

# What is good scientific practice for research software?

... and how can we make it part of our research culture?

Konrad U. Förstner

@konradfoerstner

University of Würzburg

May 10<sup>th</sup>, 2017

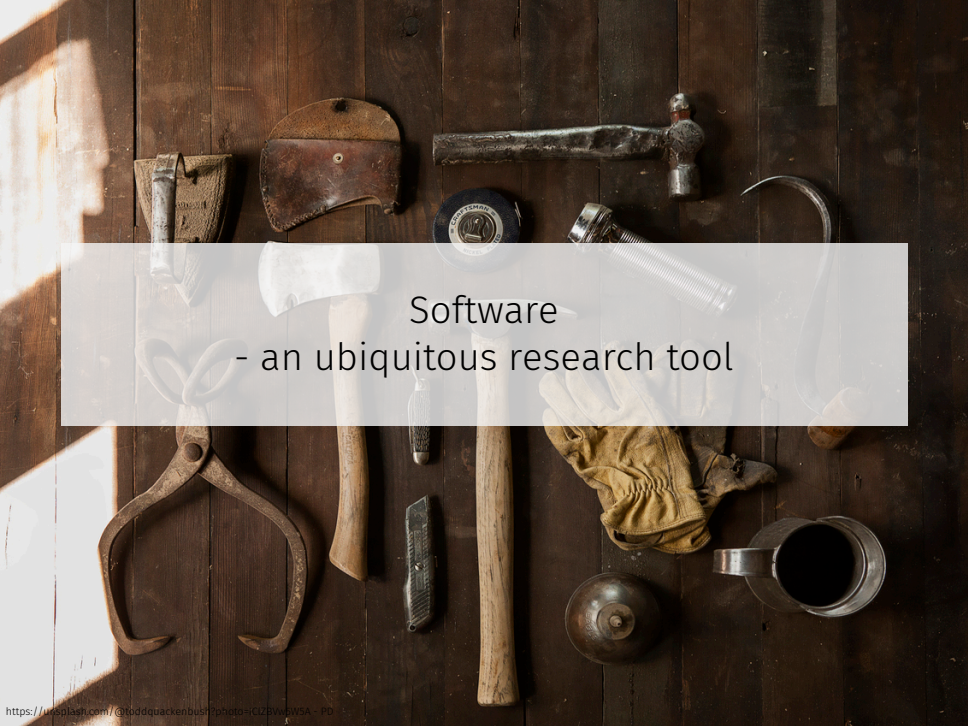
Science  $\Leftrightarrow$  Technology



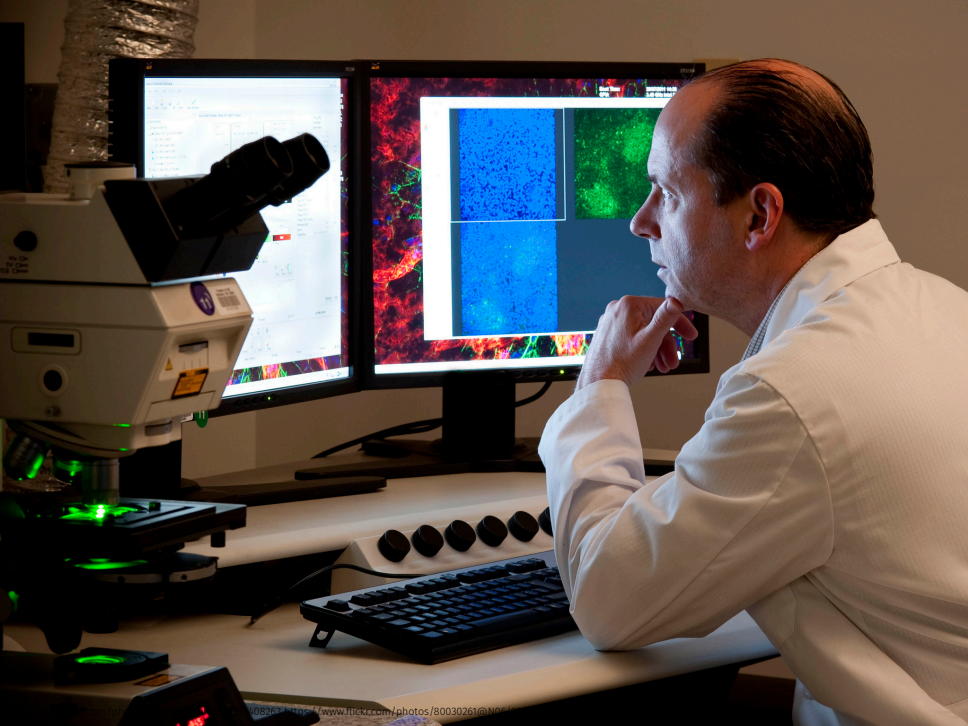






A top-down photograph of various tools and objects arranged on a dark wooden surface. The tools include a hammer, a wrench, a pair of pliers, a pair of gloves, a cup of coffee, a utility knife, a saw, and a pair of shears. The scene is lit from the left, creating a strong shadow on the left side of the frame.

Software  
- an ubiquitous research tool







It is unquestionable that there is a strong and growing dependence of research on software.

Software is also a result of the  
scientific work.

Quality, accessibility, citability, etc.  
have to be ensured.

The importance of software for research is widely ignored.

## [\[HTML\] A Statistical Analysis of Peptide Electron Transfer Dissociation Fragmentation Mass Spectrometry](#)

[RJ Chalkley](#), [KF Medzhradszky](#), [AJ Lynn](#)... - *Analytical ...*, 2010 - [ncbi.nlm.nih.gov](#)

... An **in-house script** was used to extract the ion types that were matched in all these 2149 spectra. The cumulative number of each ion type matched was determined and these values were normalized to determine the relative frequency of observation of different ion types. ...

Cited by 50 Related articles All 8 versions Web of Science: 32 Cite Save

## [SAINT: probabilistic scoring of affinity purification-mass spectrometry data](#)

[H Choj](#), [B Larsen](#), [ZY Lin](#), [A Breitkreutz](#)... - *Nature ...*, 2011 - [nature.com](#)

... dataset. To replicate PP-NSAF 6 , we removed 330 contaminants from the dataset using the vector magnitude approach. After filtering, probabilities were computed using an **in-house script** following the method presented in ref. 6 ...

Cited by 267 Related articles All 15 versions Web of Science: 182 Cite Save

## [High-resolution structure determination by continuous-rotation data collection in MicroED](#)

[BL Nannenga](#), [D Shi](#), [AGW Leslie](#), [T Gonen](#) - *Nature Methods*, 2014 - [nature.com](#)

... MOSFLM. Supplementary Fig. 3: Process flow from raw data to completely processed data set using MOSFLM. The raw data collected on the TEM is first converted to a file compatible with MOSFLM by an **in-house script**. MOSFLM ...

Cited by 44 Related articles All 15 versions Web of Science: 29 Cite Save

## [\[HTML\] Generation of a predicted protein database from EST data and application to iTRAQ analyses in grape \(Vitis vinifera cv. Cabernet Sauvignon\) berries at ...](#)

[J Lückner](#), [M Laszczak](#), [D Smith](#)... - *BMC ...*, 2009 - [bmcbgenomics.biomedcentral.com](#)

... combined into a second tab delimited file. Duplicate entries among exocarp or mesocarp files were identified using an **in-house script** in the R environment with 'Custom ORF ID' as the search string. Then, ratiometric data at each of ...

