

TOWARDS OSGEO BEST PRACTICES FOR SCIENTIFIC SOFTWARE CITATION

INTEGRATION OPTIONS FOR PERSISTENT IDENTIFIERS IN OSGEO PROJECT REPOSITORIES



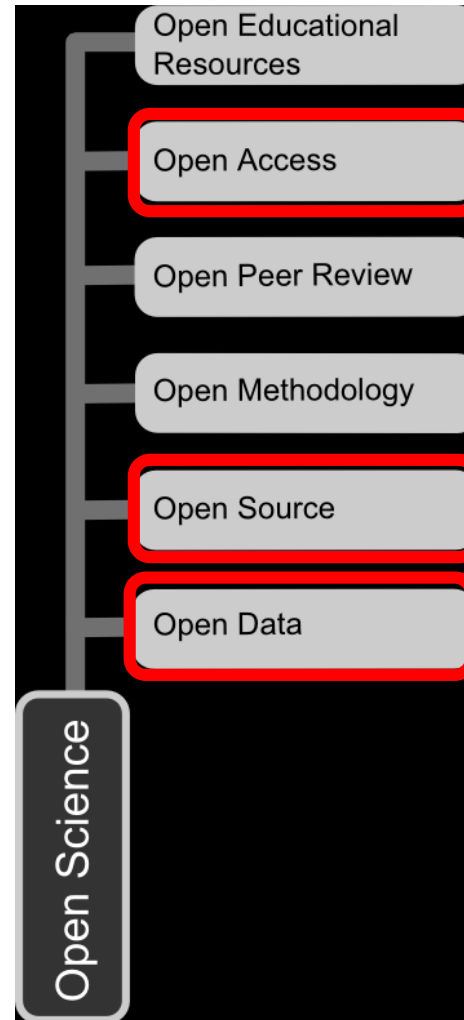
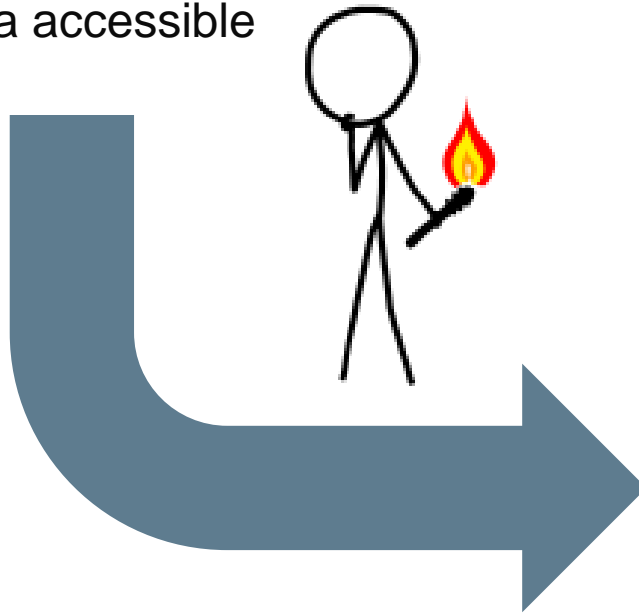
Peter Löwe, Markus Neteler, Jan Goebel, Marco Tullney
Boston, August 17 2017



Does Prometheus receive due credit for his creativity ?

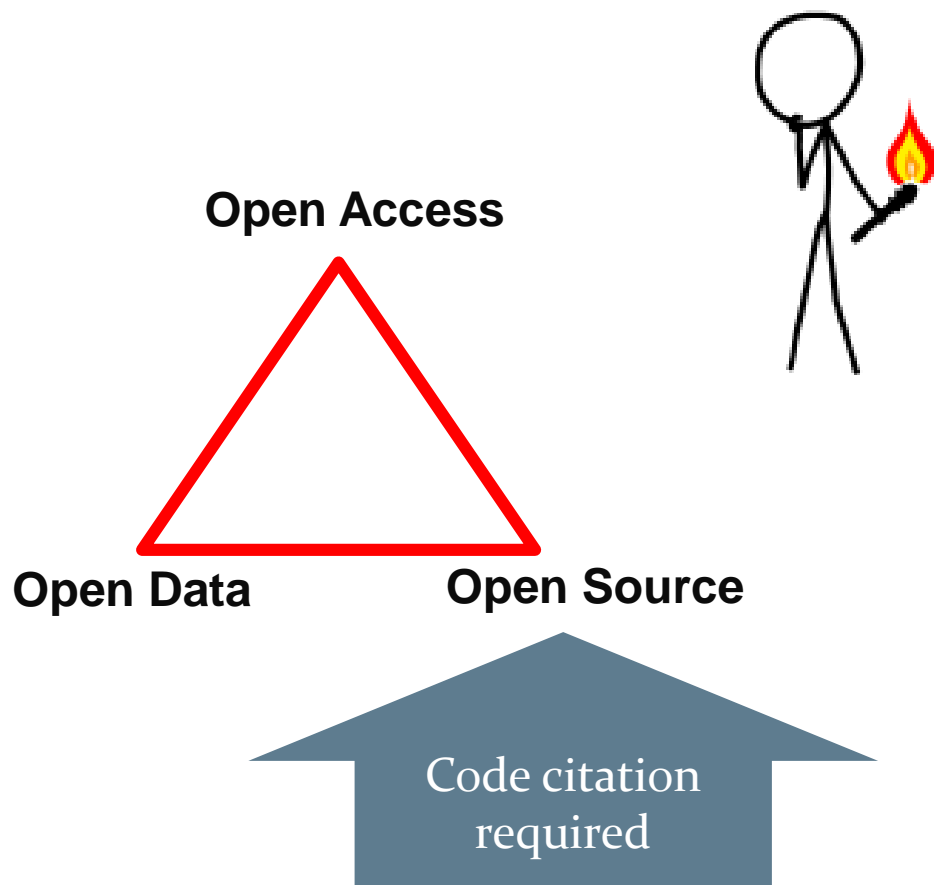
<https://xkcd.com/1228/>

Open Science is the movement to make scientific research and data accessible to all



https://en.wikipedia.org/wiki/Open_science#/media/File:Open_Science_-_Prinzipien.png

Open Science Triangle: Science-related benefits



Society:

- **Greater availability** and accessibility of **publicly funded scientific research outputs**
- **Greater reproducibility** and transparency of scientific works

Community:

- Possibility for **rigorous peer-review**

Individual:

- **Greater impact** of scientific research

Code citation: Requires standards and infrastructure

Motivation for Code Citation

Understanding research fields: code as important part of the record of research and progress in science (*no „throwaway code“*)

Credit: Researchers on all levels(including students!) deserve credit in their coin of the realm (aka citation), especially when this work enables further research by others.

