

Edmond – The Open Research Data Repository of the Max Planck Society



Dr. Yves Vincent Grossmann
grossmann@mpdl.mpg.de

Max Planck Digital Library



MAX PLANCK
digital library

Agenda

- What is Edmond?
- for Researchers
- for Librarians and Research Manager
- for IT Staff

What is Edmond?

Mission Statement

Edmond – Open Research Data Repository of the Max Planck Society

Edmond – serves the **publication of research data** from all disciplines

Edmond – offers scientists the ability to **create citable research objects**

Edmond – enables **secure** preservation: provision, describing, documentation, linking, publishing and archiving of all kinds of data

Edmond – supports **standardized metadata profiles**

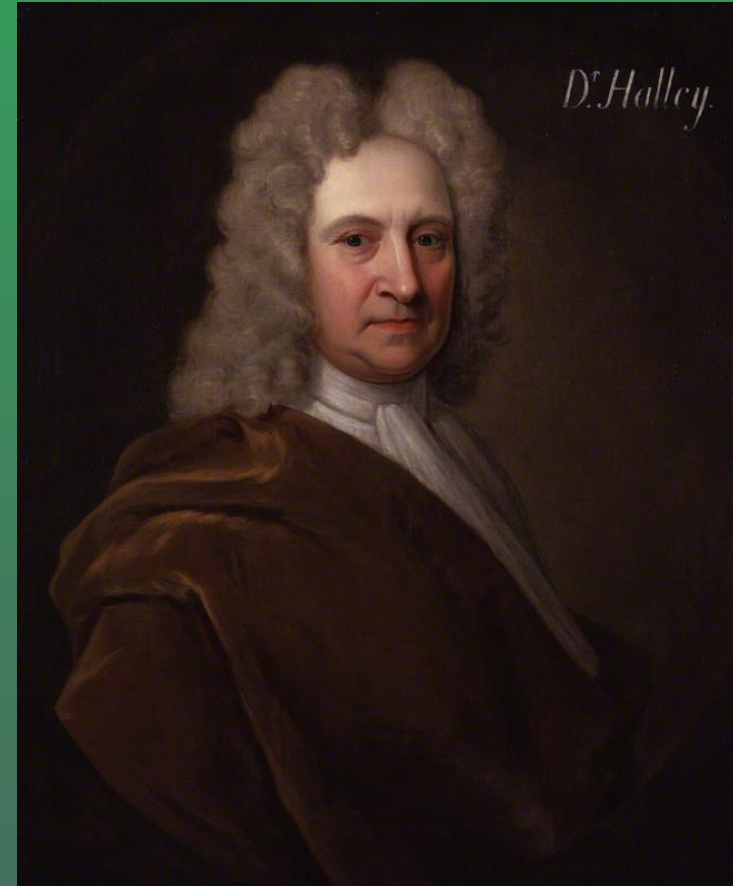
Edmond – gives an **insight into the work** of scientists at the Max Planck Institutes

Background

- institutional repository since 2014 provided by Max Planck Digital Library
- part of the Max Planck Digital Library portfolio in the field of Research Data Management
- built on Dataverse Software

The Name “Edmond”

- the first collection in Edmond consists of images of the nucleus of the comet Halley.
- therefore the repository was named after Edmond Halley (1656-1742)
- was an English astronomer
- best known for computing the orbit of the eponymous Halley's Comet and after whom the comet is named



Edmond Halley by Richard Philips,
Public Domain.

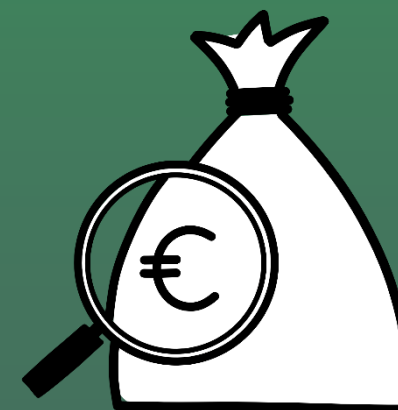
Target Group and User Community

- Max Planck researchers and information specialists
- external collaboration partners



Costs

- usage and also the support is free of charge to all employees of the Max Planck Society and their collaboration partners