Cited by 49 Related articles All 21 versions Web of Science: 33 Cite Save More

## [Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic Escherichia coli](#)

[KG Joensen](#), [F Scheutz](#), [O Lund](#)... - *Journal of clinical ...*, 2014 - *Am Soc Microbiol*

... Briefly, 1,647 (in a later update 5,029) complete bacterial genomes were downloaded from NCBI, and each k-mer (k = 16) with the prefix ATGAC was saved in a database using an **in-house script**. ... Another **in-house script** was used to search the database. ...

Cited by 149 Related articles All 12 versions Web of Science: 79 Cite Save

## [\[PDF\] LIUM's systems for the IWSLT 2011 speech translation tasks.](#)

[A Rousseau](#), [F Bougares](#), [P Deléglise](#)... - ..., 2011 - [ai2-s2-pdfs.s3.amazonaws.com](#)

... We call these sets LIUM dev2010 and LIUM tst2010. We then introduced different input types, after the baseline system had been fixed. Moreover, all of our data was processed by a newer version of our **in-house script** first described in [12] and based on previous work by [13]. ...

Cited by 35 Related articles All 10 versions Cite Save More

## **Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates.**

[Eklund A](#)<sup>1</sup>, [Nichols TE](#)<sup>2</sup>, [Knutsson H](#)<sup>3</sup>.

### **Author information**

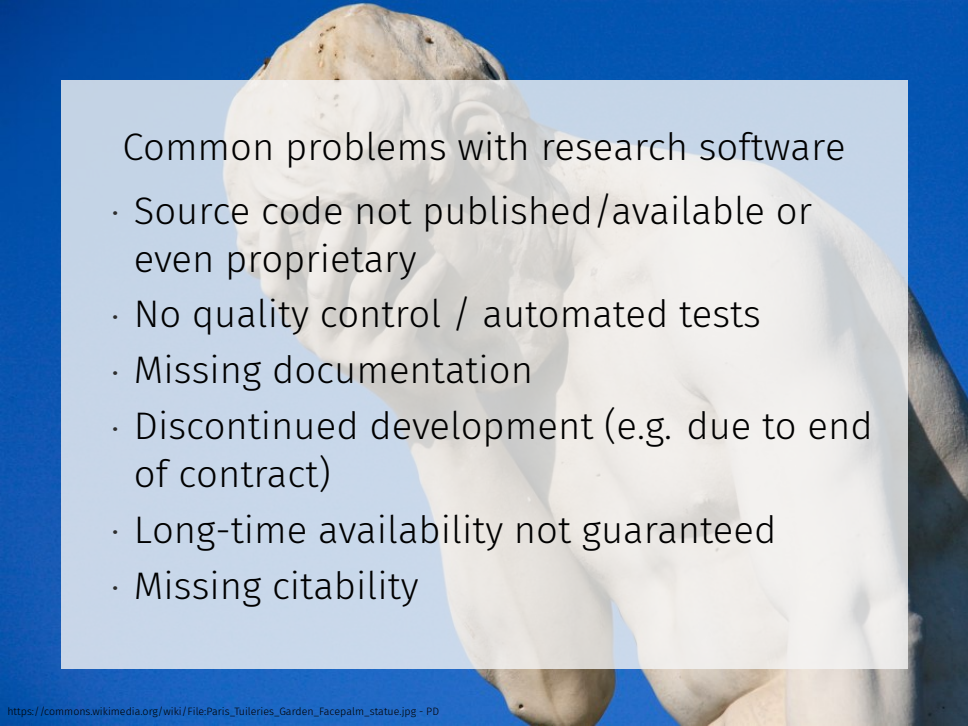
### **Erratum in**

Correction for Eklund et al., Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. [Proc Natl Acad Sci U S A. 2016]

### **Abstract**

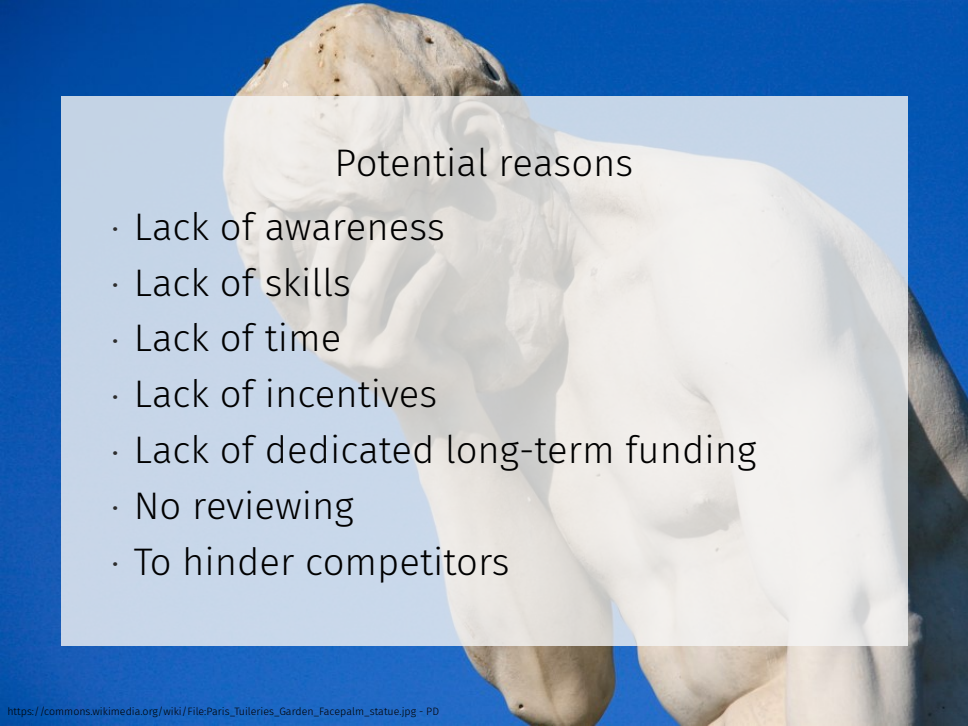
The most widely used task functional magnetic resonance imaging (fMRI) analyses use parametric statistical methods that depend on a variety of assumptions. In this work, we use real resting-state data and a total of 3 million random task group analyses to compute empirical familywise error rates for the fMRI software packages SPM, FSL, and AFNI, as well as a nonparametric permutation method. For a nominal familywise error rate of 5%, the parametric statistical methods are shown to be conservative for voxelwise inference and invalid for clusterwise inference. Our results suggest that the principal cause of the invalid cluster inferences is spatial autocorrelation functions that do not follow the assumed Gaussian shape. By comparison, the nonparametric permutation test is found to produce nominal results for voxelwise as well as clusterwise inference. These findings speak to the need of validating the statistical methods being used in the field of neuroimaging.



A white marble statue of a woman, likely the 'Facepalm' statue in the Tuileries Garden in Paris. She is depicted with her head buried in her hand, a gesture of despair or frustration. The statue is set against a clear blue sky. The image is used as a background for a slide about research software problems.