Discoverability: Citation enables finding and reuse

Reproducibility: Citation of a specific software is required, but also information about underlying software stack and configurations are needed



Business

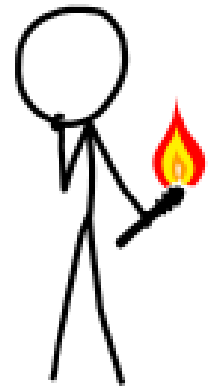
Research,
Education,

Memory
Organisations

**OSGeo
infrastructure
& best practices**

Data Centers, Code
Repositorories,
Libraries

Code citation ?



Software Citation Best Practices

according to FORCE 11



General best practices for code citation already exist!

- **Importance** *Software matters in Science*
- **Credit and Attribution** *Get due credit for your work*
- **Unique Identification** *Unique, persistent, interoperable*
- **Persistence** *Identifier & metadata never expire*
- **Accessibility** *Code & documentation, interop.*
- **Specificity** *Reference to specific code versions*

<https://www.force11.org/>

In comparison to the actual magnitude of research code being produced, only a fraction is being communicated by journals.

As a result, advances in scientific software are not being properly communicated and therefore remain inaccessible to other scientists.

Reality Check: OSGeo Journal

OSGeo Journal

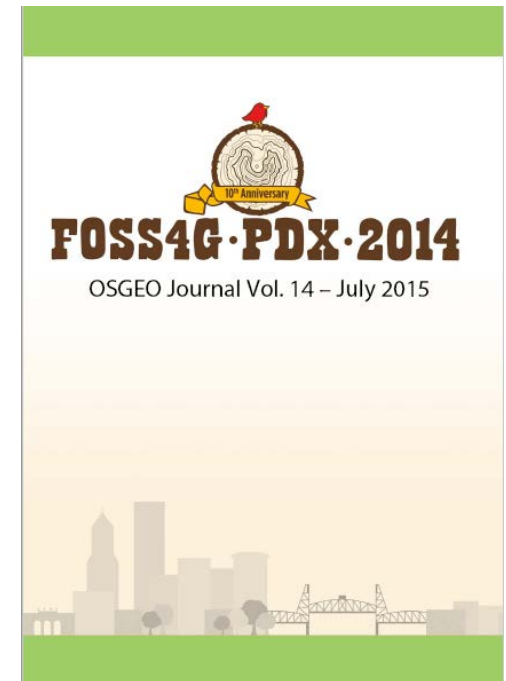
The Journal of the Open Source Geospatial Foundation

<http://www.osgeo.org/journal>

OSGeo Journal

The OSGeo Journal is a digital publication containing case studies, news, tutorials, project updates and more. With a general aim at promoting, highlighting and educating readers about open source geospatial applications in general, but also provides updates on OSGeo projects. The Journal is also available in French.

- Founded in 2007
- Online Journal
- ISSN
- Publishes FOSS4G proceedings
- ***No defined standards for software citation (yet)***



Journal of Open Source Software

-a Role Model ?



The Journal of Open Source Software

The Journal of Open Source Software

A developer friendly journal for research software packages.

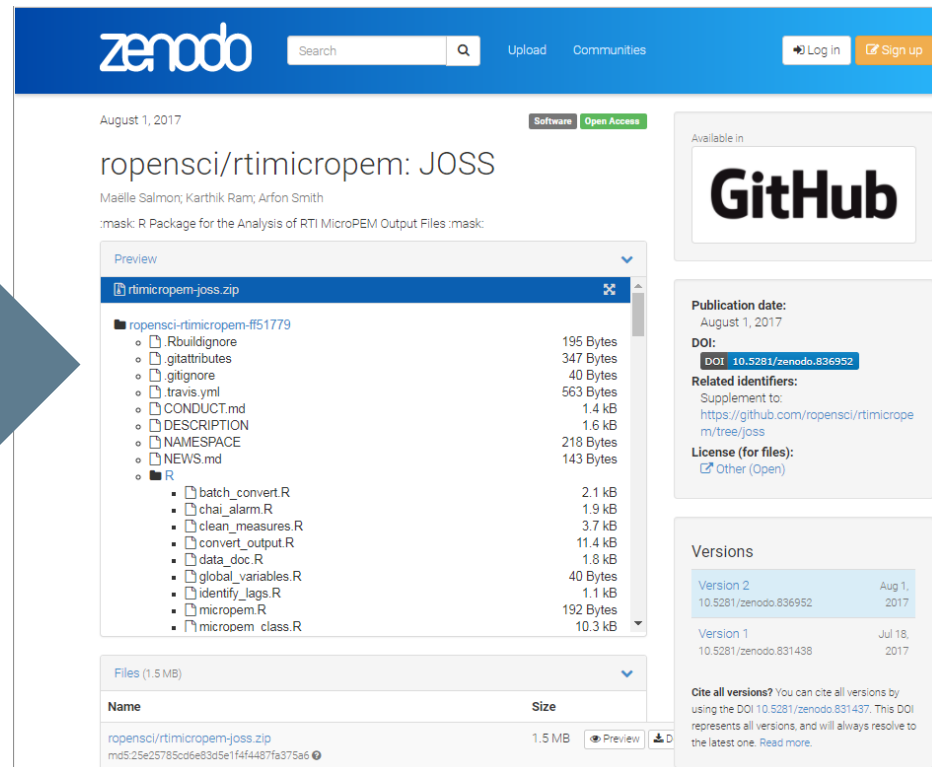
 ropensci / rtimicropem JOSS 10.21105/joss.00333

Supports the input and reproducible analysis of RTI MicroPEM output files.

[10.5281/zenodo](https://zenodo.org/doi/10.5281/zenodo.836952)



„DOI-Link“ points to code within github repository.
Metadata stored in Zenodo.



zenodo Search Upload Communities Log in Sign up

August 1, 2017 Software Open Access

ropensci/rtimicropem: JOSS

Maëlle Salmon; Karthik Ram; Arfon Smith

:mask: R Package for the Analysis of RTI MicroPEM Output Files :mask:

Preview

- rtimicropem-joss.zip
- ropensci-rtimicropem-#51779
 - Rbuildignore 195 Bytes
 - gitattributes 347 Bytes
 - gitignore 40 Bytes
 - travis.yml 563 Bytes
 - CONDUCT.md 1.4 kB
 - DESCRIPTION 1.6 kB
 - NAMESPACE 218 Bytes
 - NEWS.md 143 Bytes
 - R
 - batch_convert.R 2.1 kB
 - chai_alarm.R 1.9 kB
 - clean_measures.R 3.7 kB
 - convert_output.R 11.4 kB
 - data_doc.R 1.8 kB
 - global_variables.R 40 Bytes
 - identify_lags.R 1.1 kB
 - micropem.R 192 Bytes
 - micropem.class.R 10.3 kB

Files (1.5 MB)

| Name | Size |
|--------------------------------------|--------|
| ropensci/rtimicropem-joss.zip | 1.5 MB |
| md5:25e25785c06e83d5e1f4f4487fa375a6 | |

Available in **GitHub**

Publication date: August 1, 2017

DOI: [DOI 10.5281/zenodo.836952](https://doi.org/10.5281/zenodo.836952)

Related identifiers: Supplement to: <https://github.com/ropensci/rtimicropem/tree/joss>

License (for files): [Other \(Open\)](#)

Versions

| Version | Date |
|-----------|--------------|
| Version 2 | Aug 1, 2017 |
| Version 1 | Jul 18, 2017 |

Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.831437. This DOI represents all versions, and will always resolve to the latest one. [Read more.](#)

Long term perspective:

- Data and code will move within the WWW,
- URL links to webpages will expire over time .

