

# Current Developments in the Research Data Repository RADAR

- **Felix Bach\*** - Head of Research Data Department  
[felix.bach@fiz-karlsruhe.de](mailto:felix.bach@fiz-karlsruhe.de)
- **Kerstin Soltau\*** - Product manager RADAR
- **Sandra Göller\*** - Data Steward, Metadata Specialist
- **Christian Bonatto Minella\*** - Data Steward,  
Project Manager

\*



# Outline

- **RADAR** - advanced generic **R**esearch **D**ata **R**epository
- RADAR in the National Research Data Infrastructure (**NFDI** - “*Nationale Forschungsdaten-Infrastruktur*”)
  - RADAR4Chem
  - RADAR4Culture, ...
- **Concept** of RADAR -
- Current **developments**
- **Closing** the **gaps** for an implementation of **FDOF** in RADAR



# RADAR – Research Data Repository

RADAR (**R**esearch **D**ata **R**epository) is a

- cross-disciplinary research data repository
- for the archival and publication of research data
- of completed scientific studies and projects.

Funded by DFG from September 2013 until August 2016

- Emphasis on „Long tail“



# RADAR Service Levels

## Data Archival

- „dark archive“
- Flexible retention period (5, 10, 15 years)
- Data providers control access rights (private/shared/public)