## Common problems with research software

- Source code not published/available or even proprietary
- No quality control / automated tests
- Missing documentation
- Discontinued development (e.g. due to end of contract)
- Long-time availability not guaranteed
- Missing citability

A white marble statue of a man, likely a classical figure, with his right hand pressed against his forehead in a gesture of distress, contemplation, or despair. The statue is set against a solid blue background. The image is semi-transparent, allowing the text to be overlaid.

## Potential reasons

- Lack of awareness
- Lack of skills
- Lack of time
- Lack of incentives
- Lack of dedicated long-term funding
- No reviewing
- To hinder competitors

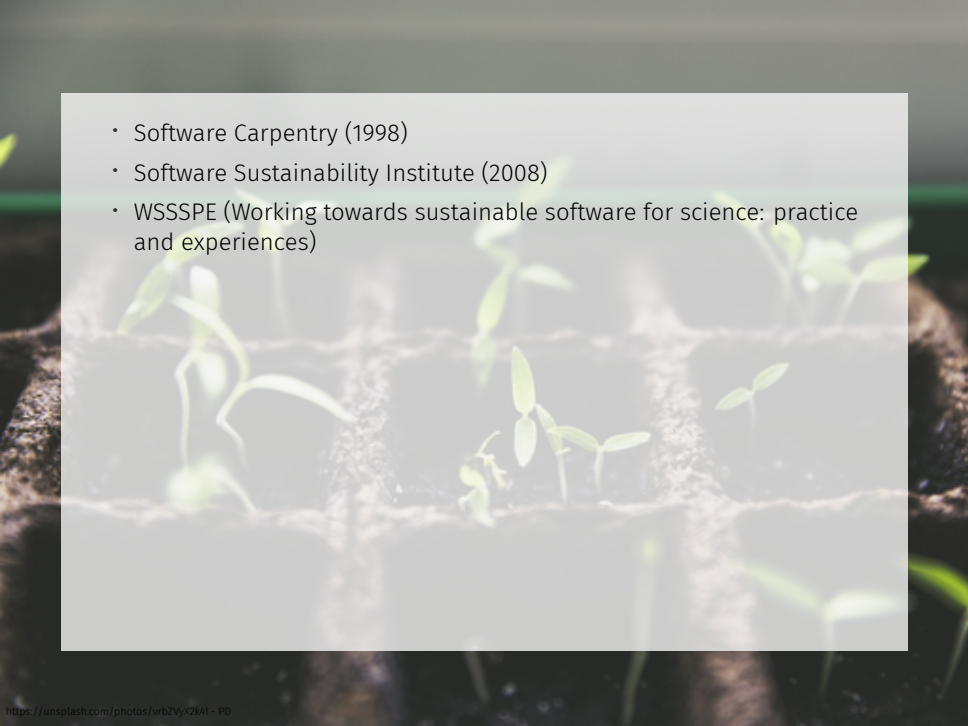
A close-up photograph of a seedling tray with several small green plants growing in individual compartments. The plants are in various stages of growth, with some showing two leaves and others just starting to emerge. The soil is dark and rich. The background is blurred, showing a green horizontal line.

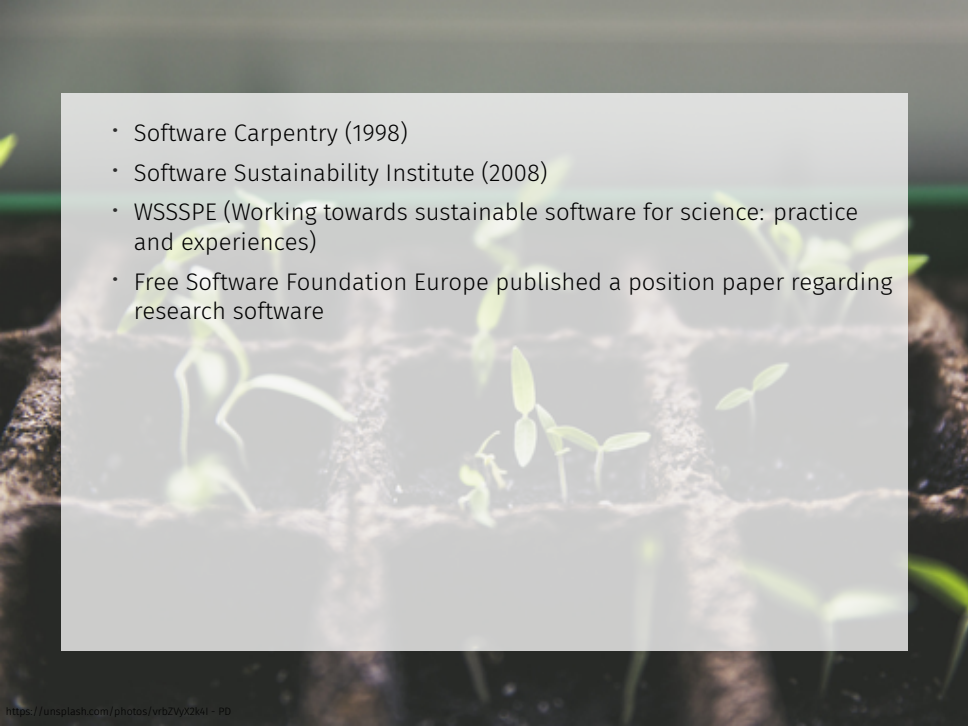
Several initiatives have been launched to address these issues.

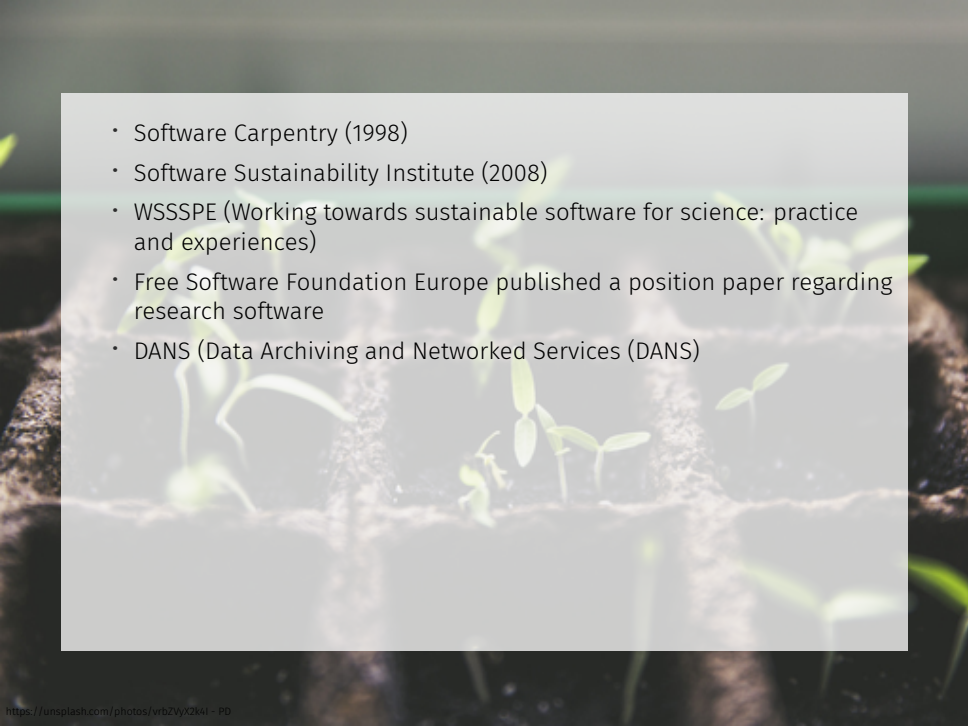
- Software Carpentry (1998)

