

# Science 3D project

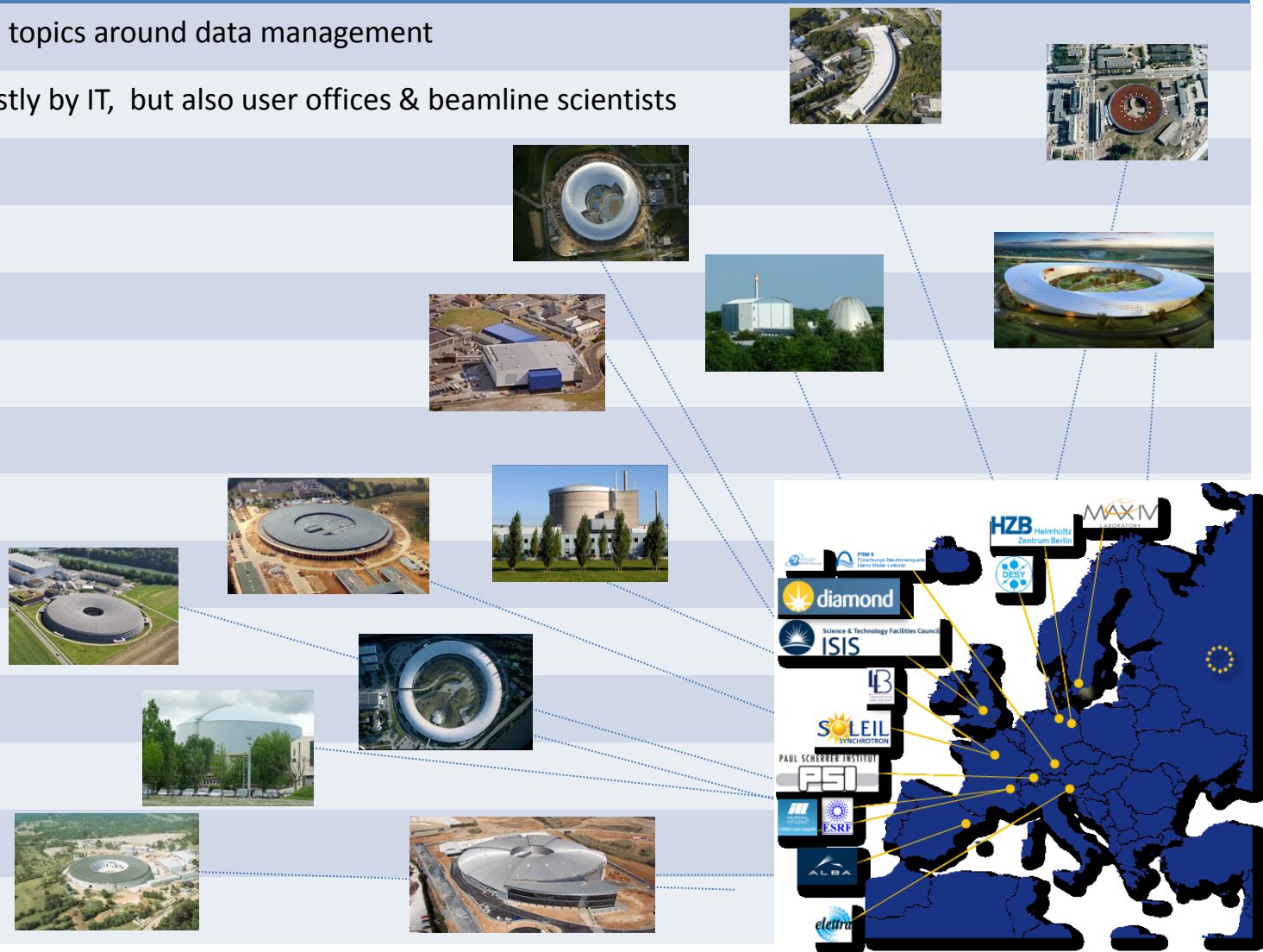
---

Frank Schlünzen  
DESY-IT



## Consortium of 13 Photon and Neutron RIs

- **Focus** entirely on topics around data management
- **Represented** mostly by IT, but also user offices & beamline scientists