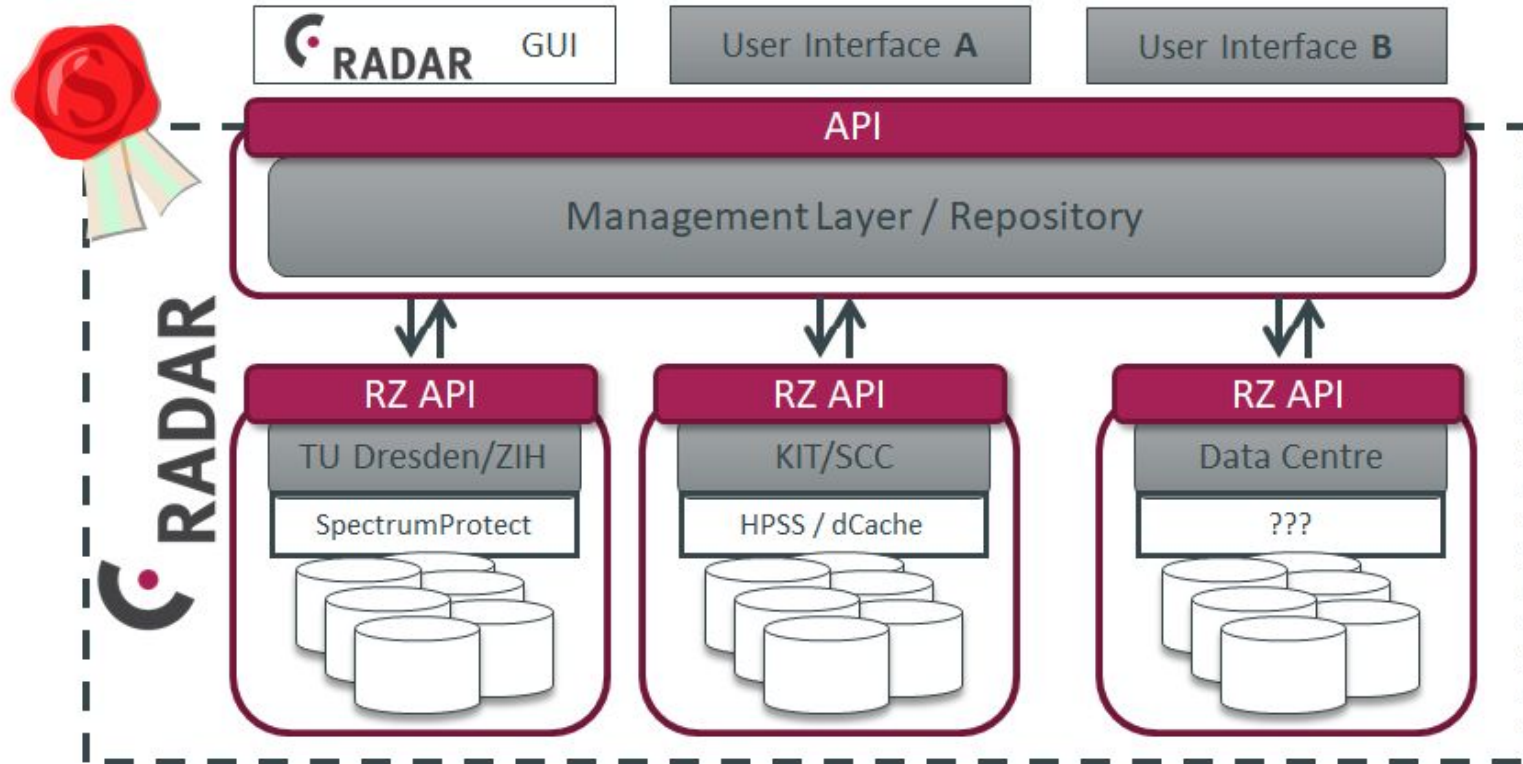


Source: Wikipedia Commons, Simon A. Eugster. CC 3.0 BY SA

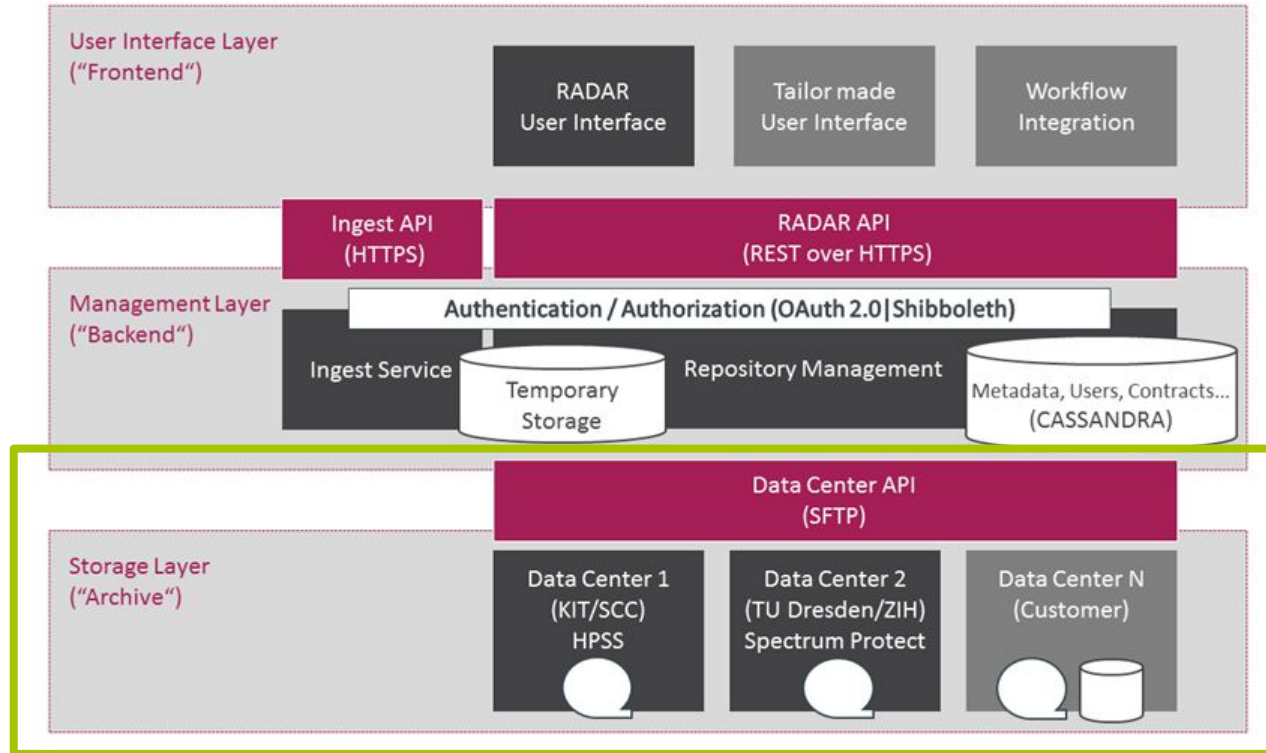
## Publication (incl. Data Archival)