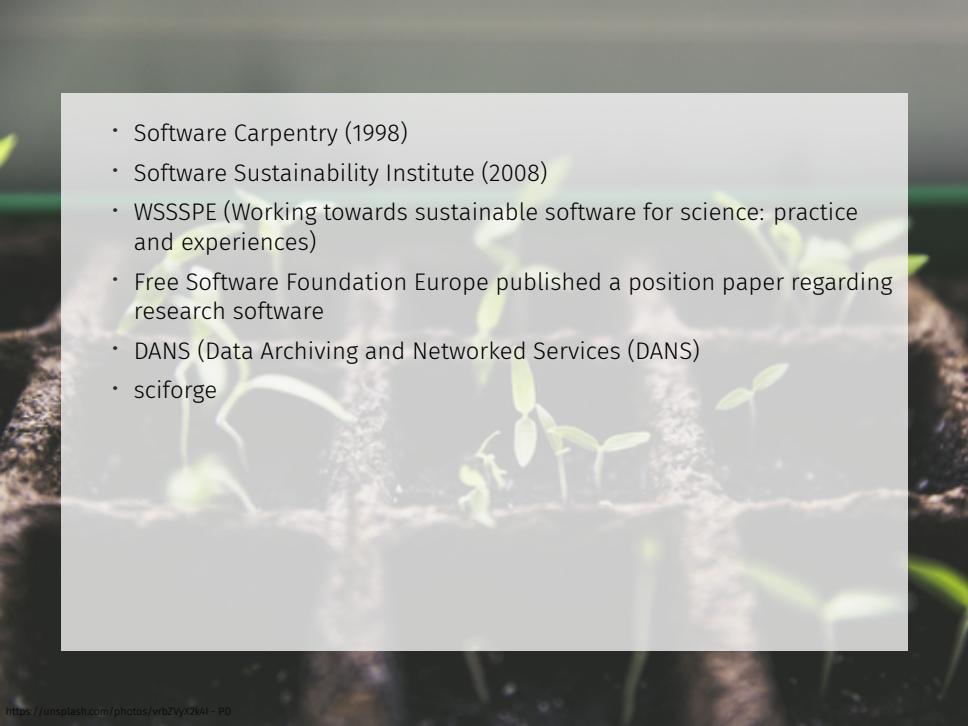
- Software Carpentry (1998)
- Software Sustainability Institute (2008)



- 
- A background image showing several small green seedlings with two leaves each, growing in a dark brown, textured soil tray. The seedlings are arranged in a grid pattern, and the lighting is soft, highlighting their vibrant green color against the dark soil.
- Software Carpentry (1998)
  - Software Sustainability Institute (2008)
  - WSSSPE (Working towards sustainable software for science: practice and experiences)

- 
- Software Carpentry (1998)
  - Software Sustainability Institute (2008)
  - WSSSPE (Working towards sustainable software for science: practice and experiences)
  - Free Software Foundation Europe published a position paper regarding research software

- 
- The background of the slide is a close-up photograph of several small green seedlings with two leaves each, growing out of dark, rich soil. The lighting is soft, highlighting the texture of the soil and the vibrant green of the young plants.
- Software Carpentry (1998)
  - Software Sustainability Institute (2008)
  - WSSSPE (Working towards sustainable software for science: practice and experiences)
  - Free Software Foundation Europe published a position paper regarding research software
  - DANS (Data Archiving and Networked Services (DANS))

- 
- The background of the slide features a close-up, slightly blurred photograph of several small green seedlings with two leaves each, growing out of dark, rich soil. The lighting is soft, highlighting the texture of the soil and the vibrant green of the young plants.
- Software Carpentry (1998)
  - Software Sustainability Institute (2008)
  - WSSSPE (Working towards sustainable software for science: practice and experiences)
  - Free Software Foundation Europe published a position paper regarding research software
  - DANS (Data Archiving and Networked Services (DANS))
  - sciforge

- Software Carpentry (1998)
- Software Sustainability Institute (2008)
- WSSSPE (Working towards sustainable software for science: practice and experiences)
- Free Software Foundation Europe published a position paper regarding research software
- DANS (Data Archiving and Networked Services (DANS))
- sciforge
- DFG program "Research Software Sustainability"  
(7M €, 130 applications)



- Software Carpentry (1998)
- Software Sustainability Institute (2008)
- WSSSPE (Working towards sustainable software for science: practice and experiences)
- Free Software Foundation Europe published a position paper regarding research software
- DANS (Data Archiving and Networked Services (DANS))
- sciforge
- DFG program "Research Software Sustainability" (7M €, 130 applications)
- Helmholtz Association Task Group "Access to and re-use of research software" formed; organized a workshop about research software

- Software Carpentry (1998)
- Software Sustainability Institute (2008)
- WSSSPE (Working towards sustainable software for science: practice and experiences)
- Free Software Foundation Europe published a position paper regarding research software
- DANS (Data Archiving and Networked Services (DANS))
- sciforge
- DFG program "Research Software Sustainability" (7M €, 130 applications)
- Helmholtz Association Task Group "Access to and re-use of research software" formed; organized a workshop about research software
- de-RSE founded (de-rse.org)

- Software Carpentry (1998)
- Software Sustainability Institute (2008)
- WSSSPE (Working towards sustainable software for science: practice and experiences)
- Free Software Foundation Europe published a position paper regarding research software
- DANS (Data Archiving and Networked Services (DANS))
- sciforge
- DFG program "Research Software Sustainability" (7M €, 130 applications)
- Helmholtz Association Task Group "Access to and re-use of research software" formed; organized a workshop about research software
- de-RSE founded (de-rse.org)
- Several more ... lot of them here!

Since 2008: Priority Initiative "Digital Information"  
of the  
Alliance of Science Organisations in Germany



## Priority areas of the initiative

- Research Data
- Virtual Research Environments
- National Licensing
- National Hosting Strategy
- Legal Frameworks
- Open Access



## Priority areas of the initiative

- Research Data
- Virtual Research Environments
- National Licensing
- National Hosting Strategy
- Legal Frameworks
- Open Access

Some fame due to the recent "DEAL" negotiations.