## Consortium of 13 Photon and Neutron RIs

- **Focus** entirely on topics around data management
- **Number of instruments** > 200
- **Number of users** > 30.000 users/yr
- **Investment** > €4.000.000.000\*
- **Running costs** > €500.000.000/yr\*
- **Publications** > 10.000/yr\*
- **RCosts/Publication** ~ €50.000\*%
- **Data volume** >> 10PB/yr\*



% ESRF: operation budget 80-100Mio  
# publications: ~1850

\* wild guess



## Basic Idea

### Any user could

- submit or manage proposals and beamtime
- manage, access, control his/her data
- combine x-ray and neutron data
- share data within collaborations
- archive, index & publish data
- analyze data
- access resources
- capture the process in data catalogues



## Basic Idea

### Any user could

- submit or manage proposals and beamtime
- manage, access, control his/her data
- combine x-ray and neutron data
- share data within collaborations
- archive, index & publish data
- analyze data
- access resources
- capture the process in data catalogues

regardless where experiments were conducted



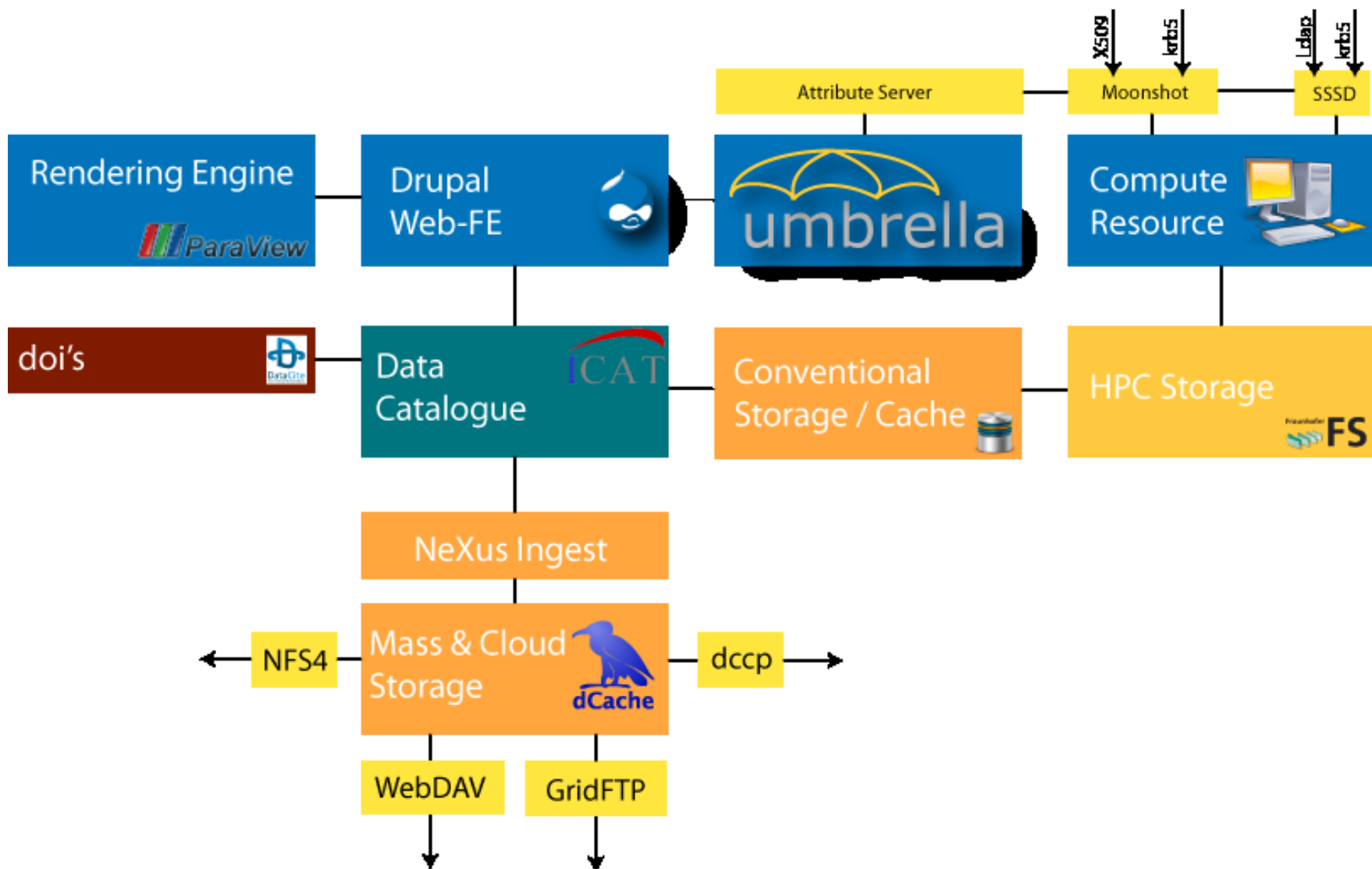
- One of the tasks:
  - A demonstrator for a data catalogue integrating several services
- The drawback:
  - Projects don't last (this one ends 09/2014)
  - Demonstrators hardly ever survive beyond the end of a project
    - Unless there is a real use case
    - So often just a waste of time & effort (& tax payers money)
- The idea:
  - Turn the demonstrator into something useful
  - Create the use case: Open Access Tomography Database