Digital Object Identifiers (DOIs) as a way to ensure stable links, preventing:

Link rot

Very bad

From Wikipedia, the free encyclopedia

For link rot in Wikipedia, see [Wikipedia:Link rot](#).

Link rot (or linkrot) is the process by which [hyperlinks](#) on individual [websites](#) or the [Internet](#) in general point to [web pages](#), [servers](#) or other resources that have become permanently unavailable. The phrase also describes the effects of failing to update out-of-date [web pages](#) that clutter [search engine](#) results.





- DOI System ISO Standard 26324 (2012)
- International DOI Foundation (1998).
- Based on the Handle system.
- **Long-term persistence and accessibility of information.**
- **Global infrastructure provider for research data and code**
DataCite (non-profit, software infrastructure is FOSS):



<https://www.datacite.org/>



DOI: Acronym for "digital object identifier".

A DOI identifies the object itself and not the place where it is located.

What you see: **alphanumeric string** (never changes)

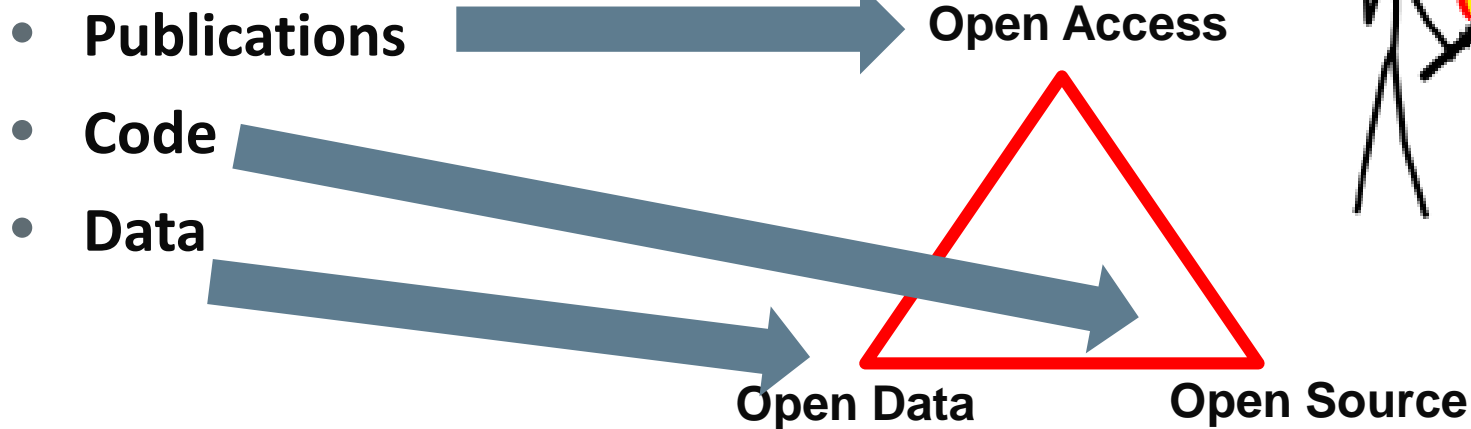
Associated with: **location** (such as **URL**)

Accompanied with: who, what, when... (**metadata**)

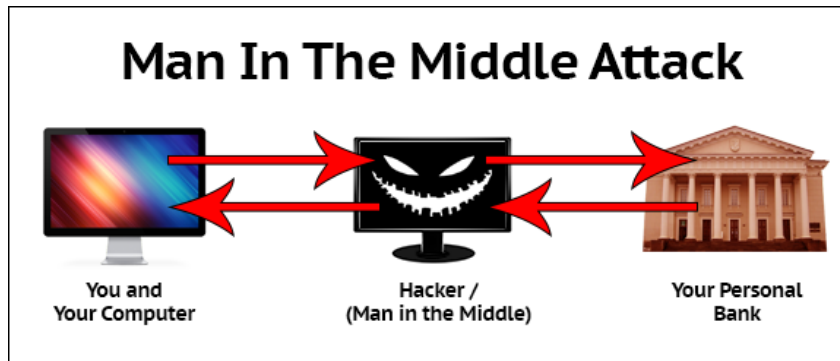


WHAT TO USE DOIs FOR ?

DOIs can be used to reference



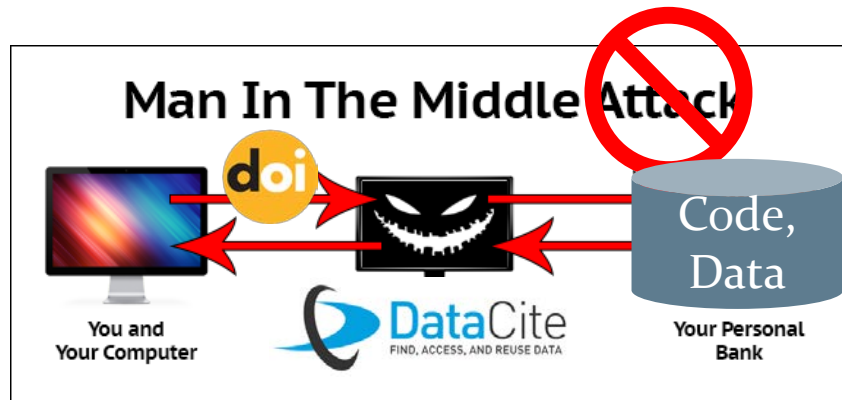
DOI magic explained: Man in the middle – can be friendly...



- **DOIs** are resolved by a **resolving entity** („man in the middle“).
- The resolving entity does not host the data itself.
- It receives **updates from the hosting data repository** whenever the data changes location (**new URL**).
- A DOI will then always resolve to a valid landing page.