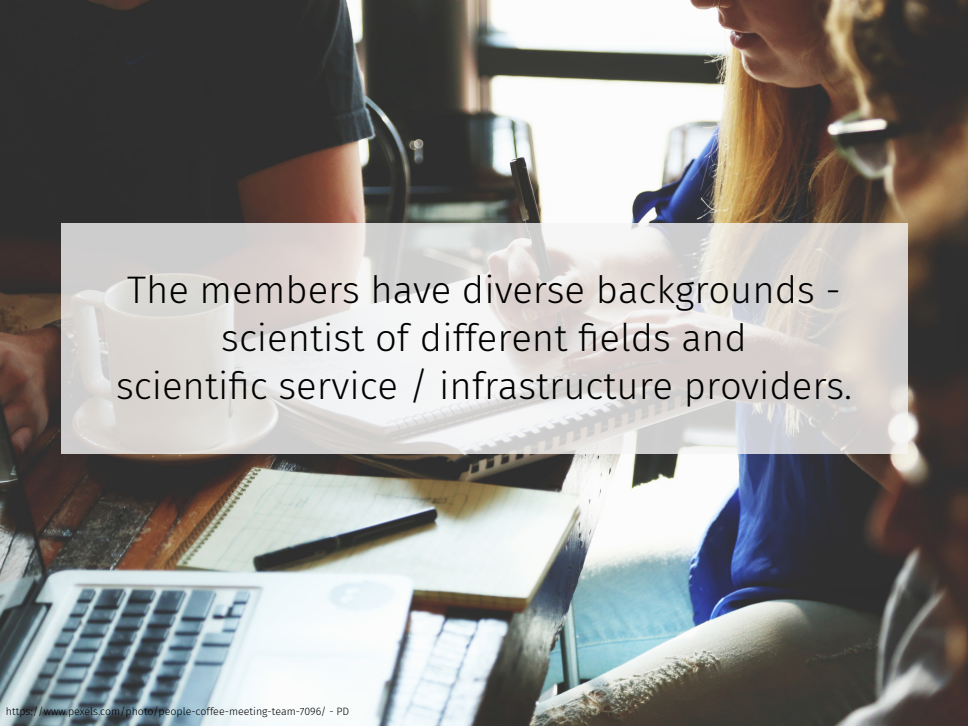


Since 2016

The cross-disciplinary *ad-hoc* working group  
"Research Software"

Mathias Bornschein  
Dr. Matthias Katerbow  
Prof. Dr. Andreas Zeller  
Dr. Bernadette Fritzsch  
Dr. Uwe Konrad  
Dr. Georg Feulner  
Dr. Jürgen Fuhrmann  
Michael Franke  
Stephan Janosch  
Dr. Michael Erben-Russ  
Dennis Zielke  
Prof. Dr. Björn Brembs  
Dr. Konrad Förstner

Ressortforschung des Bundes  
German Research Foundation  
German Research Foundation  
Helmholtz Association  
Helmholtz Association  
Leibniz Association  
Leibniz Association  
Max Planck Society  
Max Planck Society  
Fraunhofer Society  
Fraunhofer Society  
German Rectors' Conference  
German Rectors' Conference



The members have diverse backgrounds -  
scientist of different fields and  
scientific service / infrastructure providers.



*Our modus operandi*

Compile recommendations and carry them  
back into our research organisations.



## Guiding principle

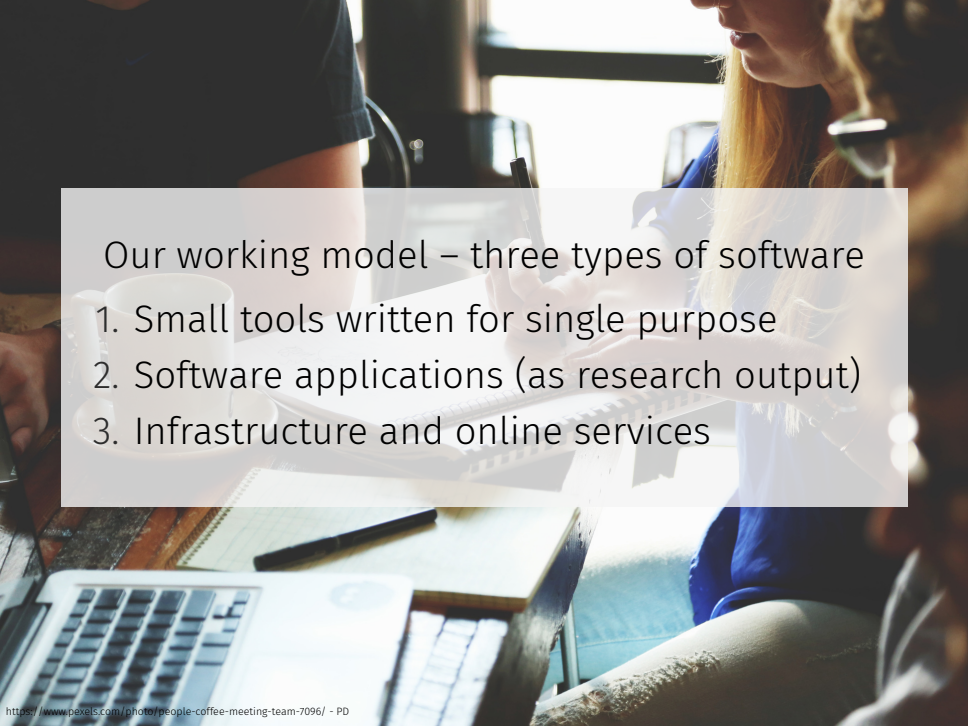
The concept of Good Scientific Practice (GSP) must be also applied to research software.

A close-up photograph of a person with long blonde hair, wearing a blue shirt, sitting at a wooden table. They are holding a black pen and writing in a spiral-bound notebook. On the table, there is a white coffee cup on a saucer, a laptop, and another notebook. The background is slightly blurred, showing a bright window and other people in a meeting or office setting.

But what can Good Scientific Practice mean  
for research software?




- 
- Reproducibility
  - Confirmability
  - Transparency
  - Quality
  - Re-usability



Our working model – three types of software

1. Small tools written for single purpose
2. Software applications (as research output)
3. Infrastructure and online services

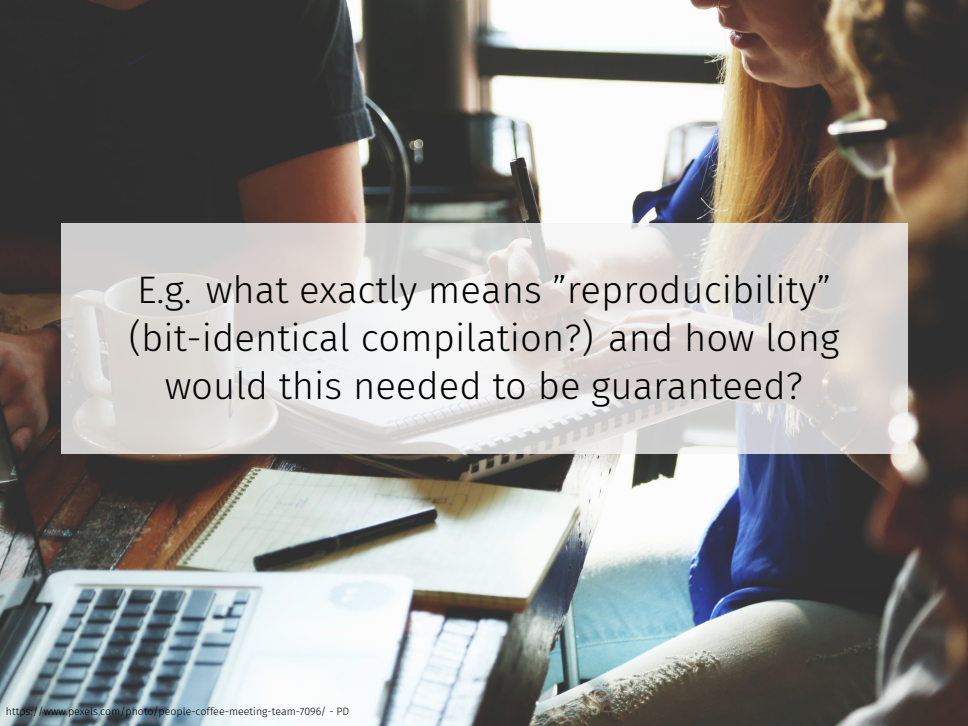


All three levels are relevant and  
have to be addressed.



