Scientific Software Center

IWR Interdisciplinary Center for Scientific Computing

UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

A research software engineering department at a German university: Four years and counting

Liam Keegan, Dominic Kempf, **Inga Ulusoy**
Scientific Software Center, Heidelberg University

ssc.uni-heidelberg.de

The Scientific Software Center

Mission Statement

The Scientific Software Center strives to improve research software development practices at Heidelberg University and beyond, to promote reproducible science and research software sustainability.

White paper:
10.5281/zenodo.10867903

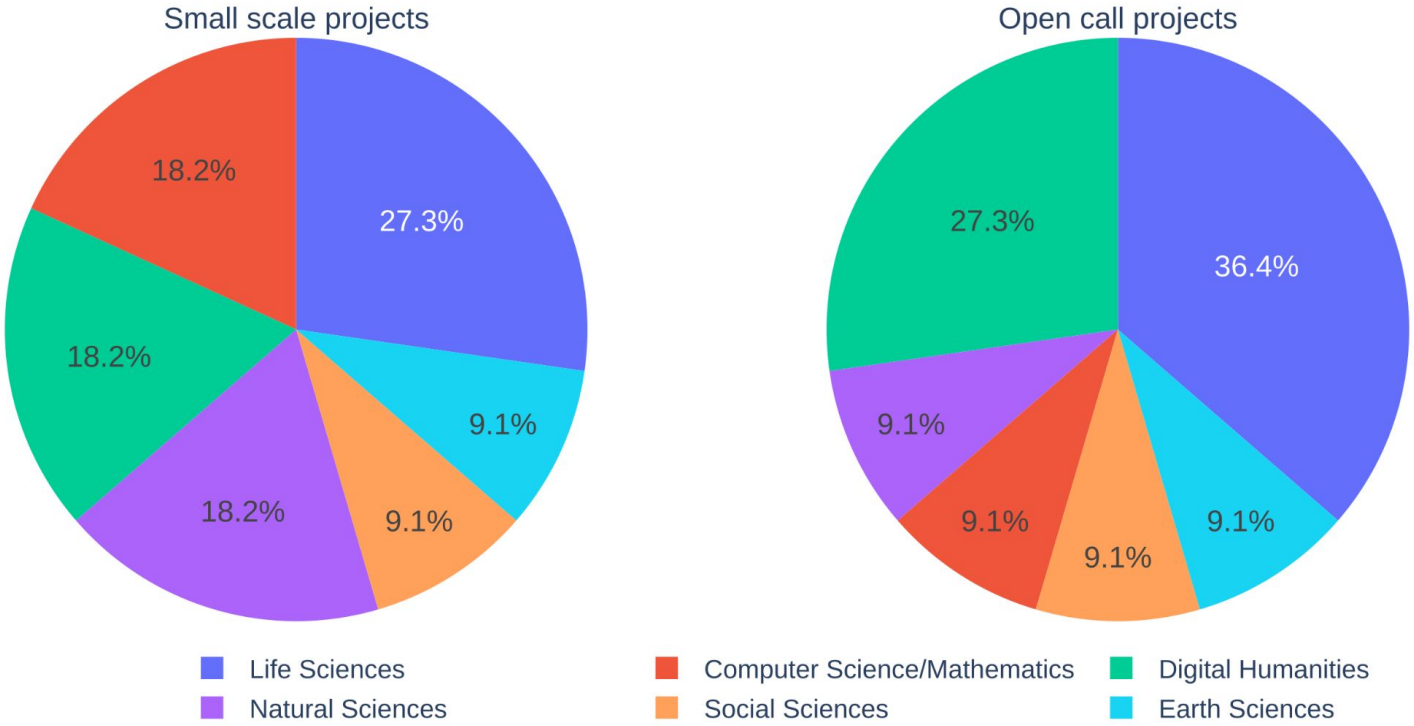
Scientific Software Center

Structure and organization of the SSC

Who we are

- Founded late 2020 with 3 RSEs by excellence funding (incubator)
- Growth through third party funding
- Team of 11 RSEs by early 2025, plus one sysadmin, one admin
- One scientific coordinator, one administrative coordinator
- Scientific board (12 people from various backgrounds/organizational roles)