<https://image.slidesharecdn.com/doi-100203060339-phpapp01/95/doi-in-he-11-728.jpg?cb=1265177093>
<https://www.deepdotweb.com/wp-content/uploads/2016/10/word-image-19.png>

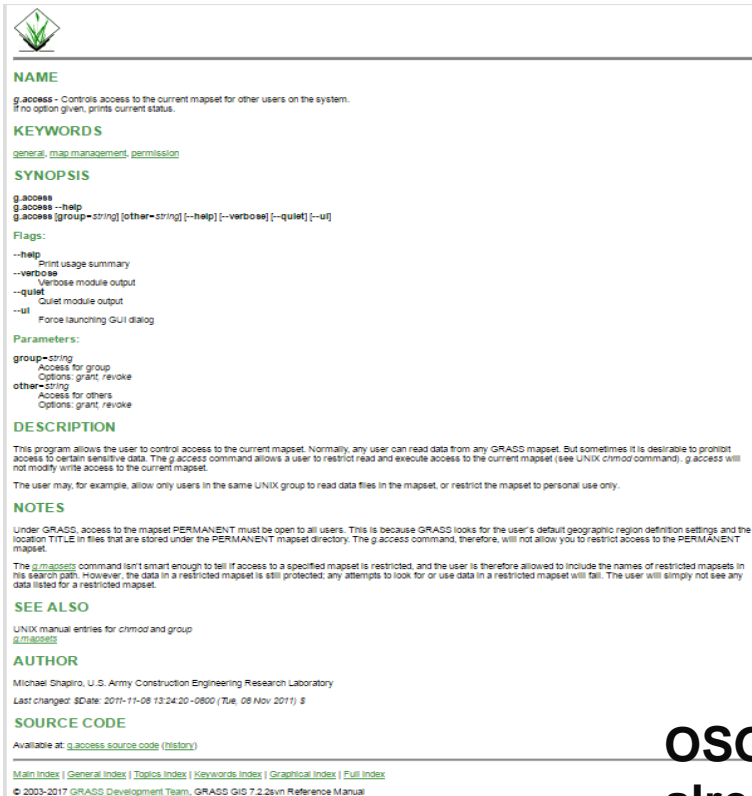
DOI magic explained: Man in the middle – can be friendly...



- **DOIs** are resolved by a **resolving entity** („man in the middle“).
- The resolving entity does not host the data itself.
- It receives **updates from the hosting data repository** whenever the data changes location (**new URL**).
- A DOI will then always resolve to a valid landing page.

<https://image.slidesharecdn.com/doi-100203060339-phpapp01/95/doi-in-he-11-728.jpg?cb=1265177093>
<https://www.deepdotweb.com/wp-content/uploads/2016/10/word-image-19.png>

Requirement: DOIs resolve to landing pages



```
NAME
g.access - Controls access to the current mapset for other users on the system.
If no option given, prints current status.

KEYWORDS
general, map management, permission

SYNOPSIS
g.access
g.access --help
g.access [group=string] [other=string] [--help] [--verbose] [--quiet] [--ui]

Flags:
--help          Print usage summary
--verbose       Verbose module output
--quiet         Quiet module output
--ui            Force launching GUI dialog

Parameters:
group=string    Access for group
                Options: grant, revoke
other=string    Access for others
                Options: grant, revoke

DESCRIPTION
This program allows the user to control access to the current mapset. Normally, any user can read data from any GRASS mapset. But sometimes it is desirable to prohibit access to certain sensitive data. The g.access command allows a user to restrict read and execute access to the current mapset (see UNIX chmod command). g.access will not modify write access to the current mapset.

The user may, for example, allow only users in the same UNIX group to read data files in the mapset, or restrict the mapset to personal use only.

NOTES
Under GRASS, access to the mapset PERMANENT must be open to all users. This is because GRASS looks for the user's default geographic region definition settings and the location TITLE in files that are stored under the PERMANENT mapset directory. The g.access command, therefore, will not allow you to restrict access to the PERMANENT mapset.

The g.mapsets command isn't smart enough to tell if access to a specified mapset is restricted, and the user is therefore allowed to include the names of restricted mapsets in his search path. However, the data in a restricted mapset is still protected; any attempts to look for or use data in a restricted mapset will fail. The user will simply not see any data listed for a restricted mapset.

SEE ALSO
UNIX manual entries for chmod and group
g.mapsets

AUTHOR
Michael Shapiro, U.S. Army Construction Engineering Research Laboratory
Last changed: $Date: 2011-11-08 13:24:20 -0800 (Tue, 08 Nov 2011) $

SOURCE CODE
Available at: g.access source code \(history\)

Main Index | General Index | Topics Index | Keywords Index | Graphical Index | Full Index
© 2003-2017 GRASS Development Team, GRASS GIS 7.2.2svn Reference Manual
```

- Every DOI resolves to a **landing page**.
- Landing pages provide **metadata and further content**
- DOIs are designed to **outlive their content**,

OSGeo content like GRASS module manual pages already qualify as landing pages for DOIs

A digital object with a DOI has to be:

Stable[°] (i.e. not going to be modified)

Complete (i.e. not going to be updated)

Permanent – by assigning a DOI we're committing to make the dataset available for posterity

Good quality – by assigning a DOI its receiving the data centre's stamp of approval, saying that it's complete and all the metadata is available



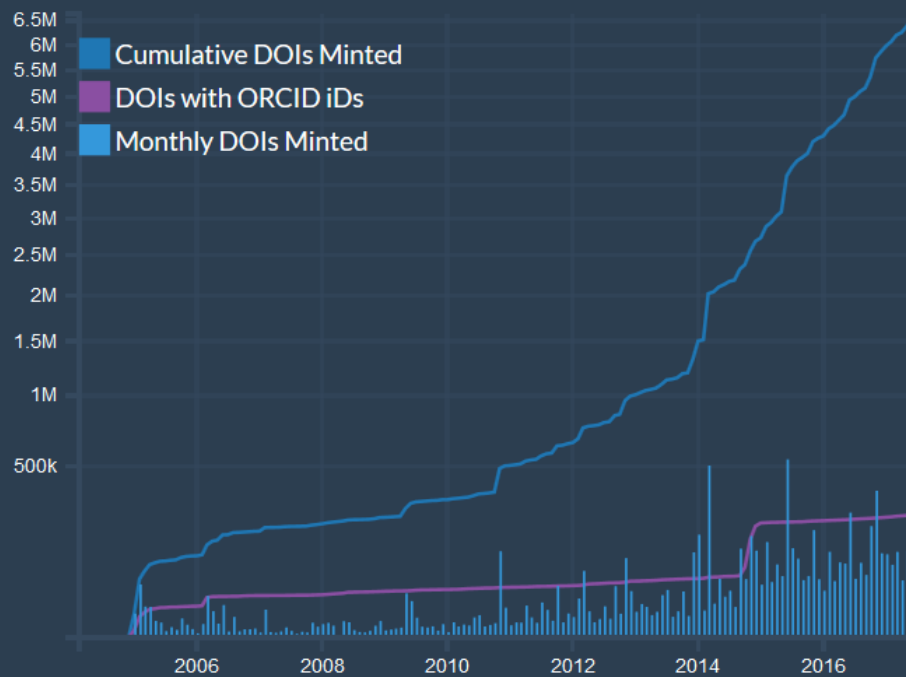
([°] DOI can handle software-versioning)

DOIs are on the rise

RESEARCH OBJECT IDENTIFIER METRICS (DATACITE)

[VIEW RAW DATA](#)

overview



www.datacite.org

Example: DOI for Journal Articles



Environmental Modelling &
Software

Volume 31, May 2012, Pages 124-130



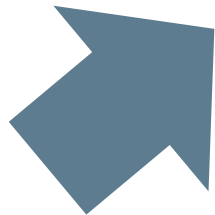
GRASS GIS: A multi-purpose open source GIS