Exact needs and possibilities might differ  
between scientific communities.

Discourse must also happen  
inside of these communities.

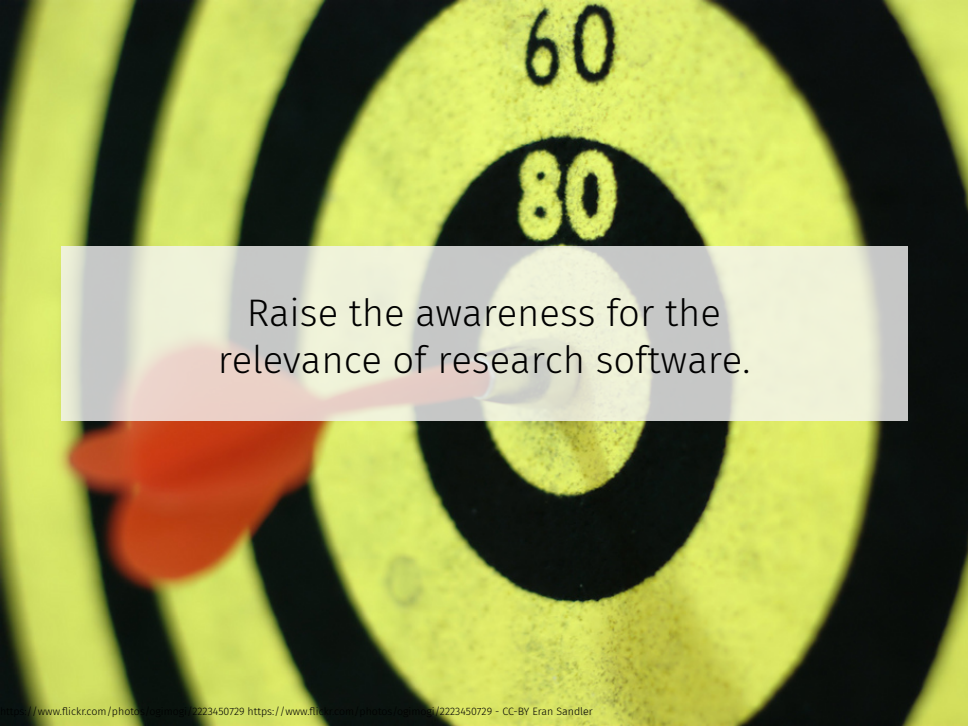


E.g. what exactly means "reproducibility" (bit-identical compilation?) and how long would this needed to be guaranteed?

A close-up photograph of a dartboard. The board is yellow with black concentric circles. The numbers '60' and '80' are printed in black on the yellow background. A red dart with a white stem is hitting the center bullseye. A semi-transparent white rectangular box is overlaid on the center of the image, containing the text 'Our aims / wishes'.

Our aims / wishes




A close-up photograph of a dartboard. The board is yellow with black concentric circles. The numbers '60' and '80' are printed in black on the yellow background. A red dart with a white tip is embedded in the center bullseye. A semi-transparent white rectangular box is overlaid on the center of the image, containing text.

Raise the awareness for the  
relevance of research software.

A close-up photograph of a dartboard. The board is yellow with black concentric circles. The number '60' is printed in black at the top, and '80' is printed in yellow on a black background in the center. A red dart is shown in mid-air, having just hit the white bullseye in the center. The background is blurred, showing more of the dartboard's rings.

Include standards for research software into  
the common Good Scientific Practice  
recommendations.

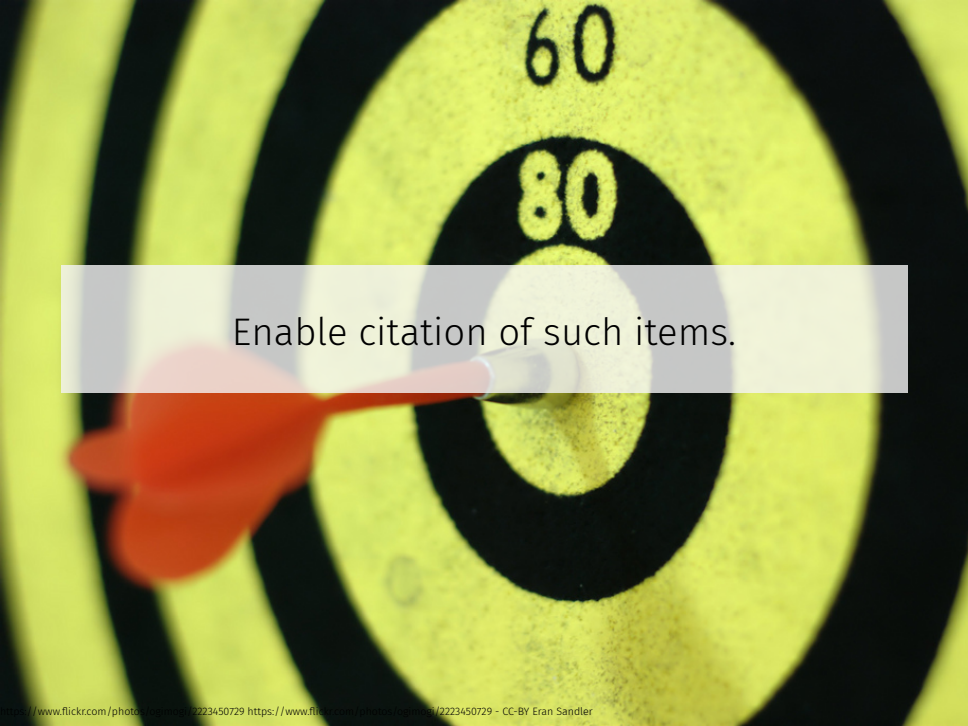


A close-up photograph of a dartboard. The board is yellow with black concentric rings. The numbers '60' and '80' are printed in black on the yellow background. A red dart is shown in the lower-left quadrant, with its white tip hitting the center bullseye. A semi-transparent white rectangular box is overlaid on the center of the image, containing text.

Introduce standards and mechanisms for  
quality control of research software.

A close-up photograph of a dartboard. The board is yellow with black concentric rings. The numbers '60' and '80' are printed in black on the yellow background. A red dart is shown in the lower-left quadrant, with its tip pointing towards the center bullseye. A semi-transparent white rectangular box is overlaid on the center of the image, containing text.