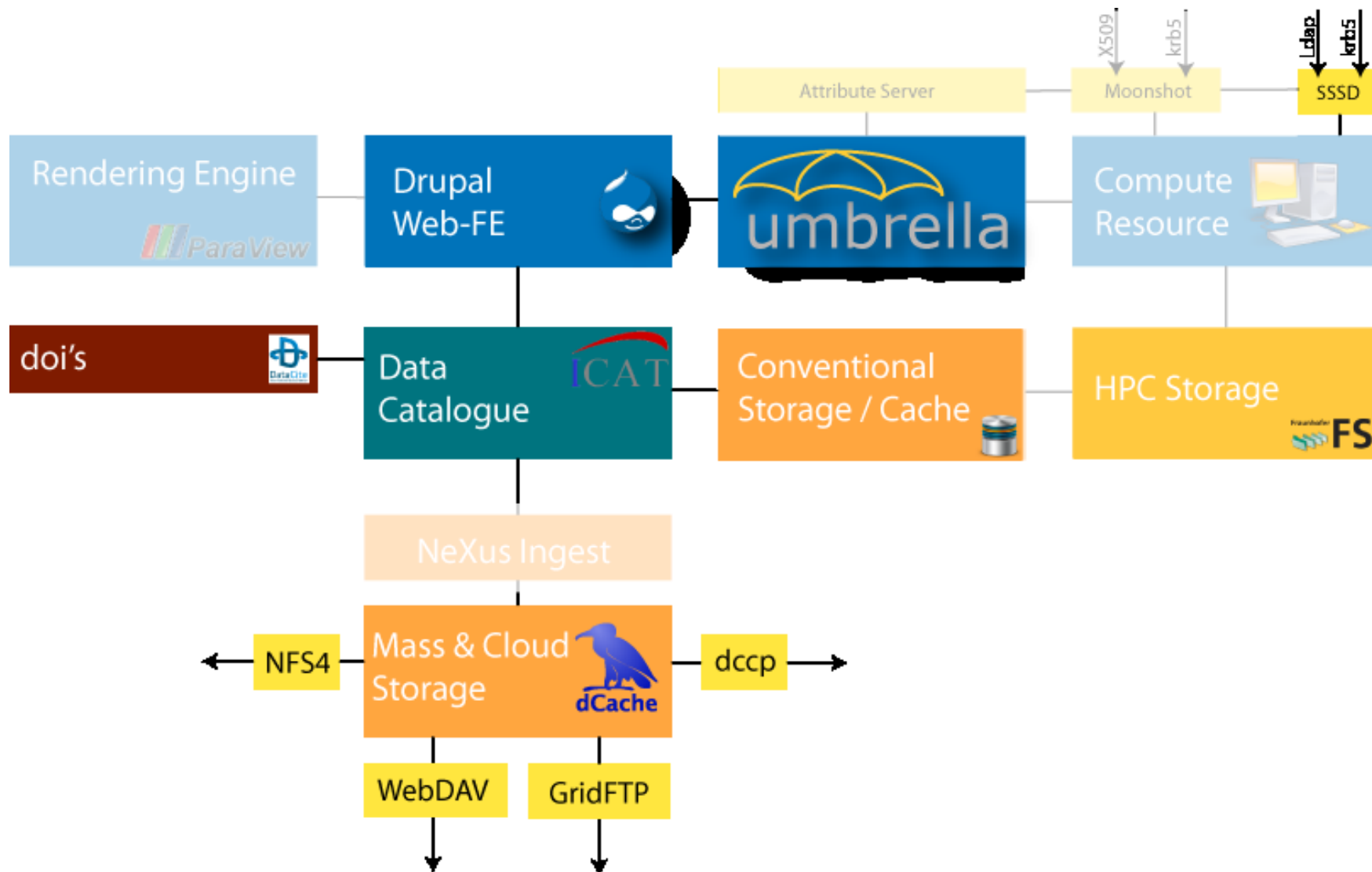
- Science 3D aims to be an **open access database** for
  - Publications
  - Raw and Derived Scientific Data
    - DOIs to make data citable
  - Supplemental Materials
    - 3D Models
    - Models suitable for Rapid Prototyping (3D prints)
    - Movies



