Software a First-Class Citizen in Research

“We argue that research software should be treated as a first-class research output, in equal footing to research data. Research software and research data are both fundamental to contemporary research. However, the recognition of the importance of research software as a valuable research output in its own right is lagging behind that of research data.”

By Mateusz Kuzak, Maria Cruz, Carsten Thiel, Shoaib Sufi, and Nasir Eisty.

<https://software.ac.uk/blog/2018-11-28-making-software-first-class-citizen-research>

28 November 2018

Why are we here today and what do we want to talk about ?

We want to highlight the importance of research software in contemporary research and its relationship to research data, open science, and reproducibility.

Our key messages to NWO

- If **open science** is to lead towards better, more transparent, and reproducible research, then **research software**, research data and publications need all to be treated on equal footing at the policy level.
- NWO should make more explicit policies for supporting and evaluating research software.
- There are some differences when it comes to research software preservation (Software Sustainability) as compared to data preservation, but it shouldn't be more difficult.
- **We are here to provide advice and to help!**

Research Software is
Fundamental to Research

Software is fundamental to research

- 92% of academics use research software
- 69% say that their research would not be practical without it
- 56% develop their own software
(worryingly, 21% of those have no training in software development)

Data and software are
intrinsically linked

Data and software are intrinsically linked

- Digital data are completely inaccessible without software – a fact that is often underappreciated.
- Data stewardship and software sustainability are distinguished mostly by the notion that data need to be kept *as is* while software needs to be maintained in order to remain useful.
- **It is essential for the future use and re-use of data to process and manage data and software on equal footing, policy-wise and practically.**

Source: P. Aerts (NLeSC) and P. Doorn (DANS), A conceptual approach to data stewardship and software sustainability, https://dans.knaw.nl/nl/over/organisatie-beleid/informatiemateriaal/AConceptualApproachtoDataStewardshipandSoftwareSustainability_DEF.pdf

Meanwhile at the policy level....

RESEARCH DATA – OPEN BY DEFAULT

European
Commission



HORIZON 2020 GRANTEES ARE REQUIRED

... the focus is very much on research data.

The FAIR Guiding Principles for scientific data management and stewardship

Scientific Data volume 3, Article number: 160018 (2016) DOI: 10.1038/sdata.2016.18

SCIENTIFIC DATA

OPEN

SUBJECT CATEGORIES

- » Research data
- » Publication characteristics

Comment: The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson *et al.*[#]

“Importantly, it is our intent that the principles apply not only to ‘data’ in the conventional sense, but also to the algorithms, tools, and workflows that led to that data.”

“All components of the research process must be available to ensure transparency, reproducibility and reusability.”



Netherlands Organisation
for Scientific Research

Open (FAIR) data

< Open Science

> **Data management**

Data management section

> Contacts

> Plan S

Besides publications also research data that emerges from NWO-funded research should be as accessible and reusable as possible. The idea behind open science is 'Open as possible, closed if necessary.' Due consideration is given to aspects such as privacy, public security, ethical limitations, property rights and commercial interests.

What about Research Software?

Research Software and Reproducibility

Why is this important?

IS THERE A REPRODUCIBILITY CRISIS?



What factors contribute to irreproducible research?

- Selective reporting
- Pressure to publish
- Supporting data / methods/ **code not available**
- Insufficient oversight and mentoring