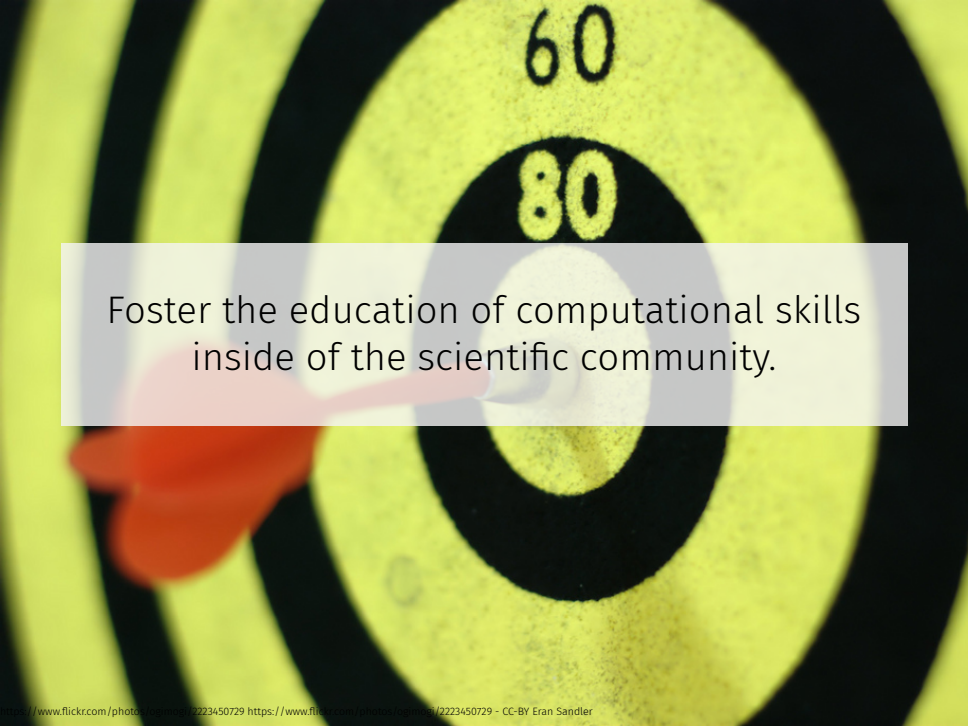
Create institutional platforms to publish and archive software/source code/workflows.

A close-up photograph of a dartboard. The board is yellow with black concentric rings. The numbers '60' and '80' are printed in black on the yellow background. A red dart with a white stem is embedded in the center bullseye. A semi-transparent white rectangular box is overlaid on the center of the image, containing the text 'Enable citation of such items.'


Enable citation of such items.

A close-up photograph of a dartboard. The board is yellow with black concentric rings. The numbers '60' and '80' are printed in black on the yellow background. A red dart with a white tip is embedded in the center bullseye. A semi-transparent white rectangular box is overlaid on the center of the image, containing text.

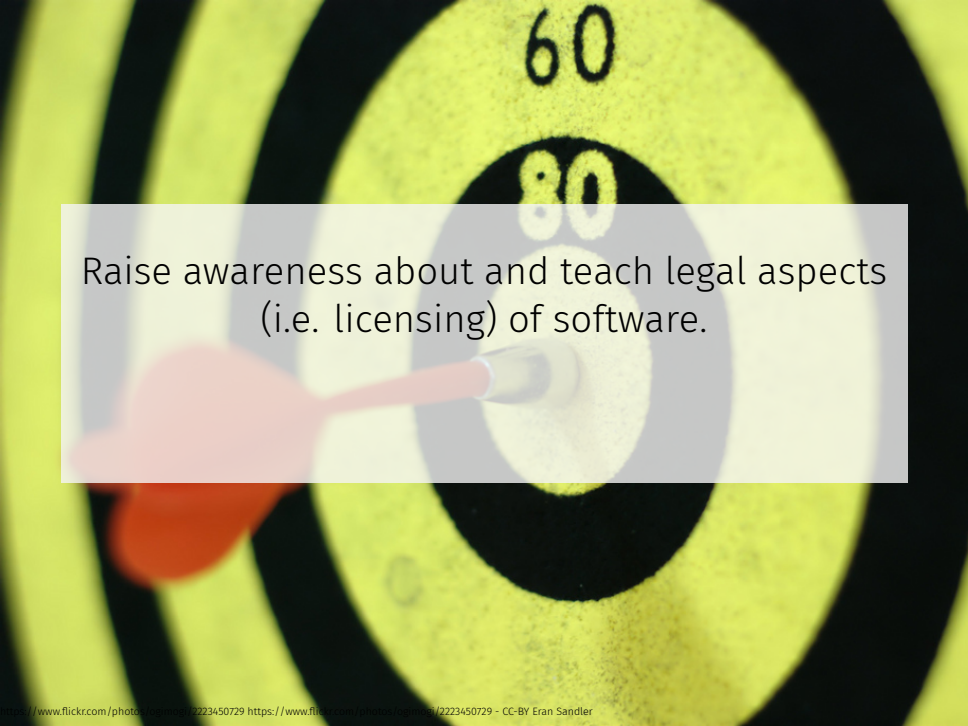
Make these citations part of the scientific reputation system.

A close-up photograph of a dartboard. The board is yellow with black concentric rings. The numbers '60' and '80' are printed in black on the yellow background. A red dart with a white tip is embedded in the center bullseye. A semi-transparent white rectangular box is overlaid on the center of the image, containing text.

Foster the education of computational skills  
inside of the scientific community.

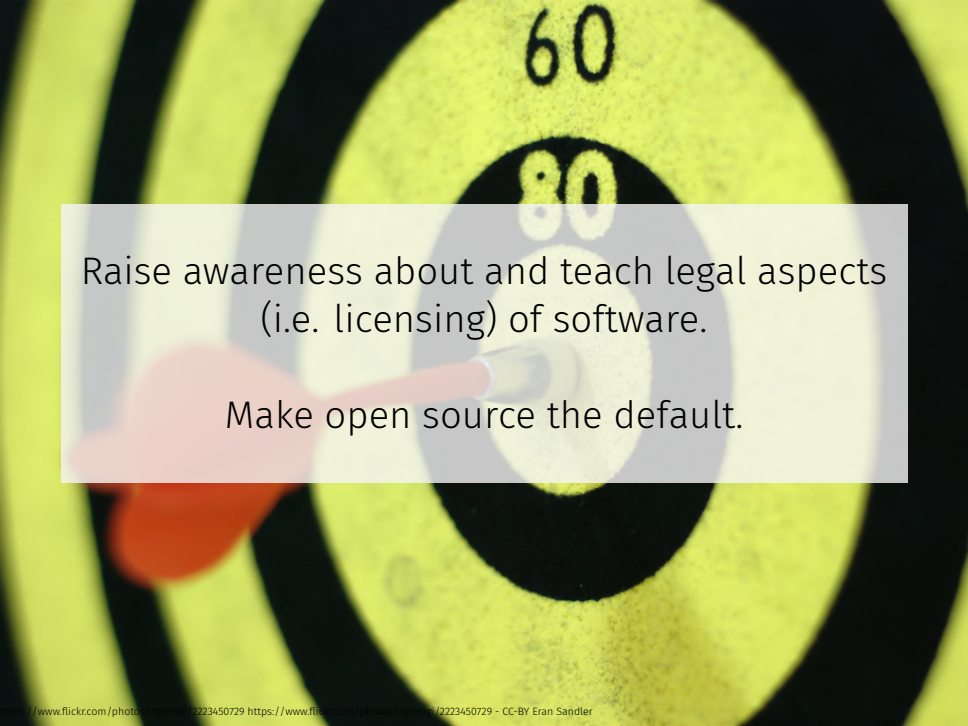
A close-up photograph of a dartboard. The board is yellow with black concentric circles. The numbers '60' and '80' are printed in black on the yellow background. A red dart is shown in the lower-left quadrant, with its tip pointing towards the center bullseye. A semi-transparent white rectangular box is overlaid on the center of the image, containing text.