Distribution of small-scale and open call projects 2021-2023



Exemplary RSE projects

A successful project: ammico

Researcher: From social (political) sciences

Research question: Analyze social media posts to identify characteristics of misinformation

Data: Social media posts (screen shots) of varying resolution, language, content

Prior work: Take screenshots from wayback machine, crop manually

Development project tech stack: Python (pandas, spaCy, transformers, LAVIS, retinaface, deepface, OpenCV, dash/plotly)

User interface: Jupyter notebook

https://github.com/ssciwr/AMMICO
DOI doi: 10.31235/osf.io/v8txj

Example: User interface

ammico

AI Media and Misinformation Content Analysis Tool (Social/Political Sciences)

TextDetector

Analyse text

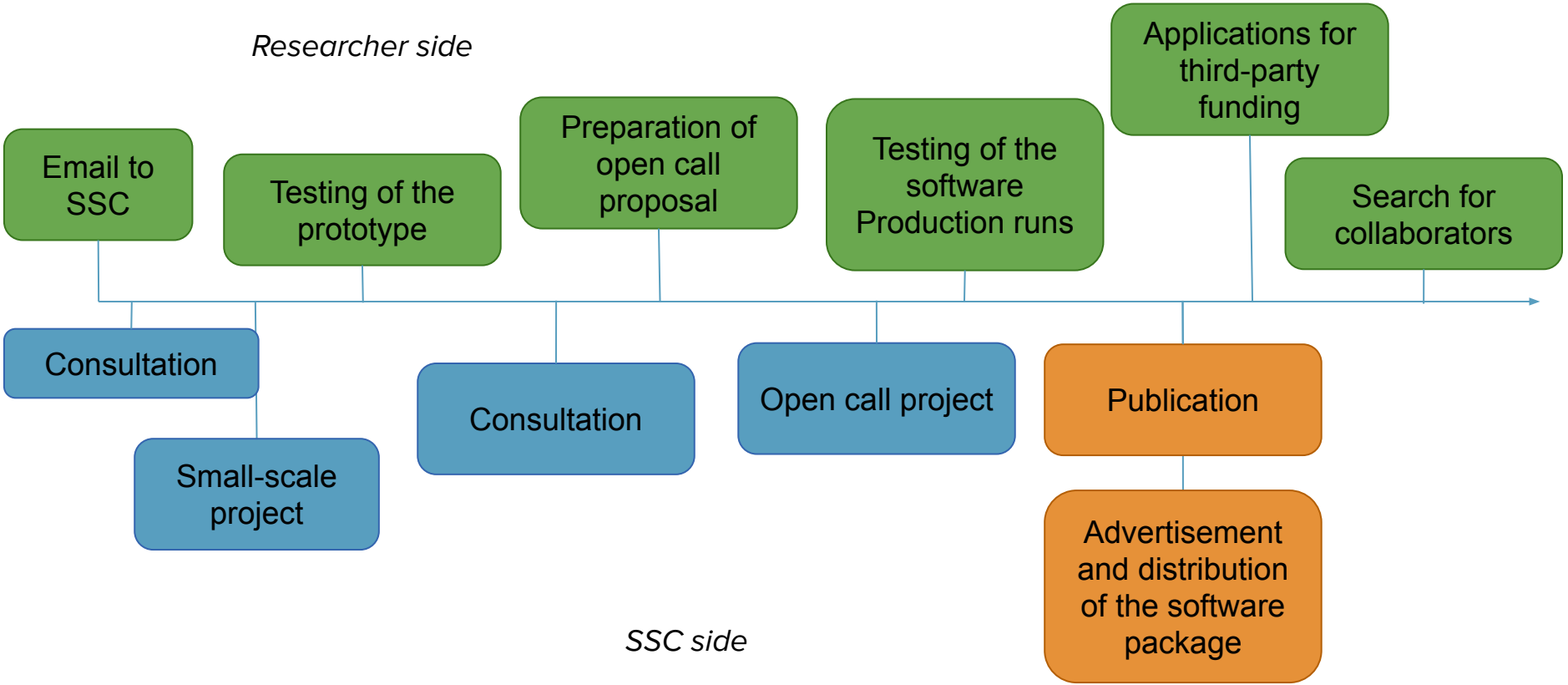
Select models for text_summary, text_sentiment, text_NER or leave blank for default:

Select model revision number for text_summary, text_sentiment, text_NER or leave blank for default:

Run Detector

filename	../../data/Image_some_text/109237S_spa.png
text	29 de septiembre CONFÍAN EN LA REUNIÓN DE HOY 0:10/0:14
text_language	es
text_english	September 29th THEY TRUST IN TODAY'S MEETING 0:10/0:14
text_clean	September 29th THEY TRUST IN TODAY 'S MEETING
text_summary	September 29th THEY TRUST IN TODAY'S MEETING 0:10/
sentiment	POSITIVE
sentiment_score	0.99
entity	
entity_type	

ammico: A timeline



Shifting requirements

Objectives July 2022:

- Text extraction and translation
- Text cleaning
- Person, face and emotion recognition, detection of gender, type of clothing
- Object detection
- Color (hue) detection

Initial requirements
Feasibility

Communication
Testing
Test data
Test results
Reference results

End of open call project June 2023:

- Text extraction and translation
- Text cleaning, summary, named entity recognition, sentiment analysis, topic analysis
- Person, face and emotion recognition, detection of gender, ethnicity, age
- Image captioning and visual question answering
- Color (hue) detection
- Multimodal search
- Cropping tool

Implemented features
Accuracy

Characteristics of an “unsuccessful project”

Researcher: Not engaged in the development process/continuous delivery, not open to adapt their own development process

Research question: Too complex or outdated, not relevant for their future research

Data: Incomplete, not available at all, not of the required accuracy or size

Prior work: Incomplete and/or not appropriate approach/”too legacy” (complex)

Development project tech stack: Not appropriate, too niche, no one with dedicated expertise available / too steep learning curve for the available amount of time

User interface: Not appropriate for the targeted user community

Many more projects on SSC homepage

Establishing a knowledge graph community in biomedical science A unifying framework for biomedical knowledge graphs	Systems Biomedicine	Funded Project	2024 - 2027	→
faunanet Build a prototype for a flexible bioacoustic platform on edge devices (raspberry pi zero in particular)	bioacoustics, edge computing, biology	Funded Project	2024	→
Kinetic Data Fitting Improve algorithm for kinetic data fitting to achieve higher accuracy and provide more descriptors of the quality of the fit.	Molecular Biology, Chemistry	Small Scale Project	2024	
Matlab Web App Deployment Consultation on deploying Matlab Web Apps both via CI-built standalone installers and via Matlab Web App Server.	Medicine	Small Scale Project	2024	↗
MONDEY MONDEY (Milestones of Normal Development in Early Years) website	Psychology	Open Call Project	2024	→
NeuroSeq Create prototype of a MIDI-like GUI to edit and simulate neuron models	neuroscience	Funded Project	2024	→
parzivAI Chatbot for German medieval language and history interaction.	linguistics, German language studies	Small Scale Project	2024	
predicTCR Web service for users to upload samples, process them, download results	medical	Funded Project	2024	→
Project W A self-hosted platform for audio transcription with OpenAI's Whisper model. Puts large emphasis on data protection in order to allow transcription of sensitive research data.	Linguistics	Software Practical, Small Scale Project	2024 - 2025	↗
QHCC Webpage for a project providing a hepatocellular carcinoma imaging data set and machine learning models.	Medical	Funded Project	2024	↗
Schreiben nach der Briefkultur: E-Mails - Dynamiken der Normierung und Standardisierung Web portal where donors can upload emails. These emails are then anonymized using the Python package anonymizer and placed in a database together with the raw data and metadata.	linguistics/romanistic/digital_humanities	Funded Project	2024	↗

Scientific Software Center

SSC management and administration

Current management of “Project SSC”