- Unlimited retention period (25+ years)
- Optional embargos
- DataCite DOI
- Interface for peer review
- Choice of licenses, no CC0 required
- Metadata indexed, e.g. by RADAR, DataCite, Clarivate, Google

# Schematic Architecture



# Modular, open system architecture



# RADAR users in Germany

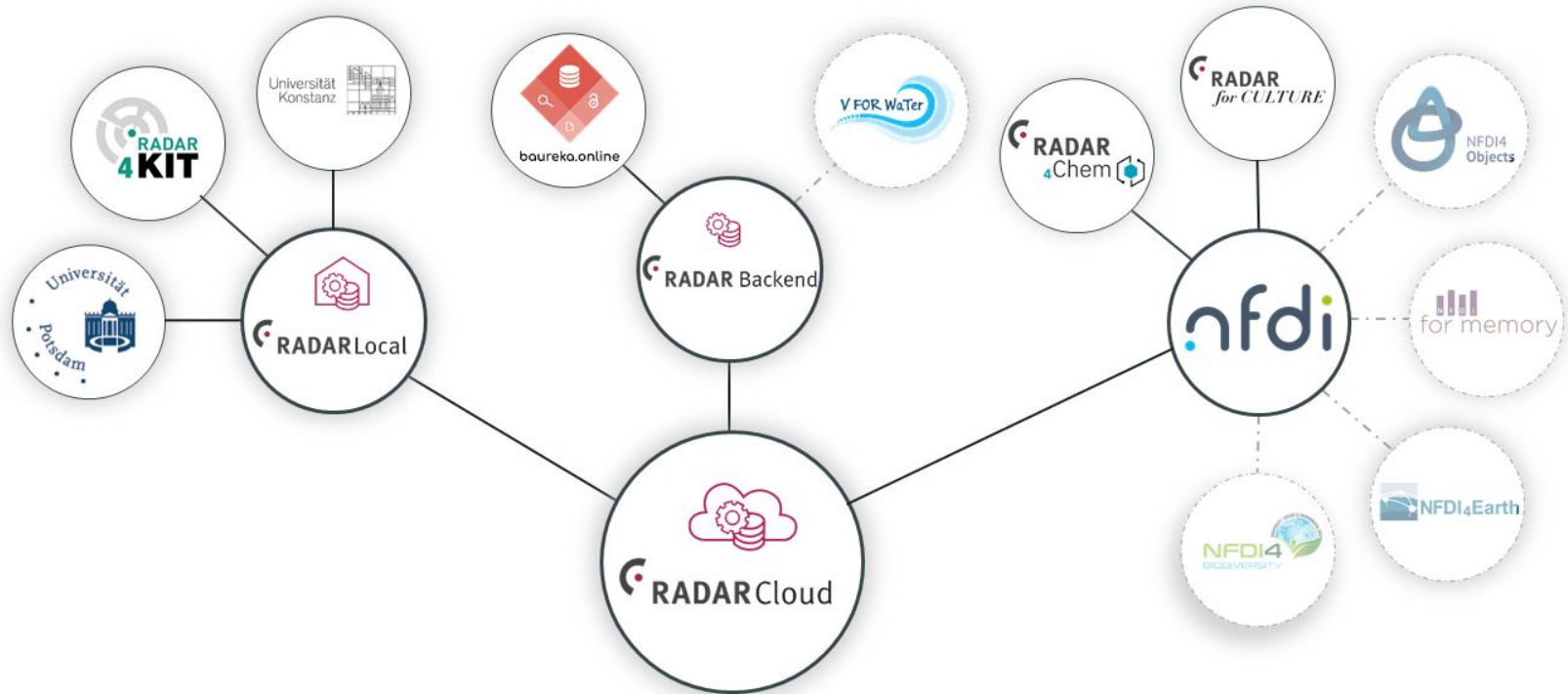


FIZ Karlsruhe – Leibniz Institute for Information Infrastructure





# Diverse offers





# RADAR4Chem, RADAR4Culture

## Core functions and features



like RADAR cloud...