Tomas Knapen

*PI, Model-based Neuroimaging,
VU Amsterdam, Spinoza KNAW*

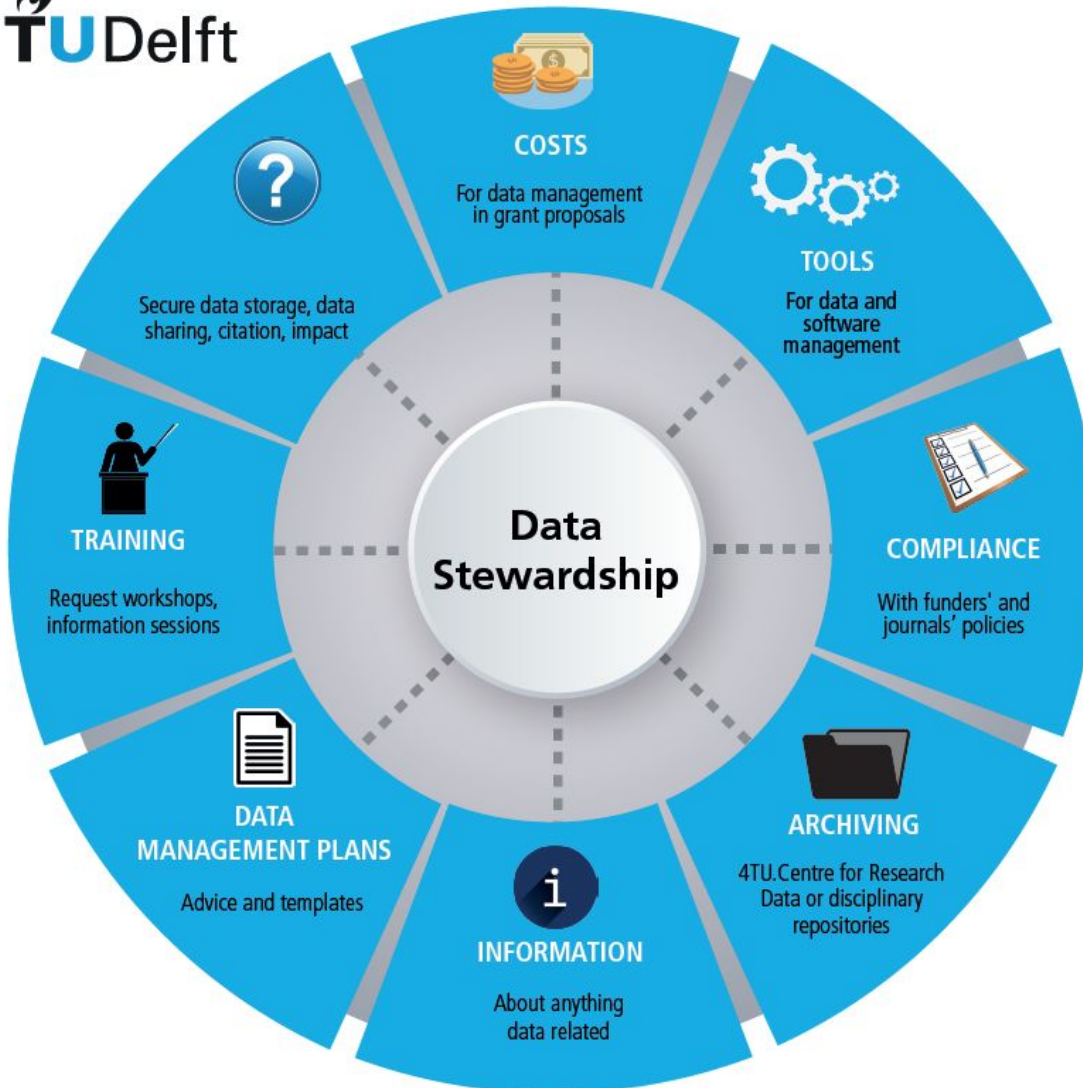
*In Brain Imaging, Open Data is becoming
mainstream, but software needs to catch up.*

*Teaching is starting, and the first true community
open-source projects are showing their research
potential, but....*

*Especially in terms of incentive structure, software
contributions are second-rate relative to
publications: this needs to change.*



Data Stewardship at TU Delft:



- Research software is an important aspect
- Joined to The Carpentries
 - Training
 - Community building

Data Champions at TU Delft: Testimony from the Lab



Anton Akhmerov

Associate Professor

Faculty of Applied Sciences

Department: Quantum Nanoscience

Areas of expertise:

Quantum nanoscience, numerical simulations

“Currently, the publication of papers, data and software are evaluated differently by the research community. Publications are the most visible output of research, followed by research data, and then finally there is software.”



Frans van der Meer

Assistant Professor

Faculty of Civil Engineering and Geosciences

Department: Materials, Mechanics, Management & Design

Areas of expertise:

Computational mechanics, coding, simulations

“In our group we develop code and use it for simulations. For the code to be re-usable even within our own research group (students come and go), it is important that it is findable and documented.”

<https://www.tudelft.nl/library/our-datachampions>

Possible policy changes

Anton Akhmerov (TU Delft)

- Allow to request an RSE position in addition to a researcher position in NWO Klein-1
Software development and maintenance is a significant effort that does not fit into a common researcher career
- Require that the software produced within NWO-funded projects is published under an open-source license, unless the researchers request a waiver.
Right now even if researchers request funding for developing commercial software, this is ignored in the evaluation.
- Allow to list software (including software maintenance) in CV, potentially alongside with other research outputs.
- Create grants for training/building software development community
- Support RSE positions in universities

The Netherlands Research Software Engineer Community

Niels Drost (NLeSC)



The Research Software Engineer

- A complement to the existing postdoctoral career structure
- Fully a part of both the scholarly community *and* an a professional software developer
- Interested in the research context and capable of understanding the scientific literature and research questions
- Measured by software output as much as through traditional academic metrics



NL-RSE community

- Following initiative in UK.
- Founded in 2018, 149 Members as of March 2019.
- Joined by DE-RSE, NORDIC-RSE, US-RSE, ...