Develop new career paths like  
Research Software Engineers, Software  
Librarians, Data Scientists.



Raise awareness about and teach legal aspects  
(i.e. licensing) of software.

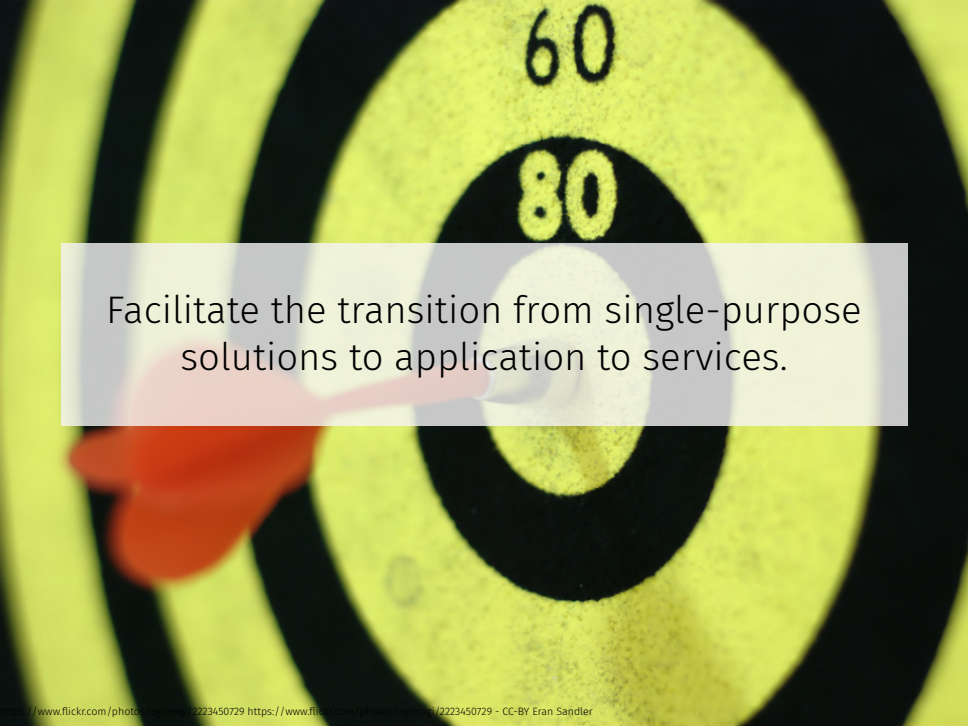


A hand holding a red dart pointing at a target with numbers 60 and 80. The target is a circular board with a yellow outer ring and a black inner ring. The number 60 is written in black at the top, and the number 80 is written in white on the black ring. A hand is holding a red dart, pointing it towards the center of the target. The background is a blurred yellow and black pattern.


Raise awareness about and teach legal aspects  
(i.e. licensing) of software.

Make open source the default.

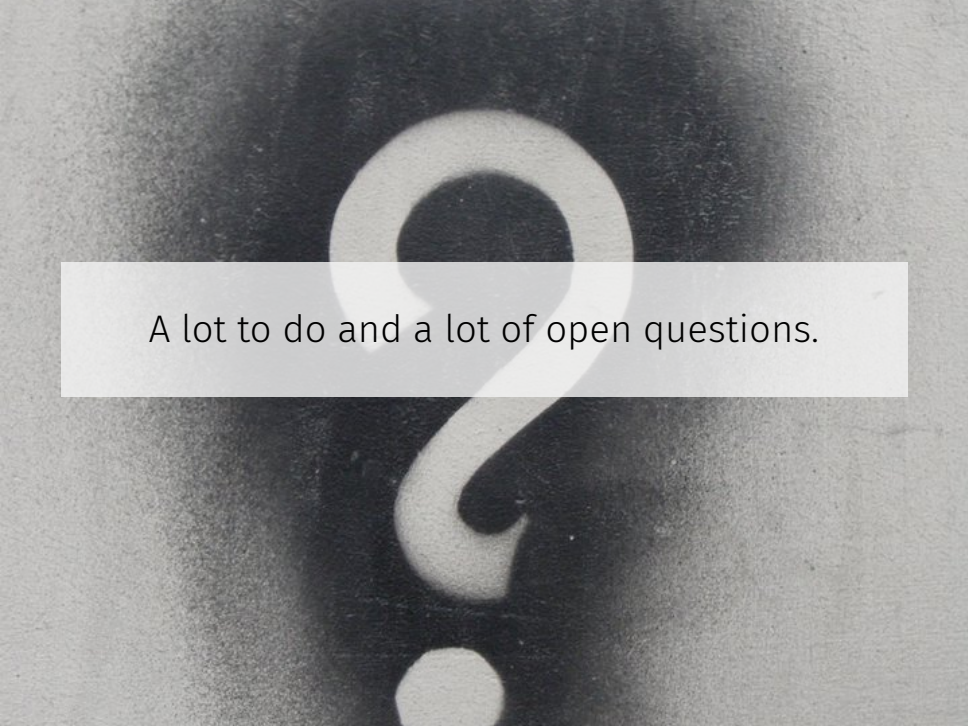


A close-up photograph of a dartboard. The board is yellow with black concentric rings. The numbers '60' and '80' are printed in black on the yellow background. A red dart is shown in the lower-left quadrant, with its white tip hitting the center bullseye. A semi-transparent white rectangular box is overlaid on the center of the image, containing text.

Facilitate the transition from single-purpose solutions to application to services.


A close-up photograph of a dartboard. The board is yellow with black concentric circles. The numbers '60' and '80' are printed in black on the yellow background. A red dart with a white tip is embedded in the center bullseye. A semi-transparent white rectangular box is overlaid on the center of the image, containing text.

Provide long-term funding to enable sustainable software development.

A large, light-colored question mark is centered on a dark, textured background. A semi-transparent white rectangular box is overlaid horizontally across the middle of the question mark, containing the text "A lot to do and a lot of open questions." in a black, sans-serif font.

A lot to do and a lot of open questions.


We represent the German  
scientific community / research organisation.



We represent the German  
scientific community / research organisation.

Ideally all these issues are addressed on an  
international level.



A classical painting depicting five muscular men in a circular, acrobatic pose. The men are rendered with detailed musculature and are holding hands or feet to form a circle. The background is dark. A semi-transparent white box is overlaid in the center of the image, containing the text "Let's do this together!".

Let's do this together!



[www.allianzinitiative.de](http://www.allianzinitiative.de)

@konradfoerstner