Functionalities

- Single Sign On for the Max Planck Society
- Direct upload to object storage
- Multiple storage locations
- Version control
- Responsive user interface



For Researchers



Single Sign On

- login via Shibboleth with Max Planck email address
→ no registration for MPG employees needed
- external collaboration partner register locally



Upload Volume

- standard web upload
- direct upload of huge single files to object storage



Usability

- Responsive design
 - desktop
 - pad
 - smartphome



Persistent Identifier I

- Digital Object Identifier (DOI) are used as persistent identifier
- DOI provider is [DataCite/TIB Hannover](#)
- automatically reserving a DOI for a new dataset



Public domain

Persistent Identifier II

- ORCID for authors
- ROR for affiliations (incl. autosuggestions focused on Max Planck Institutes)

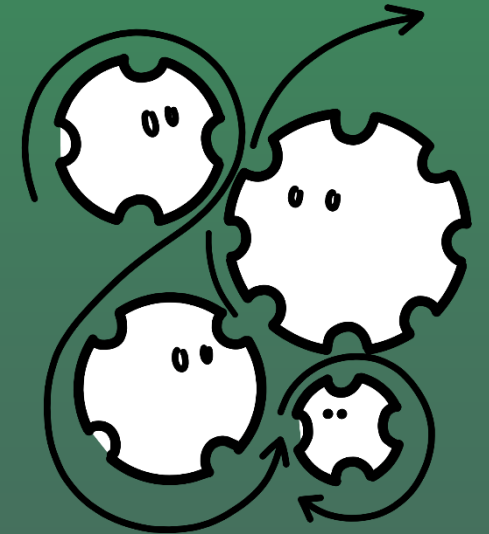
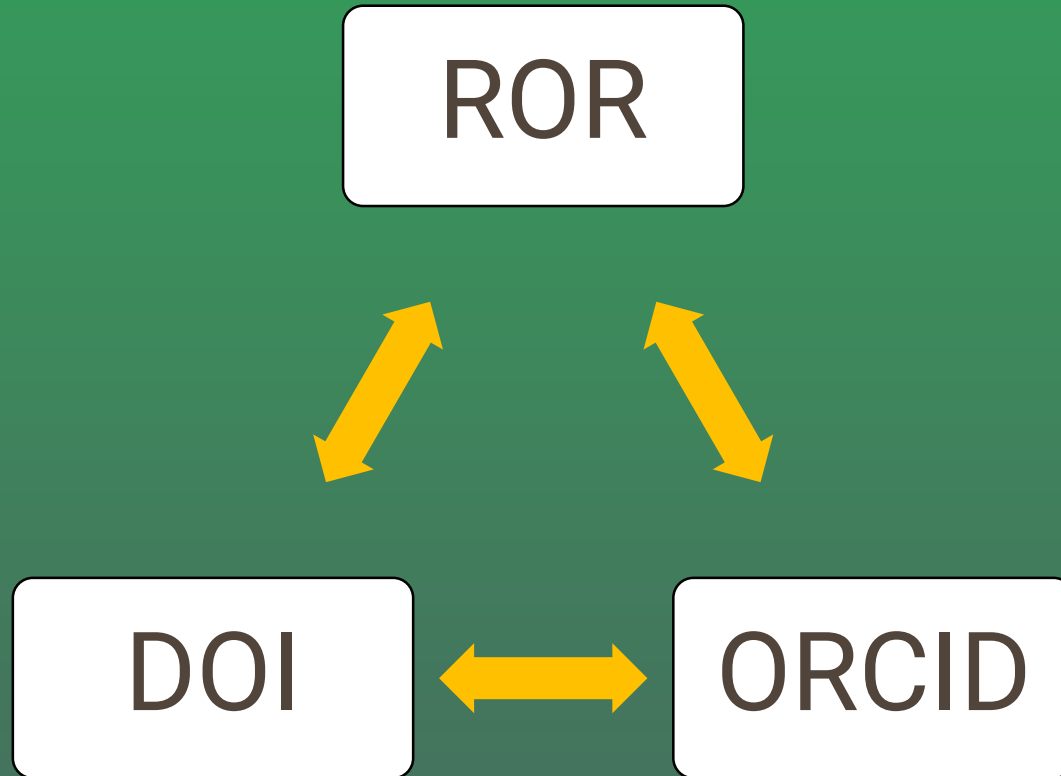