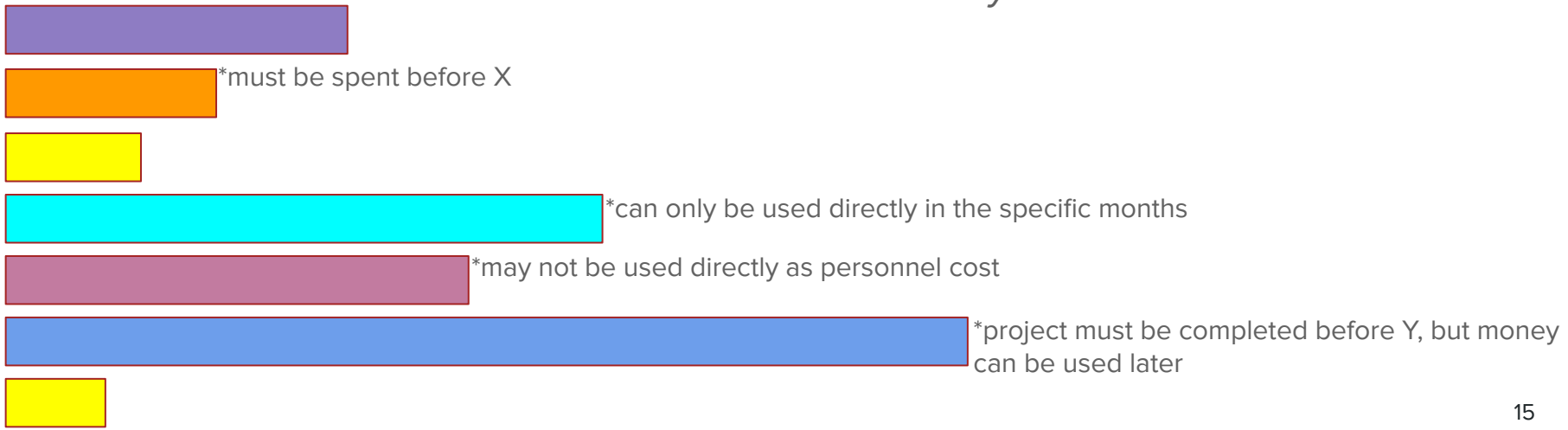
- Regular meetings with scientific and administrative coordinator (monthly)
- Regular meetings with scientific board (twice per year)
- Yearly reporting duty
- Important questions/decisions are reviewed by coordinators, critical questions/decisions reviewed by scientific board
- Day-to-day issues handled by the three core RSEs with the help of the group: A lot of ***team effort and communication/consensus***
- Assignment of projects to people and funding to people a very difficult task! We need resource pooling (money, contracts, projects)
- Recognition of RSEs and a viable career path is fundamentally important for our effectiveness

SSC: A startup in an academic setting

Administrative burden: Example

Funding (projects)

“The mythical man month”