- Data publication for **all data types** and **formats**
- Unlimited retention period (min. 25 years / **bitstream preservation**)
- **DOI registration** (DataCite) incl. DOI reservation option
- Optional **embargo** (1-12 months, endless)
- Optional **peer review** prior to publication
- **RADAR metadata schema** based on DataCite metadata schema
- Choice of **licences** for re-use (incl. CC)
- **Linking** to other digital resources ("related identifier")
- **Metadata indexing** (RADAR, DataCite, harvesting via OAI-PMH)
- **Authentication** option also via DFN-AAI (Shibboleth SSO)
- **Metrics** on access and download figures

**special for NFDI:**     - **FREE**,   - **ONLY DATA PUBLICATION** (no archiving-only option)  
   to **foster FAIR data**



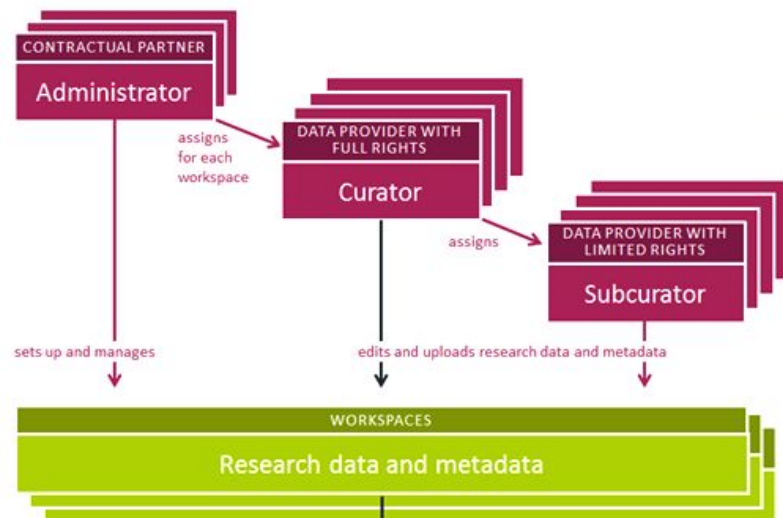
## RADAR4Chem – Use Cases

- allows the publication of all data types and formats.
- complements the portfolio of already existing subject repositories in chemistry

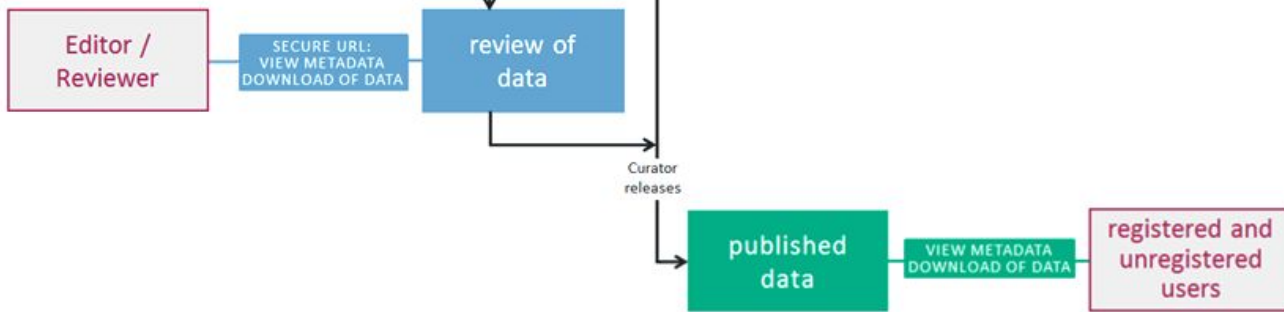
Possible data publications suitable for RADAR4Chem:

- datasets which do not fit in an existing subject-specific repository
- interdisciplinary datasets
- “distributed” datasets,
  - that do not fit into one single existing repository due to diversity of data types
  - interlinking via “related identifiers”