Research Organization Registry, CC BY 4.0,
<https://doi.org/10.5281/zenodo.4701802>



Public domain, CC0

Identifier Network



Roles and Permissions

- different roles and permissions are available:
 - Administrator
 - Contributor
 - Curator
 - File Downloader
 - Member
- detailed description is available in the section on Roles and Permissions in the [Dataverse Guide](#)



Private URL

- private URL for unpublished datasets
- e.g. for reviewing processes, where unpublished research data should already be shared



Licenses

Edmond supports the following licenses:

- **Creative Commons**

- Attribution 4.0 International (CC BY 4.0)
- Attribution ShareAlike 4.0 International (CC BY-SA 4.0)
- Public Domain Dedication (CC0 1.0)

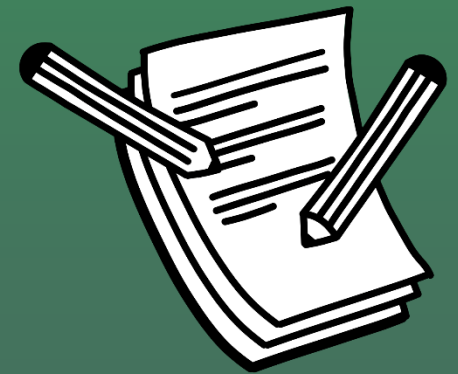
- **Open Data Commons**

- ODC Public Domain Dedication and License
- ODC Attribution License (ODC-By) v1.0
- Open Database License (ODbL) v1.0

- **Data License Germany**

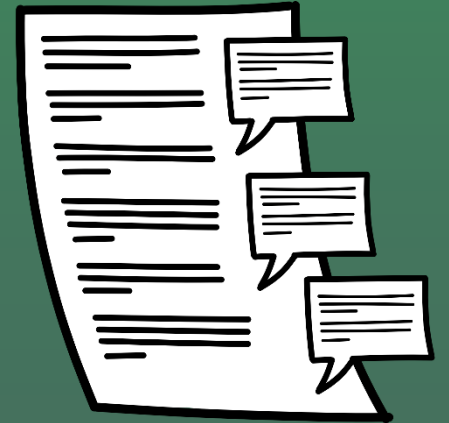
- Data license Germany – attribution – version 2.0
- Data license Germany – Zero – version 2.0

- self-defined license



Guidance

- [Edmond Help Section](#)
- [generic Dataverse Guide](#)
- [MPDL Terms of Use](#)



Edmond Support

- support team for all needs:
edmond@mpdl.mpg.de
- virtual tutorials on the service
- open for future requests and improvements by the user community



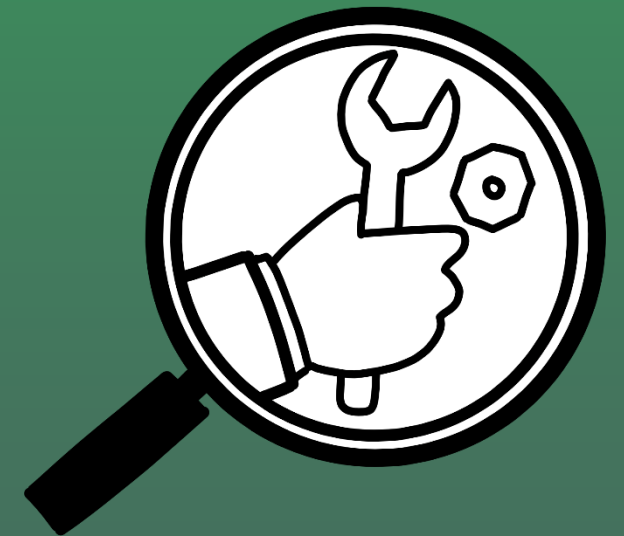
Individual Coaching and Events

- individual coaching upon request
- special [RDM support](#) for MPG researchers available
- [RDM workshop and events](#) for MPG researcher, information specialists and other interested colleagues