Including RSEs from the following organizations and institutes

- AMOLF
- Amsterdam University of Applied Sciences
- ASML
- Astron
- Data Archiving and Networked Services (DANS)
- Delft University of Technology
- RIVM
- Dutch Techcentre for Life Sciences (DTL)
- Erasmus MC
- Huygens ING
- Leiden Observatory
- Leiden University Library
- Leiden University Medical Center
- Meertens Institute
- National Library of the Netherlands (KB)
- Netherlands eScience Center
- Netherlands Institute of Ecology
- Radboud University Medical Center
- Radboud University Nijmegen
- Royal Netherlands Meteorological Institute
- Science [&] Technology Corporation
- SRON Netherlands Institute for Space Research
- SURFsara
- UMC Utrecht
- University of Amsterdam
- University of Groningen
- University of Twente
- Utrecht University
- VU University Amsterdam
- Wageningen University & Research

US RSE

The US Research Software Engineer Community

What is an RSE?

We like an inclusive definition of Research Software Engineers to encompass those who regularly use expertise in programming to



Training and Capacity Building

Mateusz Kuzak (DTL)



- **the Carpentries**

teaching foundational coding, and data science skills to researchers worldwide

- **Promoting exchange of knowledge and skills**

communities of practice: study groups, hacky hours

- **ELIXIR Software Development Best Practices WG**

Open Software > Better Software > Better Research

- **4 Recommendations for Open Source Software**

Training Researchers how to build Open Source Software Projects

- **Building Capacity**

open training resources, trainers, RSEs

Guiding researchers to make FAIR software

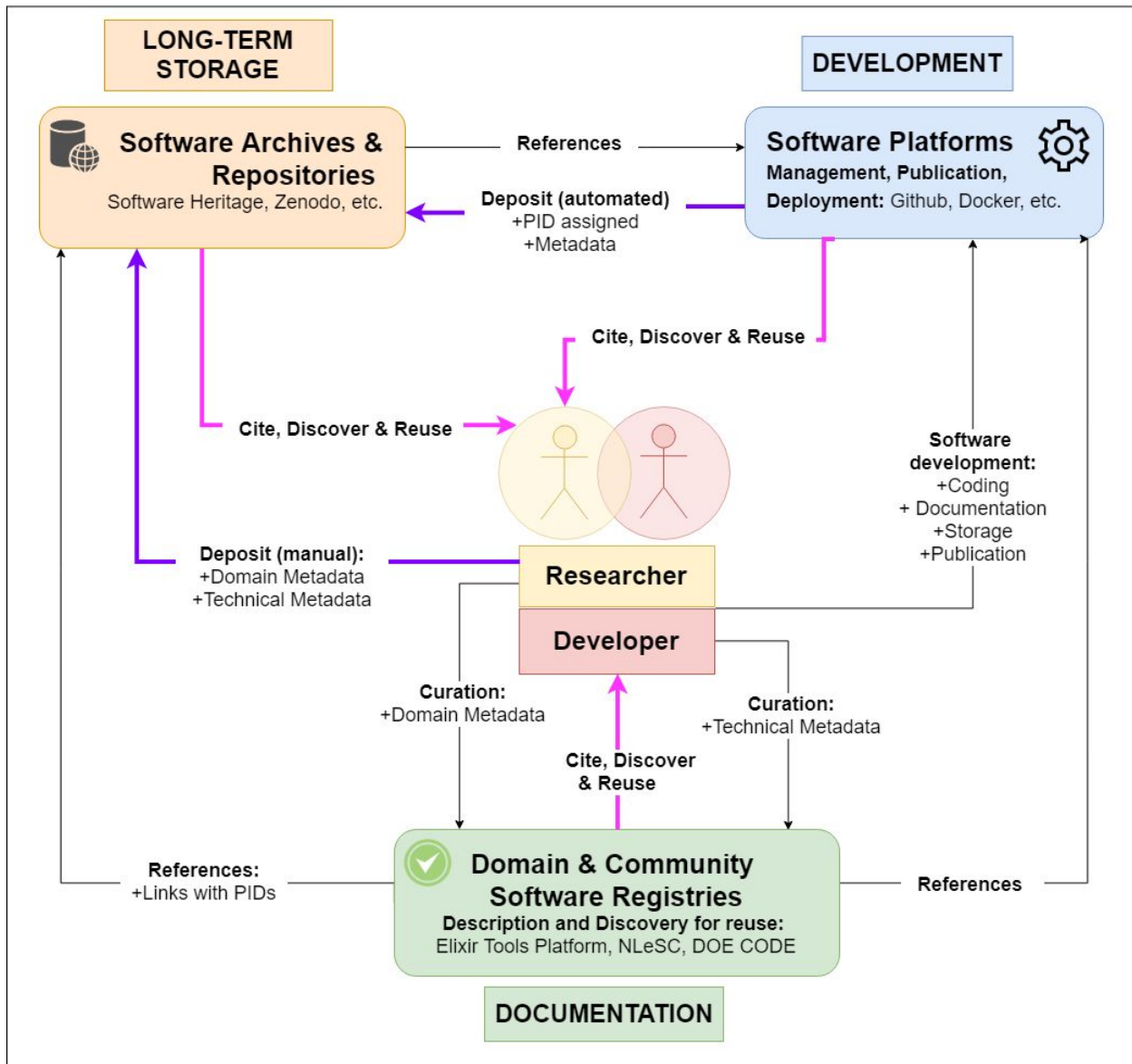
Carlos Martinez-Ortiz (NLeSC)