# Workflow: Roles and rights



Contract	-	admin
-Workspace	-	data provider
-Data Set	-	-curator
		curate
		publish/archive
		-subcurator
		upload
		annotate



# FAIR Metadata annotation

- at the level of dataset (mandatory), folders and files (optional)
- via form-based editor, XML upload, RADAR API
- RADAR MD Schema: generic, based on DataCite MD kernel
- Interfaces to ORCID, Crossref Funder Registry, ROR (in work), Open Street Map, ontology- & terminology services (e.g. TIB, lobid, GND)
- supports custom subject specific MD schemas

## 10 mandatory parameters

Identifier\* (DOI / RADAR-ID)  
Creator\*  
Title\*  
Publisher\*  
Production year or time span  
Publication year\*  
Subject area  
Resource\*  
Rights\*  
Rightsholder



## 13 optional parameters

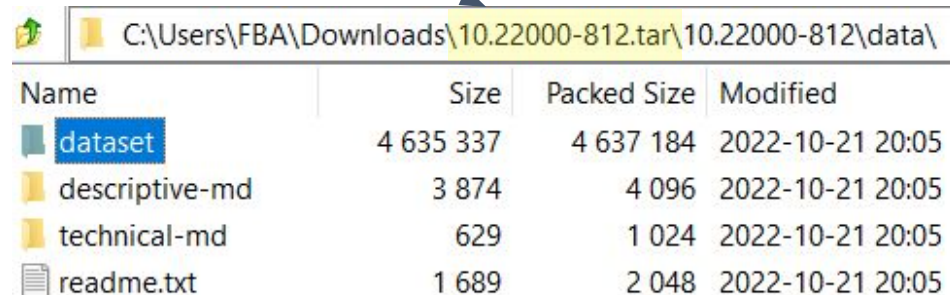
Additional title  
Description  
Keyword  
Contributor  
Language\*  
Alternate identifier\*  
Related identifier\*  
Geo location\*  
Data source  
Software type  
Data processing  
Related information  
Funder information \*

\* bases on  DataCite metadata kernel v4.0



# Data Set concept: BagIt

- Data Sets can have 2 **states**:
  - A) **Temporary** (editable) →
  - B) **Published** (with DOI) or **Archived** (no DOI) → **SIP** in archives
- **SIP** (“submission information packages”) are **BagIt** containers
  - stored as **\*.tar** (“tape archive”) file **with DOI as name** (for published data)

File Explorer Path: C:\Users\FBA\Downloads\10.22000-812.tar\10.22000-812\data\

Name	Size	Packed Size	Modified
dataset	4 635 337	4 637 184	2022-10-21 20:05
descriptive-md	3 874	4 096	2022-10-21 20:05
technical-md	629	1 024	2022-10-21 20:05
readme.txt	1 689	2 048	2022-10-21 20:05

# Current developments

- Disciplinary MD
  - Uploading of MD as XML including validation (already implemented)
  - RADAR Metadata service to allow curators to
  - create custom MD schemes & MD input forms for RADAR Frontend
- Management of Big Data
  - architectural improvements (minimize data moving, link existing archived data)
  - single file access for data sets
- integration of previews and editors
- Versioning of data sets

# Possible / Planned improvements

- Closing gaps for implementing FDO-F

- **Digital Object Interface Protocol (DOIP)** to **allow applications to interact with RADAR Data Sets** (using **existing RADAR API**)
  - “clients” - viewers, editors etc.
- **Identifier/Resolution Protocol (IRP)**
  - Currently, RADAR DOIs point to (HTML) **Landing Pages** with download options (see image)
  - **Enrich Landing Pages** (e.g. with schema.org) - point to FDO-IR?
  - **Allow to directly access**
    - **Identifier Record (FDOF-IR, to be implemented)**
    - several MD-sections in different formats (see image)
    - License


DOI: [10.22000/812](https://doi.org/10.22000/812)