Markus Neteler ^a  , M. Hamish Bowman ^b , Martin Landa ^c , Markus Metz ^a 

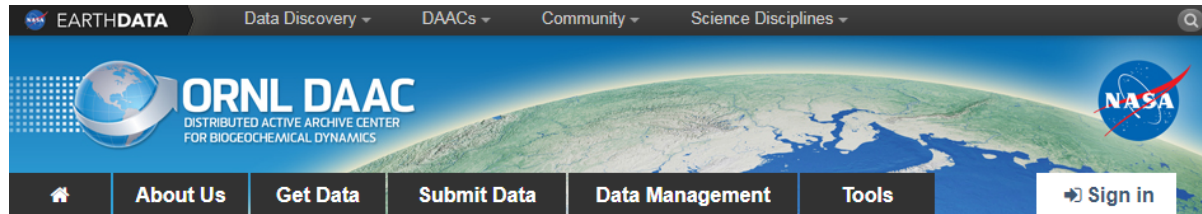
 [Show more](#)

<https://doi.org/10.1016/j.envsoft.2011.11.014>

[Get rights and content](#)



Example: DOIs for Data



Search ORNL DAAC

Search

DAAC Home

SNF Satellite Image Data Inventory

Download Data 154.3KB

Overview

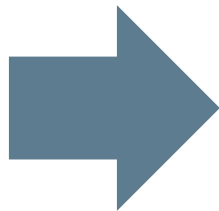
| | |
|--------------|---|
| Data set | SNF Satellite Image Data Inventory |
| DOI | https://doi.org/10.3334/ORNLDAAC/186 |
| Release date | 1996-10-24 |
| Project | Superior National Forest |
| Time period | 1986-01-01 to 1986-12-31 |

Usage Metrics

| | Count | Earliest Date | Latest Date | Data Usage |
|-----------|-------|---------------|-------------|---------------------------|
| Downloads | 31 | 2008-01-17 | 2017-05-01 | 92 total files downloaded |



Bounding box. Lat: 48.17N to 47.66N, Long: 92.51W to 91.77W



Example: DOIs for Code

The screenshot shows the Zenodo interface for a repository. At the top, the Zenodo logo is on the left, and search, upload, and community links are in the center. On the right, there are 'Log in' and 'Sign up' buttons. Below the header, the repository name 'wwolff7/SEBAL_GRASS' is highlighted with a red box. The author 'Wagner Wolff' is listed below. A description reads: 'Script to calculate daily evapotranspiration for Landsat8 images in GRASS-GIS.' To the right, a 'GitHub' logo indicates the code is available there. The 'Publication date' is November 17, 2016. The 'DOI' is 10.5281/zenodo.167350. The 'Related identifiers' section includes a GitHub link: https://github.com/wwolff7/SEBAL_GRASS/tree/v1.0.0. The 'License (for files)' is 'Other (Open)'. A 'Share' section contains icons for GitHub, a menu, Twitter, Facebook, Print, and a plus sign. The 'Cite as' section provides the citation: 'Wagner Wolff. (2016, November 17). wwolff7/SEBAL_GRASS. Zenodo. <http://doi.org/10.5281/zenodo.167350>'. A text input field for citation styles is also present.

November 17, 2016

Software Open Access

wwolff7/SEBAL_GRASS

Wagner Wolff

Script to calculate daily evapotranspiration for Landsat8 images in GRASS-GIS.

Available in

GitHub

Publication date: November 17, 2016

DOI: DOI 10.5281/zenodo.167350

Related identifiers: Supplement to: https://github.com/wwolff7/SEBAL_GRASS/tree/v1.0.0

License (for files): Other (Open)

Share

Cite as

Wagner Wolff. (2016, November 17). wwolff7/SEBAL_GRASS. Zenodo. <http://doi.org/10.5281/zenodo.167350>

Start typing a citation style...

Preview

Download

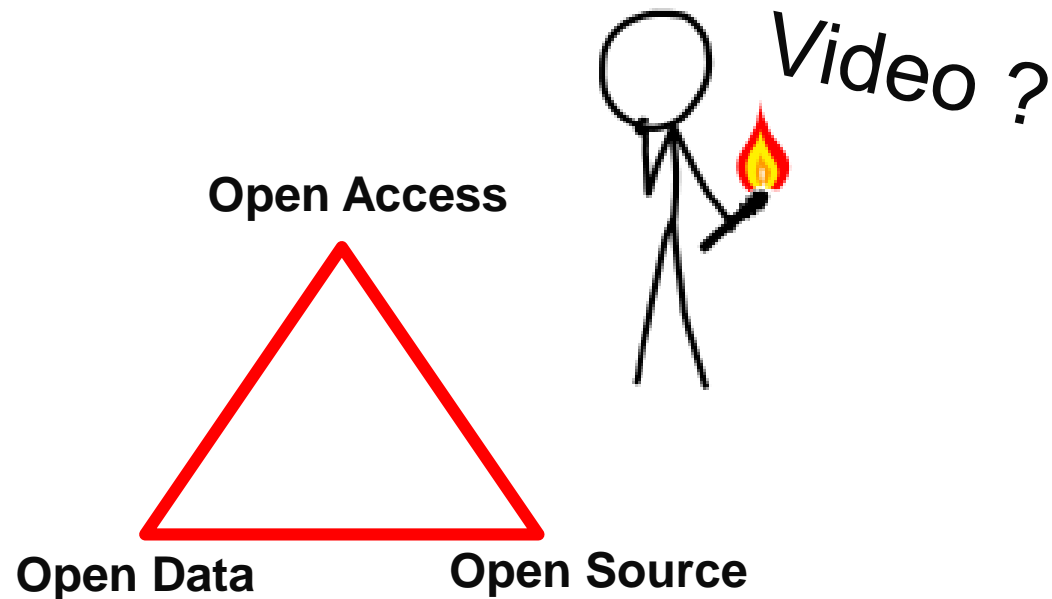
| Name | Size |
|--------------------------------------|--------|
| wwolff7/SEBAL_GRASS-v1.0.0.zip | 4.1 kB |
| md5:84a6a5057facb96be8a763d306508c8d | |

GRASS code, hosted in a personal repository.
Discoverable by the OSGeo communities ?

DOIs being currently used by OSGeo: Video

Scientific-technical video is part of the research cycle

- FOSS4G presentations deserve scientific credit by citation and long term preservation in a repository



OSGeo Videos with DOIs

- **OSGeo conference recordings** are hosted by FOSS4G media partner German National Library for Science and Technology (TIB) .
- The annual growth exceeds 100 hours of new content
- **OSGeo videos are part of the record of science**



<http://dx.doi.org/10.5446/14749#t=39:10,39:33>

DOI

Timestamp

Scientific
citation



- GRASS GIS,
- Over 3 decades experience (since 1982)
- OSGeo project
- Over 350 modules
- Additional add-on modules
- Main repository: SVN

<https://grass.osgeo.org/>

GRASS Code Citation

The GRASS GIS project wiki provides advice, how to cite versions of GRASS GIS in scientific. No coverage of DOIs (yet).