Supplementary Services

- RDM Support (individual support for researchers and information specialists)
- [RDMO for MPG](#) (online tool for data management planning)
- [RDM Information Platform](#) (website with generic and MPG RDM information)
- [Labfolder](#) (electronic laboratory notebook)
- [MPG.PuRe](#) (publication repository)
- [Keeper](#) (research project archive)

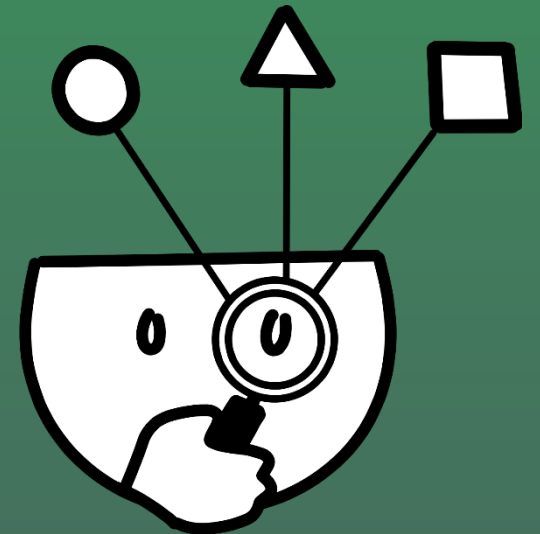


**For Librarians and
Research Managers**



Level of Curation

- no data curation
- content distributed as deposited



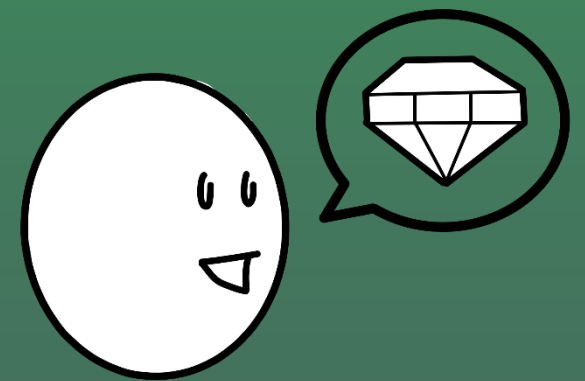
Data Handling and Changes

- changes on published datasets are always possible, e.g. add or delete files, also editing metadata
- all changes will lead to a new version of the dataset, needs to be published again
- depending on the changes, a minor or major release version of the dataset will result



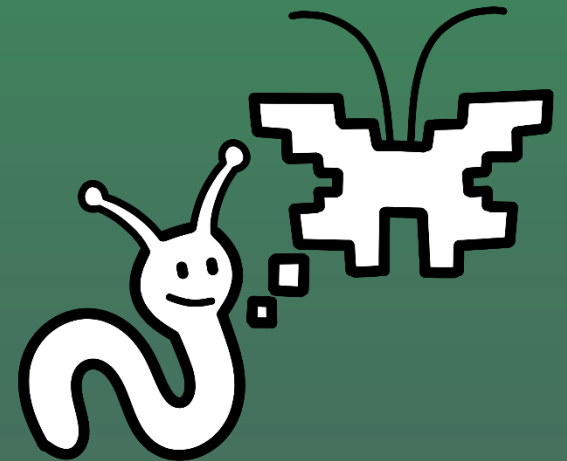
Standards

- Identifier Standard
 - [DOI](#)
 - [ROR](#)
 - [ORCID](#)
- Datacite
- Dublin Core
- Export formats (see below)
- OAIS model



Metadata

- metadata schema in Edmond is following [DataCite Metadata Schema 4.4](#)



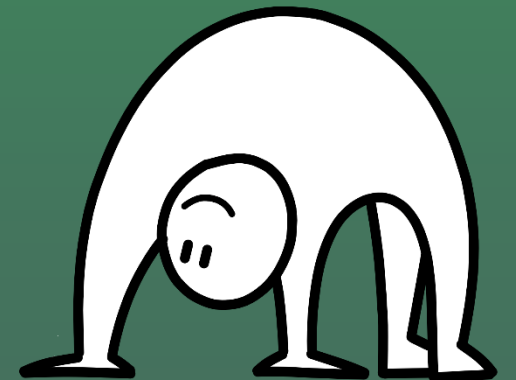
Export formats

- Download multiple files as ZIP
- no change of file formats
- Different types of metadata exports
 - Dublin Core
 - DataCite
 - DDI
 - JSON-LD (schema.org)