- Science 3D aims to be an **open access database** for
  - Publications
  - Raw and Derived Scientific Data
    - DOIs to make data citable
  - Supplemental Materials
    - 3D Models
    - Models suitable for Rapid Prototyping (3D prints)
    - Movies
- Promoting Scientific results
  - Platform to present results & data
  - Easily associate data, publications, scientists

- Science 3D aims to be an **open access database** for
  - Publications
  - Raw and Derived Scientific Data
    - DOIs to make data citable
  - Supplemental Materials
    - 3D Models
    - Models suitable for Rapid Prototyping (3D prints)
    - Movies
- Promoting Scientific results
  - Platform to present results & data
  - Easily associate data, publications, scientists
- Provide rich materials for
  - Developers
  - Teachers and Educators
  - Beamline staff and colleagues





- **Storage**

- Long term mass storage dCache (unlimited space)
- Conventional storage for the catalog (secure, fast)
- Fast cluster storage (for HPC wherever needed)
- Register data with datacite

Details in talk on  
Wednesday

- **Storage**
  - Long term mass storage dCache (unlimited space)
  - Conventional storage for the catalog (secure, fast)
  - Fast cluster storage (for HPC wherever needed)
  - Register data with datacite
  
- **Compute infrastructure**
  - Work group server without access restrictions
  - Batch farm (~8000 cores)
  - GRID farm (many more cores)
  - HPC cluster (~3000 cores) with fast interconnect (IB)
  - GPGPUS (nvidia M2050 → K20X) with slow/fast interconnect

Details in talk on  
Wednesday

- **Commercial Software**
  - Tomography specific like avizo – ask Felix
    - But aim to provide whatever can be provided
  - Matlab, IDL, Mathematica, Maple, etc
  - Fast remote graphical access

- **Commercial Software**
  - Tomography specific like avizo – ask Felix
    - But aim to provide whatever can be provided
  - Matlab, IDL, Mathematica, Maple, etc
  - Fast remote graphical access
- **Services**
  - Data upload/download
  - Infinite storage (time & space)
  - DOI registration for datasets with DataCite
  - Harvest publication data from pubdb
  - Convert / Provide data in maintainable formats
  - feeds/tweets/newsletters
  - 3D print (tbc)



- DESY & HZG have a number of resources (e.g. Schülerlabor)
- Discussions how/if get closer to beamlines
  - Talk to Felix



- DESY & HZG have a number of resources (e.g. Schülerlabor)
- Discussions how/if get closer to beamlines
  - Talk to Felix
- Tomography offers tremendous insights into evolution, anatomy, materials...
  - 3D models on the screen; 3D prints for the classroom
  - Still very difficult to harvest for pupils
  - Might well be completely unrealistic, thought to figure it out here
- Science 3D might offer a platform for scientists and teacher/pupils/students to make contacts and possibly bring science more into classrooms
- Nothing we (IT) could offer ...