GRASS Citation Repository

Contents [\[hide\]](#)

- 1 [How to cite GRASS GIS](#)
 - 1.1 [GRASS GIS Manual citations](#)
 - 1.2 [Peer reviewed general GRASS GIS articles](#)
 - 1.3 [Books](#)
 - 1.4 [Topic oriented citations](#)
 - 1.5 [Historic citations](#)
- 2 [Helpful tools and GRASS related bibliographic searches](#)

```
@Manual{GRASS_GIS_software,  
  title = {Geographic Resources Analysis Support System (GRASS GIS)  
Software, Version 7.2},  
  author = {{GRASS Development Team}},  
  organization = {Open Source Geospatial Foundation},  
  year = {2017},  
  url = {http://grass.osgeo.org},  
}
```

GRASS Code Levels



GRASS GIS

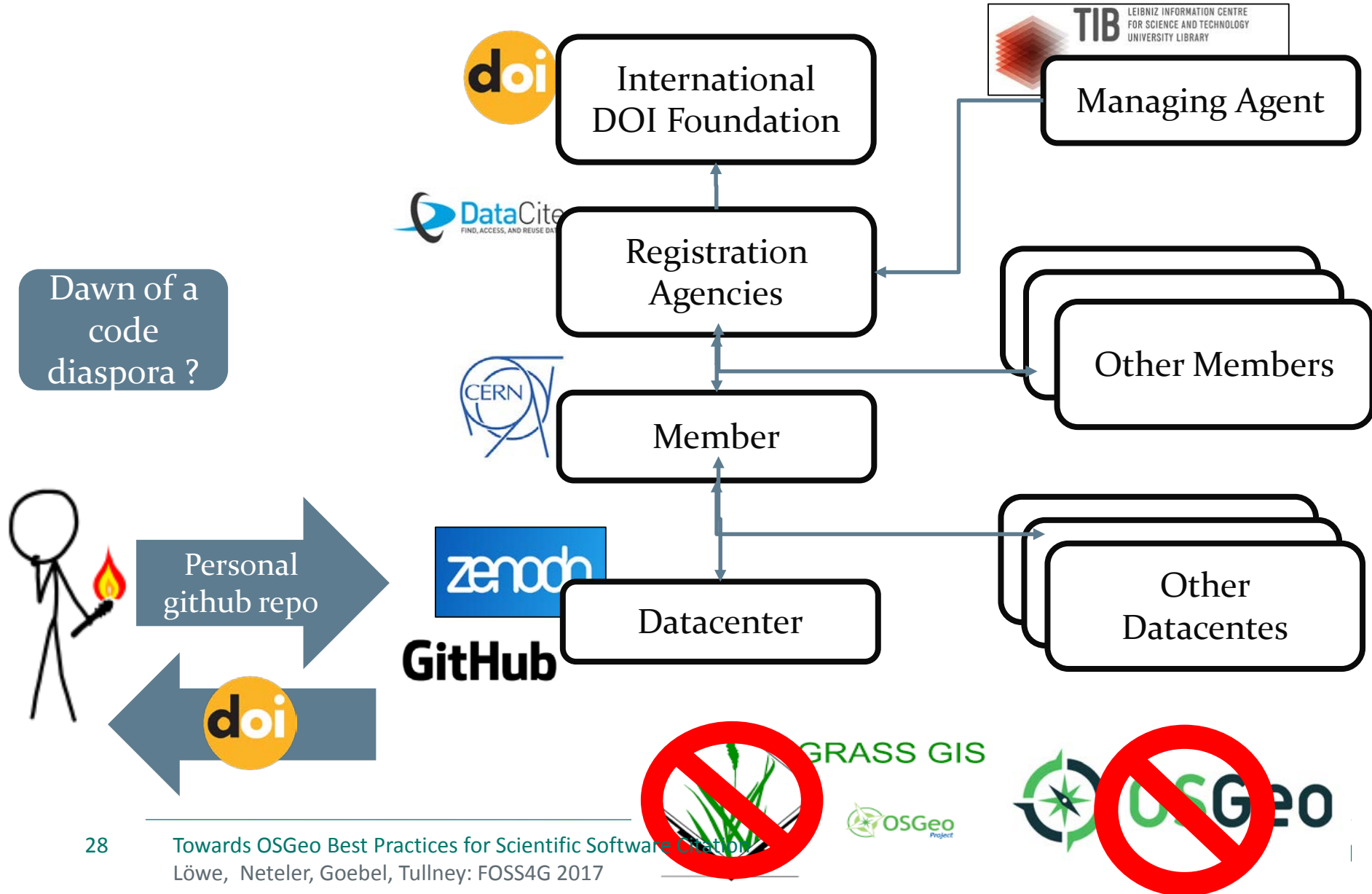


„**External code**“, based on GRASS repo, not shared with community, not hosted in OSGeo GRASS repository. ***Potentially volatile.***