Publication date: 2022-10-21

Download Dataset

**DOWNLOAD (4.7 MB)**

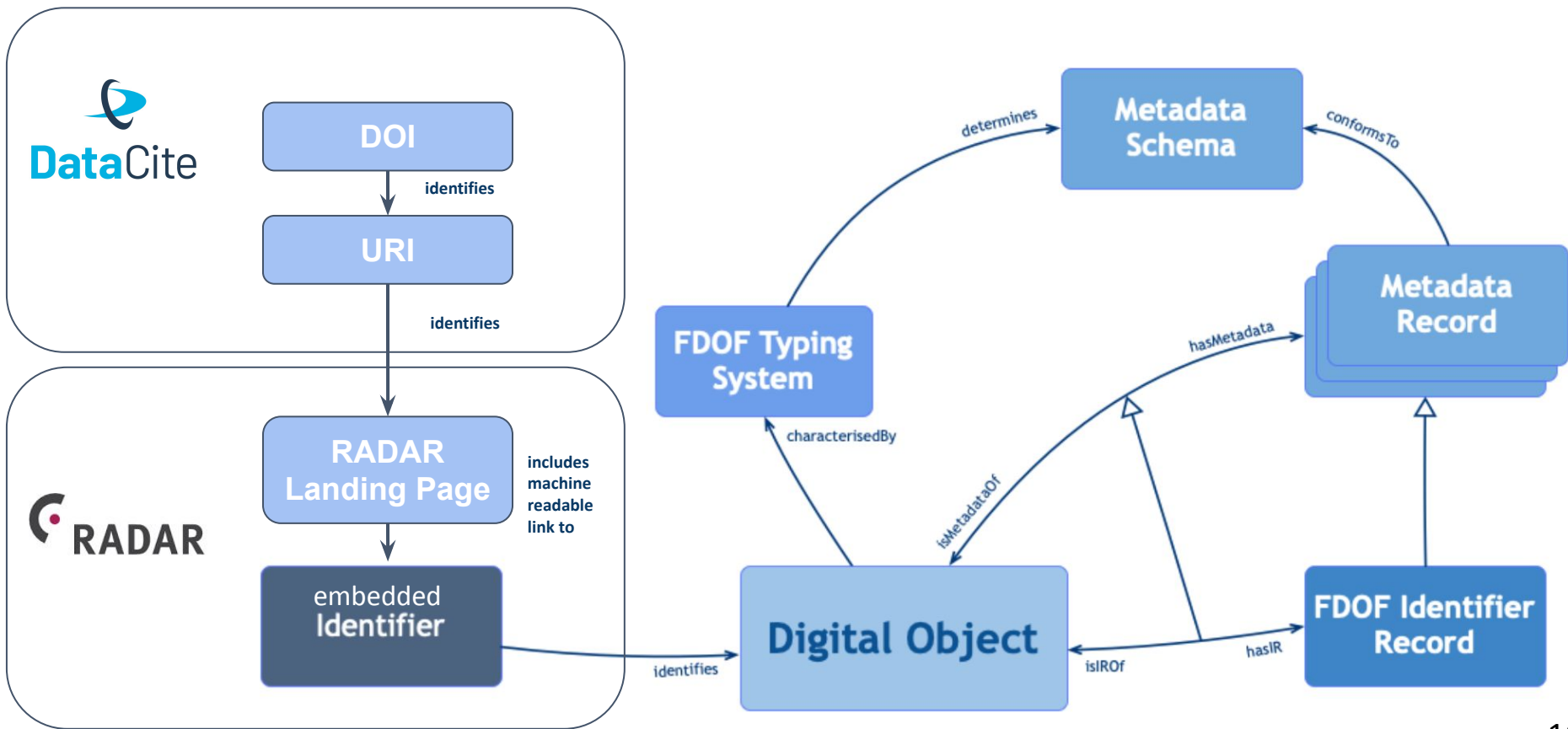
Download Metadata

<b>RADAR</b>	▼	<b>DOWNLOAD</b>
<b>RADAR</b>		
<b>DataCite</b>		
<b>OAI-DC</b>		
<b>PREMIS</b>		
<b>BibTex</b>		
<b>Endnote</b>		
<b>RIS</b>		
		