File Structure

- folder structure is possible
- can be generated via browser



Registered in re3data

re3data.org

REGISTRY OF RESEARCH DATA REPOSITORIES

<http://doi.org/10.17616/R3N33V>

Edmond



Dataverse Community

- community-driven Open Source Software
- [Dataverse Community worldwide](#)
- regular Dataverse Meetings
- annual Dataverse Conferences



Data Protection

- compliant with German Federal Data Protection Act (BDSG) and EU General Data Protection Regulation (DSGVO)
- Data Protection Officer of the Max Planck Society: datenschutz@mpg.de



Data Management Plans

- Edmond can also be used as open research data repository in your data management plan for funding applications
- special service for writing DMPs with [“RDMO for MPG”](#)

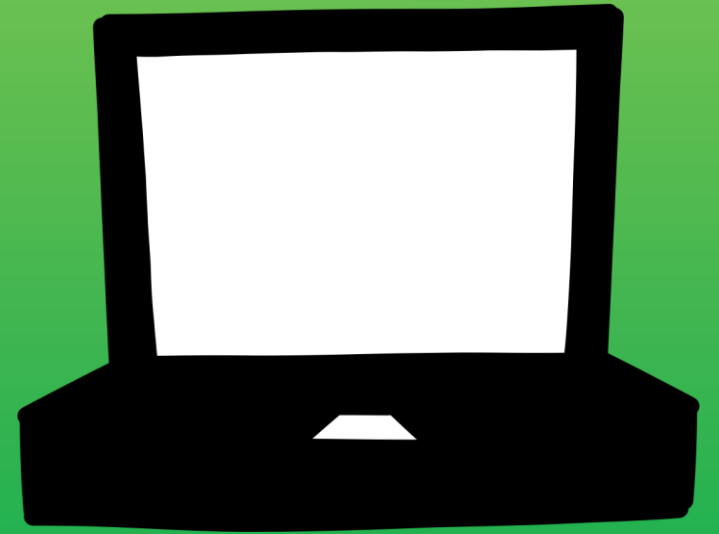


Long-term Archiving

- data, once published, will be kept and made available for at least ten years
- in concordance with “Rules of Good Scientific Practice” for the Max Planck Society
- Preservation Policy for Edmond available

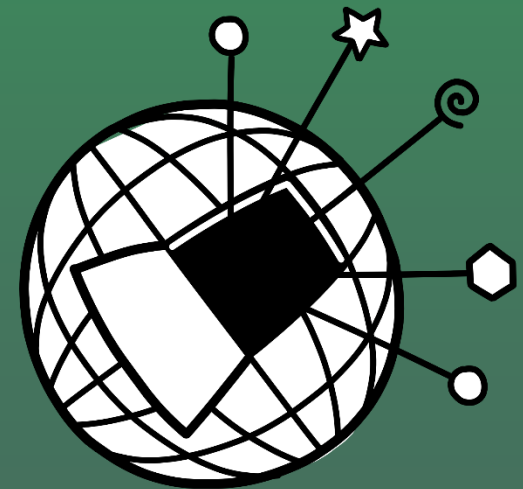


For IT Staff



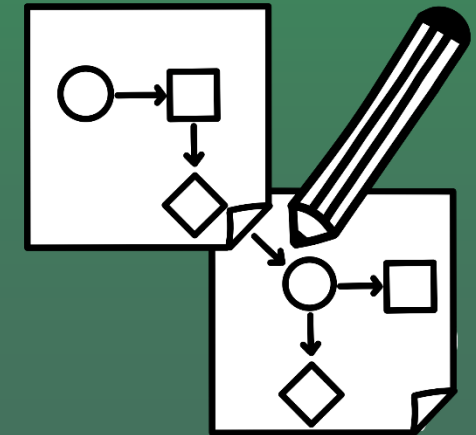
Technical Infrastructure and Data Hosting