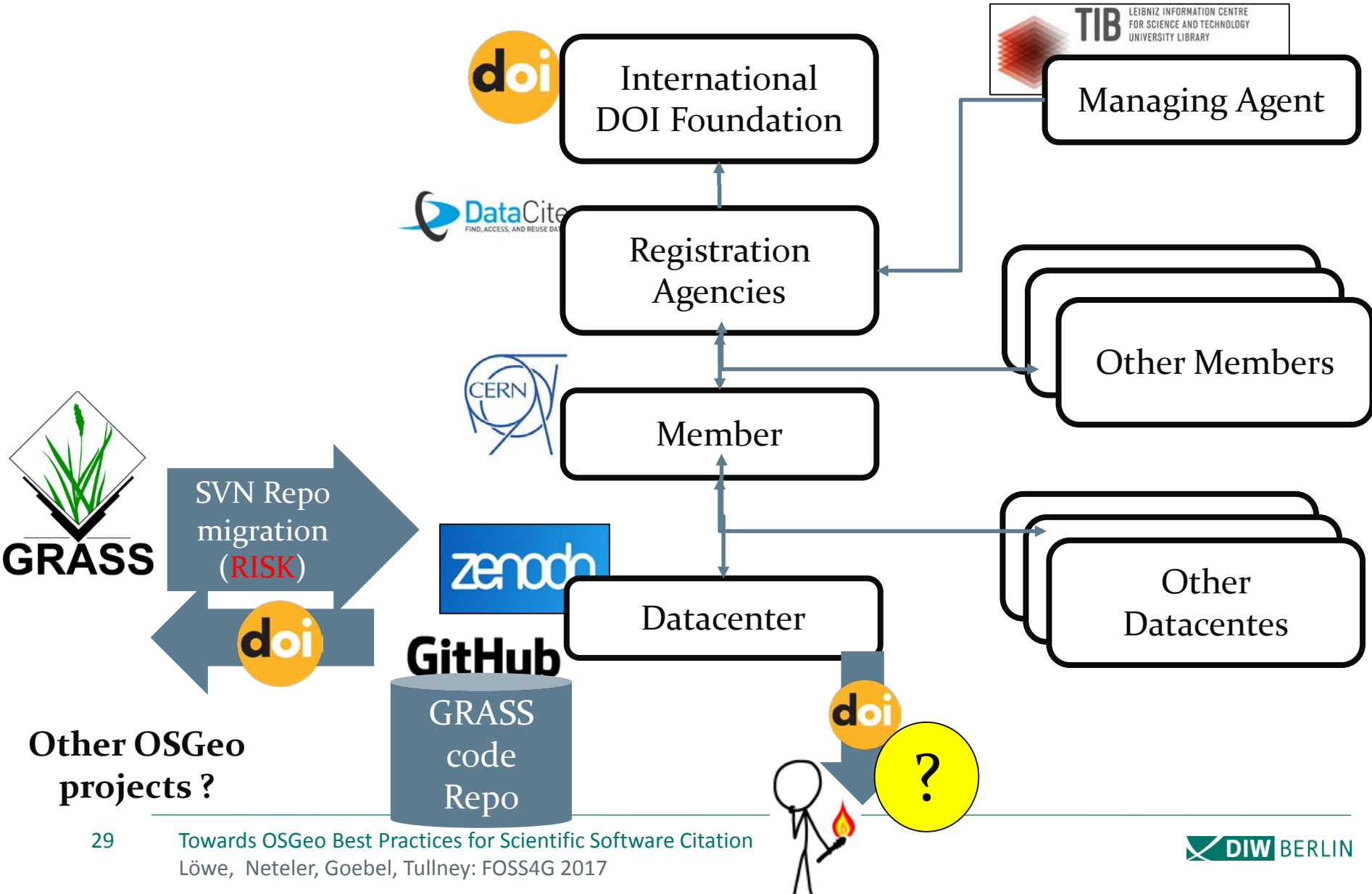
- 2. Add-on modules:** Hosted and preserved in OSGeo GRASS repository, minimal quality standards, including standardized landing page (GRASS module manual page), **limited peer review, discoverable by GRASS search functions**
- 3. Core modules:** Hosted and preserved in GRASS repository, manual page with links to previous code versions, demo data, reference to developers, **rigorous peer review by GRASS community, discoverable by GRASS search functions**

Individual Level: Zenodo Option

external code / add-ons



Community Level: Zenodo Option for GRASS Repository



Reality check: Zenodo (and figshare) are all-purpose Repositories: One size fits all ?

All
purpose.
Good ?