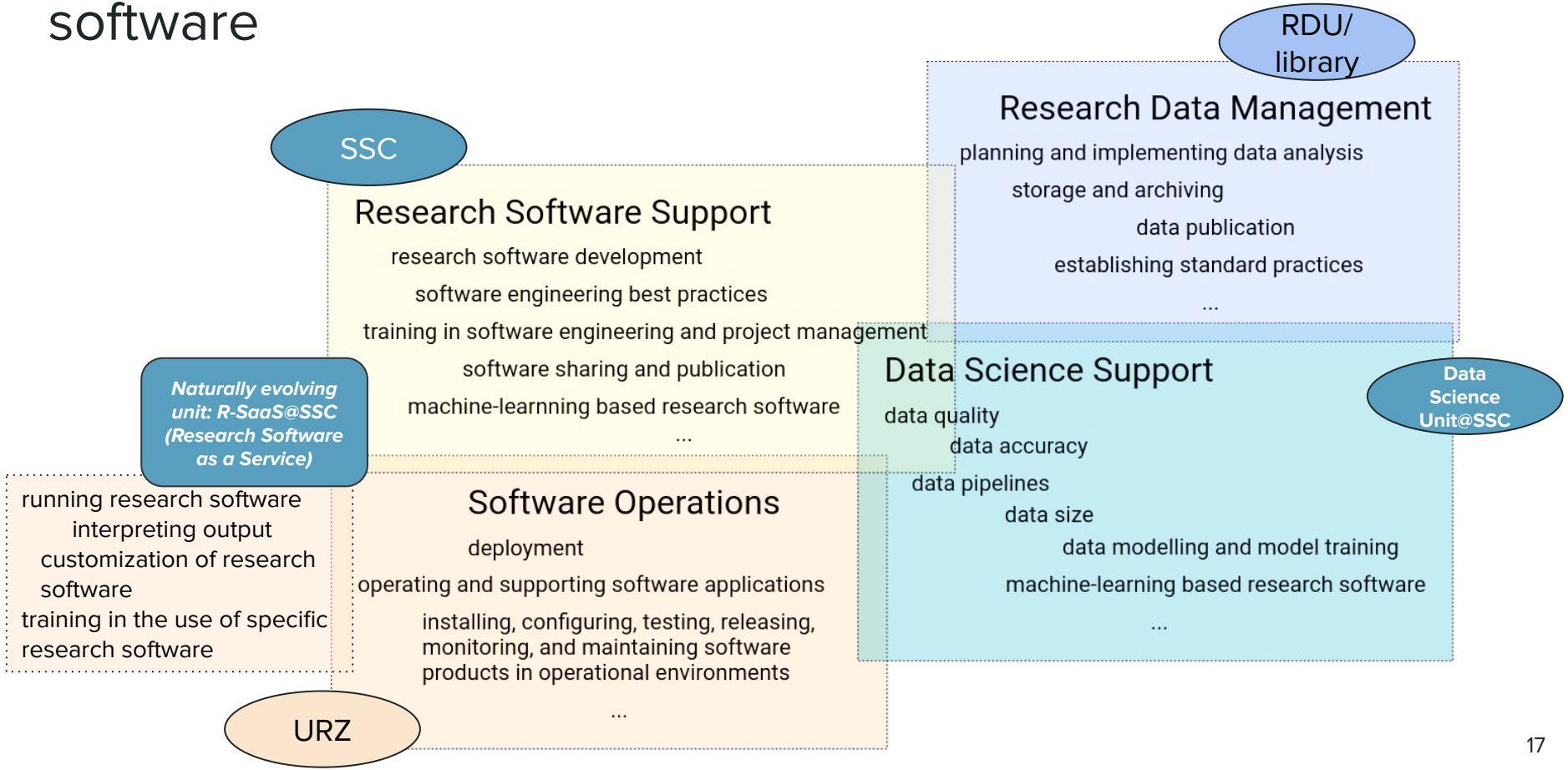


Management of SSC projects and people

- Ongoing process: Map people to projects, but also allow teams and learning
- Hiring and onboarding process has been established and is being refined with every new member
- Ownership: Every member is expected to contribute to project SSC and improve the center
- Psychological safety: Crucially important for early delivery
- Funding acquisition: A challenge in itself
- Different team constellations and team roles
- A lot more than “only IT”: There could be so much more support for the different roles (the “RSE unicorn”:
  Software Engineer + Project Manager + Team Manager + Domain Knowledge + Funding Acquisition + Teaching + Training + …)

RSE Unicorn:
Society of Research Software Engineering, Merchandise
https://socrse.myspreadshop.co.uk/all

Types of support needed for research data and software



Training and knowledge transfer

Training

- Offer many smaller compact courses that PhD students / Postdocs / Master students can squeeze in
- Very hands-on, practical, publicly available (todo: Integrate available material into software carpentries)
- Participants mostly highly motivated & excellent feedback
- Participants from all career stages, domains, and levels of expertise: extremely heterogeneous, which is challenging
- SSC Fellows mentoring program

Example: Courses

Compact Courses: Software Engineering Best Practices

- The Unix Shell
- Version Control with git
- Open Source Licensing
- Automated Testing with GitHub Actions
- A short tour of sustainable software development
- Containers in Science: Using Docker and Singularity
- Advanced Topics in Version Control with git
- Effective Software Testing
- AI in research software
- Generative AI for writing (research) software

Compact Courses: Language-specific

- Python Best Practice
- Introduction to Python Testing
- Data Exploration with Python and Jupyter
- Python Packaging

Compact Courses: High-performance computing

- Performance Benchmarking C++ Applications
- High Performance C++

Seminar Series

- Lunch-time Python

Block Courses

- Scientific Software Development

Knowledge transfer

- Handover of “finished” projects
- Responsible person must be identified early and involved in the project
- Establishing processes and tools in the research group
- Sometimes met with reluctance and unwillingness to change: not everything is perfect
- Early involvement of the researchers, back-and-forth (early delivery) and process transparency helps with the transfer
- Sometimes maintenance of completed project not guaranteed

Challenges and outlook

Challenges: Interdomain communication

I want to find xyz in my data.