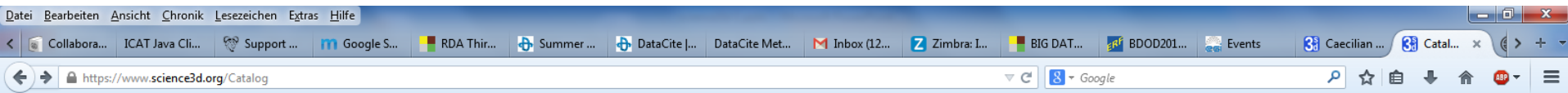


- Scientist registers a dataset (takes <5')
  - Obtain a DOI for the dataset
  - Dataset will be converted and migrated to the catalog
  - Dataset will not be accessible until released by authors
  - the entry in the catalog has to be accessible (datacite requirement)
- Scientist publishes a paper (can take quite some time)
  - Use the Data DOI to reference the dataset(s)
  - Register the publication in the database (takes <5')
  - The entry will not be accessible until the submitter hits the “publish” button
- Deposition will be announced
  - Via various channels
  - Colleagues and educators subscribing to these channels will receive updates
  - Have means to get in touch with the authors of paper and data

- Select the data
  - We are lacking the information which data precisely where used
    - Sample preparation
    - Experiment parameters
  - We can't make data public; should be responsibility of the owner/PI
    - Though ownership is an open debate
- Provide background information
  - For education
    - Prepare information suitable for teacher and students or school kids.
  - For re-use
    - Requires knowledge about the exact workflow and software stack
  - Name the sample correctly
  - Get taxonomies right
- ... in brief need your help & support (not much effort actually)!

- Authors are hesitant to contribute to an empty database
- We are hesitant to invest too much without contributors
- Authors are hesitant to contribute to a low-effort project
- ... and so on ...
- So had to kick it off & use this workshop to set the scene



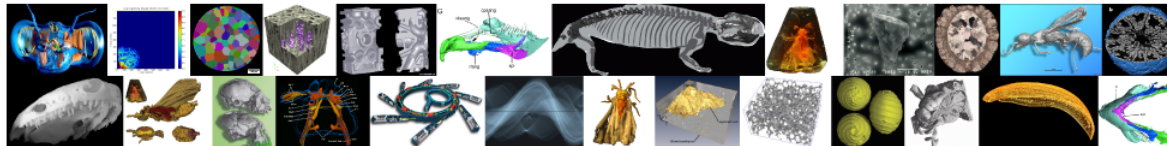


Home

## Catalog

The data catalog aims to provide original scientific data, related materials, 3D models for visualization and rapid prototyping or whatever you find useful. The data are intended to be open access for scientific or educational purposes, but subject to certain licensing models and of course best practices. By downloading and using any data you implicitly agree to these terms and conditions.

Data will be supplied with digital object identifiers (DOI), which permits to cite them in publications. And data will be safely stored for an indefinite time, so consider the catalog a safe harbor for your valuable data.



### Latest datasets

**Caecilian jaw-closing mechanics**

Thomas Kleinteich, Alexander Haas, Adam P. Summers



### Search

Search this site

Search

### Upload

publication  
reference  
tomography dataset (biology)  
tomography dataset (materials)  
electron microscopy dataset



Useful?  
Waste of time?

- Willing to contribute?
- Willing to make data open?
- Willing to add information suitable for education?
- What is missing?
- What else is needed?





Contact: [science3d@desy.de](mailto:science3d@desy.de)

Thank you for our attention !