<http://figshare.org>

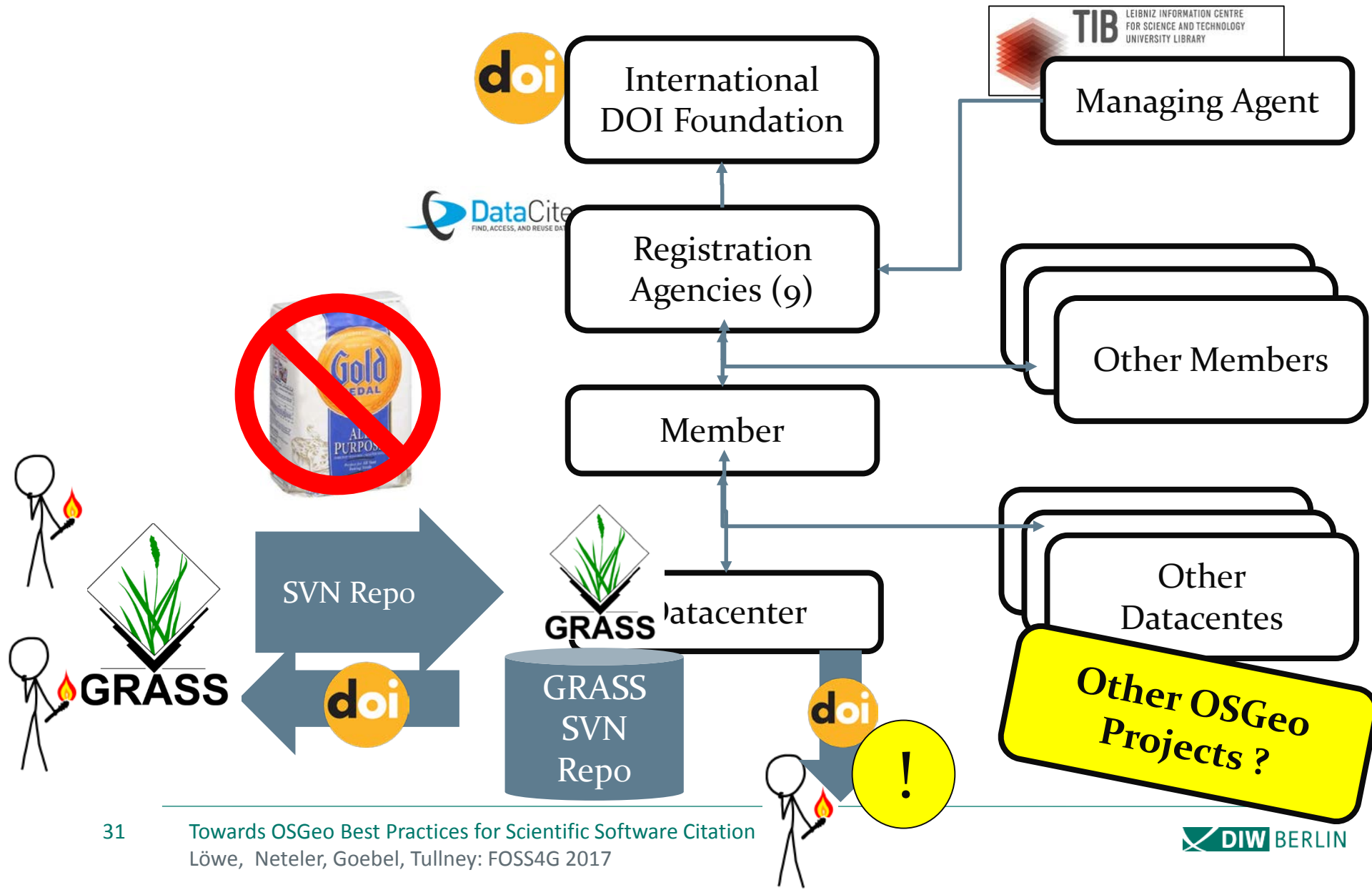


<http://zenodo.org>

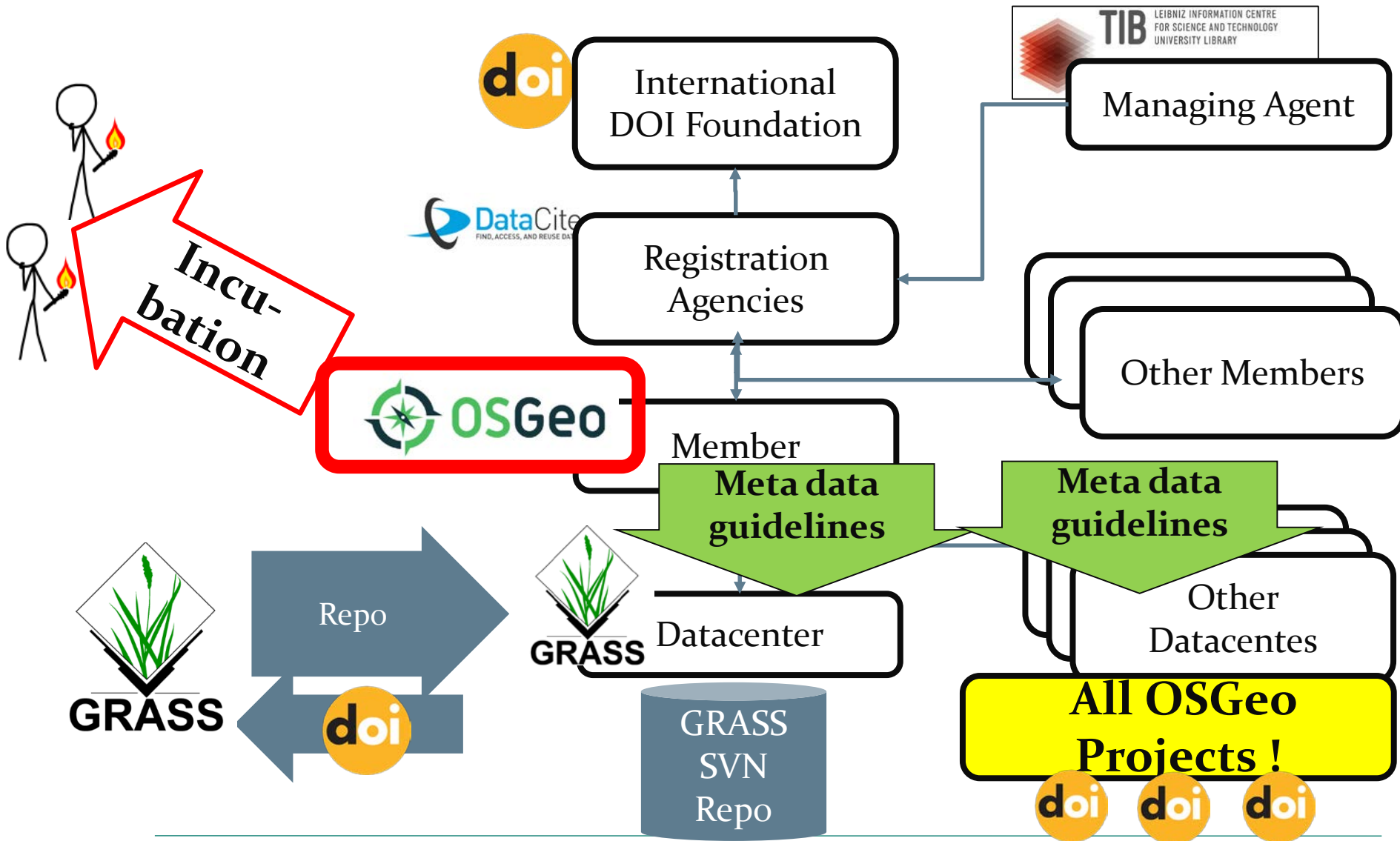


Rueda, Laura. (2017, May). Introduction to DataCite. Zenodo. <http://doi.org/10.5281/zenodo.571808>

Project Community Level: GRASS Project DataCenter



Umbrella Option: OSGeo becomes a DOI member, unlimited DOI minting for all OSGeo projects.



Opportunity: OSGeo to benefit from Datacite Services



The logo for re3data.org, with "re3" in blue and "data.org" in black.

Find an appropriate repository to access and deposit research data with re3data.org



Search our registry to find datasets, software, images, and other research material.



Generate your references automatically with our easy-to-use citation formatting tool.

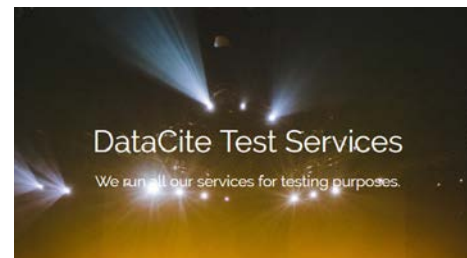
[Search.datacite.org](https://search.datacite.org)

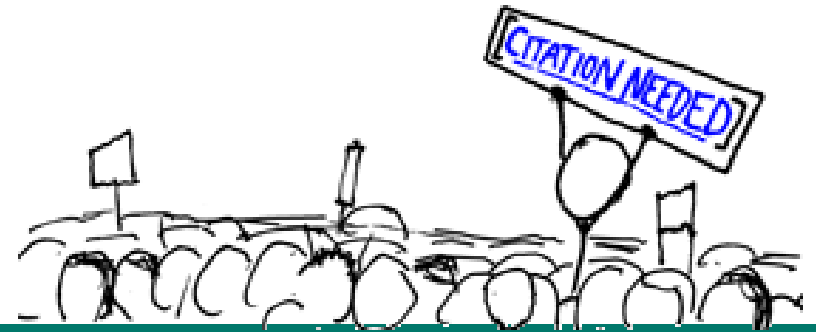
Proposal for Follow-up Action

- **Make code citation a OSGeo topic**
 - Journal
 - Projects
 - Incubation
- **Discuss DOI-/citation-related best-practices within OSGeo**



- **Explore:**
 - Conduct tests on project level





Vielen Dank für Ihre Aufmerksamkeit.



**DIW Berlin — Deutsches Institut
für Wirtschaftsforschung e.V.**
Mohrenstraße 58, 10117 Berlin
www.diw.de

Redaktion
Peter Löwe (ploewe@diw.de)

Peter Löwe¹, Markus Neteler², Jan Goebel³ and Marco Tullney⁴

¹German Institute for Economic Research
Mohrenstraße 58, 10117 Berlin, Germany
E-mail: ploewe@diw.de
<http://orcid.org/0000-0003-2257-0517>

²Mundialis GmbH & Co. KG
Kölnstraße 99, 53111 Bonn, Germany
E-mail: neteler@mundialis.de
<https://orcid.org/0000-0003-1916-1966>

³German Institute for Economic Research
Mohrenstraße 58, 10117 Berlin, Germany
E-mail: jgoebel@diw.de
<http://orcid.org/0000-0002-3243-1935>

⁴Technische Informationsbibliothek
Welfengarten 1B, 30167 Hannover, Germany
E-mail: marco.tullney@tib.eu
<https://orcid.org/0000-0002-5111-2788>