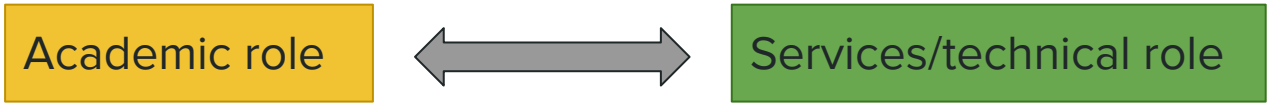
What are your requirements?

Image credits: Isabel Ulusoy

Challenges: RSE acceptance and support for funding

RSE group different from research group!

- Many different roles
- Contributions to research projects on vastly different scales
- Psychological safety vs competitive research
- Expertise should not be lost
- Learning works differently
- Funding model quite different and currently not adequate
- Recognition as academic contribution and importance of RSE

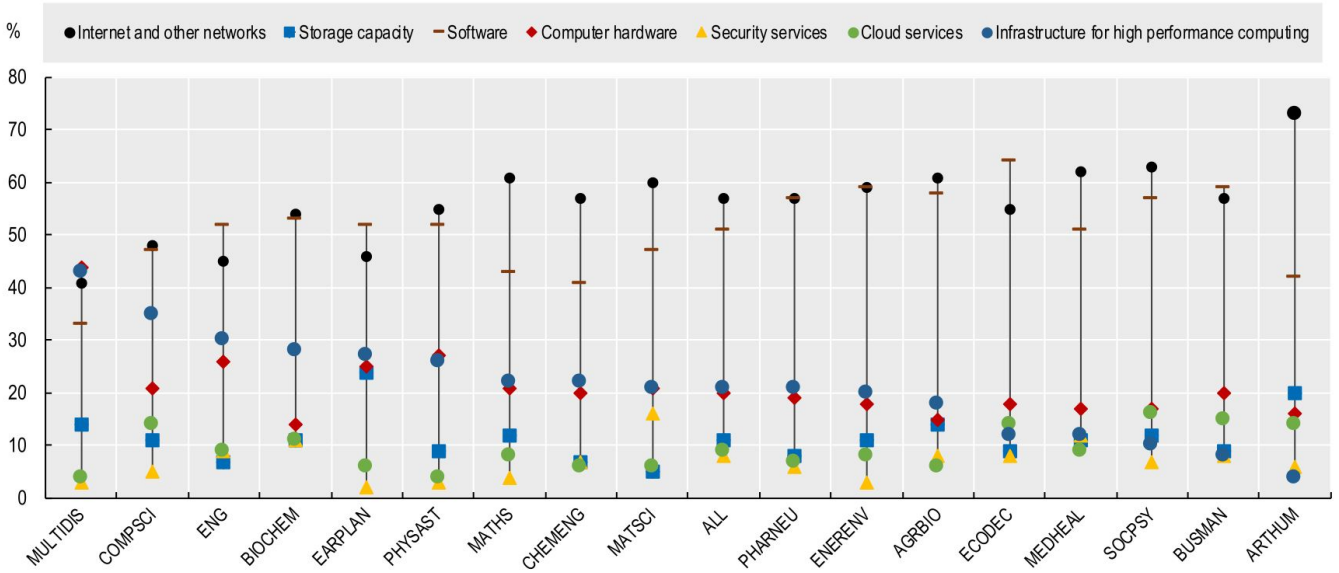


Where does an RSE fit in?

Research software = research infrastructure

Figure 5.5. Most important infrastructure for scientific authors' research work, by field

Percentage of authors deeming a given type of infrastructure as important



Note: Weighted estimates based on sampling weights adjusted for nonresponse. Respondents can select a maximum of two options.
Source: OECD International Survey of Scientific Authors (ISSA), 2018. http://oe.cd/issa.

Time for a survey regarding SMPs?
https://limesurvey.urz.uni-heidelberg.de/index.php/525728?lang=en

Getting in touch

- Our website: ssc.uni-heidelberg.de
- Our email: ssc@iwr.uni-heidelberg.de

Thank you for your attention!