- hosted at GWDG (ISO 27001 certified)
- Ceph Object Storage
- Single Sign On – MPG AAI



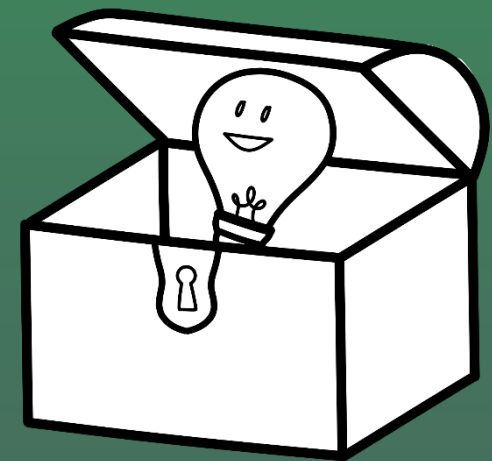
Object Storage

- Ceph Object Storage
- Software-Defined-Storage
- Open Source
- S3 Interface
- Redundant: “8 + 3” Erasure Coding



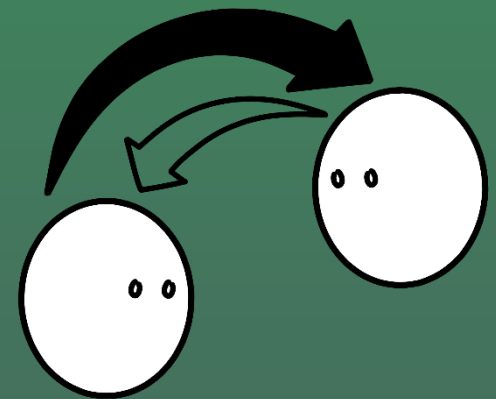
Data Safety

- diverse strategies to back up the data (back up of server and database itself)
- backups are created in frequent intervals (the shortest is every 12 hours)
- backups are stored in several locations on servers of the Max Planck Society as well as on tapes



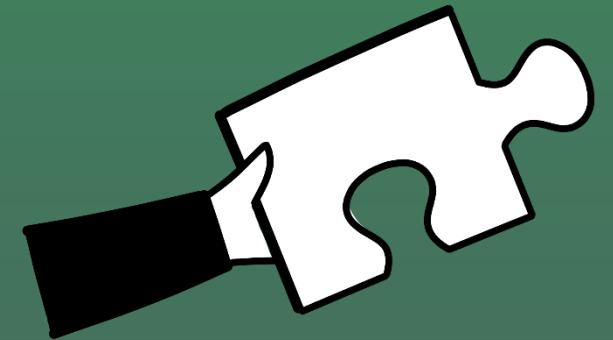
Uploader

- separate [DVUploader](#) is available; more information please see the [Dataverse Guide](#)



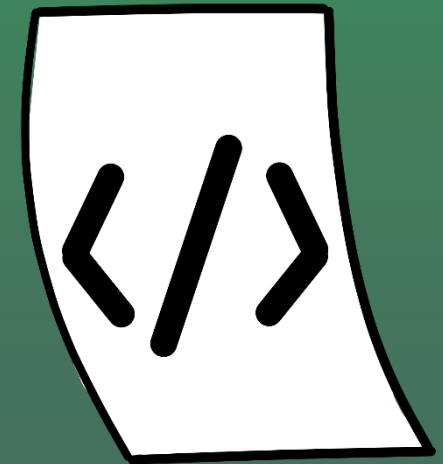
Software Components

- Linux RHEL/CentOS: Operating system
- Payara: Application server
- PostgreSQL: Application database
- Java: Programming language
- Solr: Indexing



MPDL Tech Team

- Development
 - Development of MPG-specific Dataverse features
 - Support of Dataverse community by providing bug fixes and features
 - Applying Dataverse updates and fixes from community
- QA
 - Testing on different levels (manual tests, automated tests, unit tests, integration tests)
 - Providing automated testing features
- System Administration
 - Server and storage specific maintenance and operation
 - Monitoring security and performance
 - Together with MPG affiliated computing centre GWDG



Thank you for your attention

edmond@mpdl.mpg.de