- FAIR principles for software
- Promoting / training in best practices
- Sharing experiences of RSEs between institutions

Cees Hof (DANS)

- Development of a FAIR Software Route
- Archiving Software as integrated research output
(research objects approach, workflow preservation, etc.)
- Monitoring software output in NARCIS
(work is in progress!)
- FAIR software (preservation) in EOSC.....

Development of a FAIR Software Route



Monitoring software output in NARCIS....

Based
on RSD
of
NLeSC

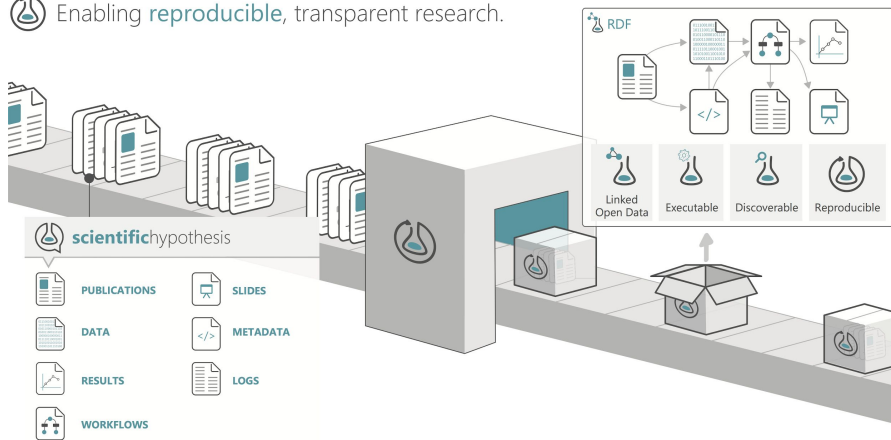
The screenshot shows the NARCIS interface for a dataset. At the top, there is a navigation bar with 'HOME', 'ABOUT NARCIS', 'LOGIN', and 'NEDERLANDS'. Below this is a blue bar with navigation links: '< BACK TO SEARCH RESULTS', '< PREVIOUS', '8 out of 8 results', and 'NEXT >'. The main content area features the 'eScience center' logo and the dataset title '3D-E-CHEM/KNIME-PYTHON-WRAPPER: V2.0.1 (2018)'. A 'Main' tab is selected. The dataset details are as follows:

Title	3D-e-Chem/knime-python-wrapper: v2.0.1
Creator	Stefan Verhoeven (Nederlands eScience Center)
Date issued	2018-02-07
Access	Open Access
Type	Dataset
Publisher	Zenodo
Abstract	<p>Abstract Python wrapper KNIME node and helpers. Used for development of KNIME nodes calling Python scripts.</p><p>Version 2 allows nodes to use Python version 2 or 3 and different serializers.</p><p>Changed</p>Replaced PythonKernel from org.knime.python to org.knime.python2 (#4)
Rights	GNU General Public License v3.0 only; Open Access
Dataset	> doi:10.5281/zenodo.1168379
Persistent Identifier	> 10.5281/zenodo.1168379
Url	> https://zenodo.org/record/1168379
DOI	> 10.5281/zenodo.1168379
Related	> https://github.com/3D-e-Chem/knime-python-wrapper/tree/v2.0.1
Related	> doi:10.5281/zenodo.597238
Metadata	> XML
Source	RSD-NLeSC

At the bottom, there is a 'CONTACT' link and a footer stating 'DANS is an institute of KNAW and NWO' with social media icons for YouTube, Twitter, and Facebook.

Archiving software/code as integrated research output

Enabling reproducible, transparent research.



Workshop on Research Objects
IEEE, October 2018, Amsterdam

Coming:

Open Repositories Conference
June 2019, Hamburg

FAIR software (preservation) in EOSC....



**EUROPEAN OPEN
SCIENCE CLOUD**

FAIR and sustainable software in:

FAIRsFAIR project
EOSC synergy project

.....
.....

Discussion