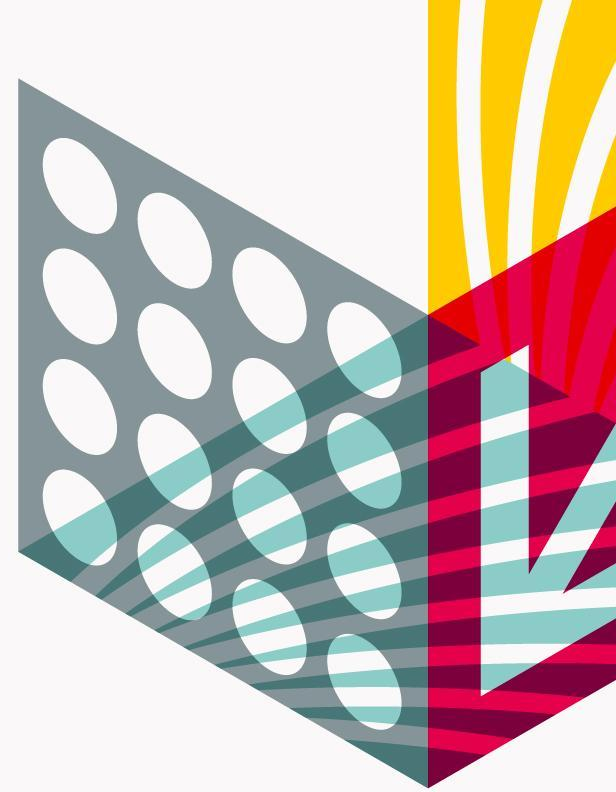
# Possible / Planned improvements

- easy implementation



# Thank you for your attention

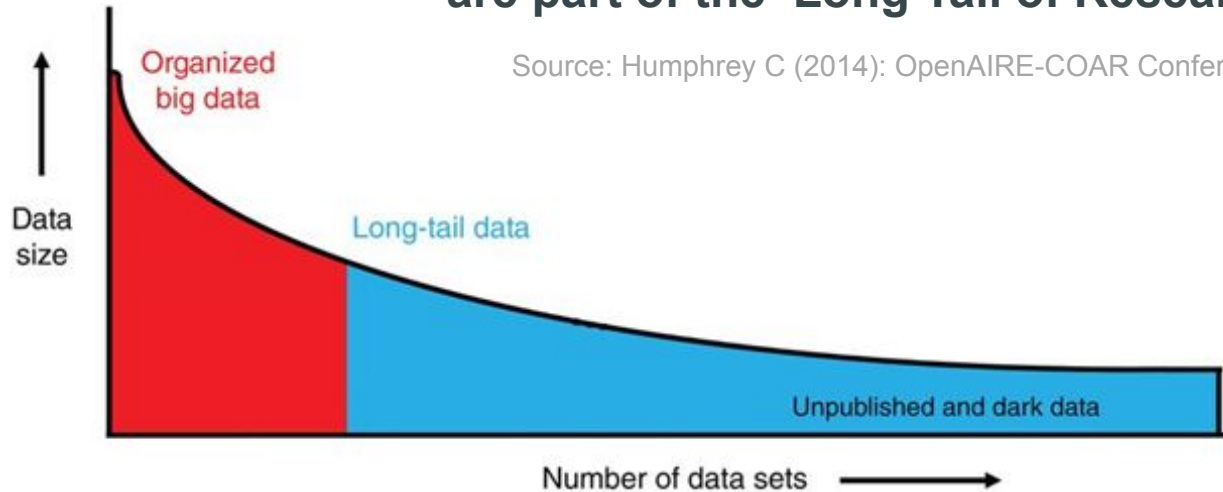
*Special Thanks to  
NFDI, NFDI4Chem, NFDI4Culture, DFG*



# The ,Long Tail' is Relevant

**“The majority of datasets produced through research are part of the ‘Long Tail of Research Data’”**

Source: Humphrey C (2014): OpenAIRE-COAR Conference, Athens

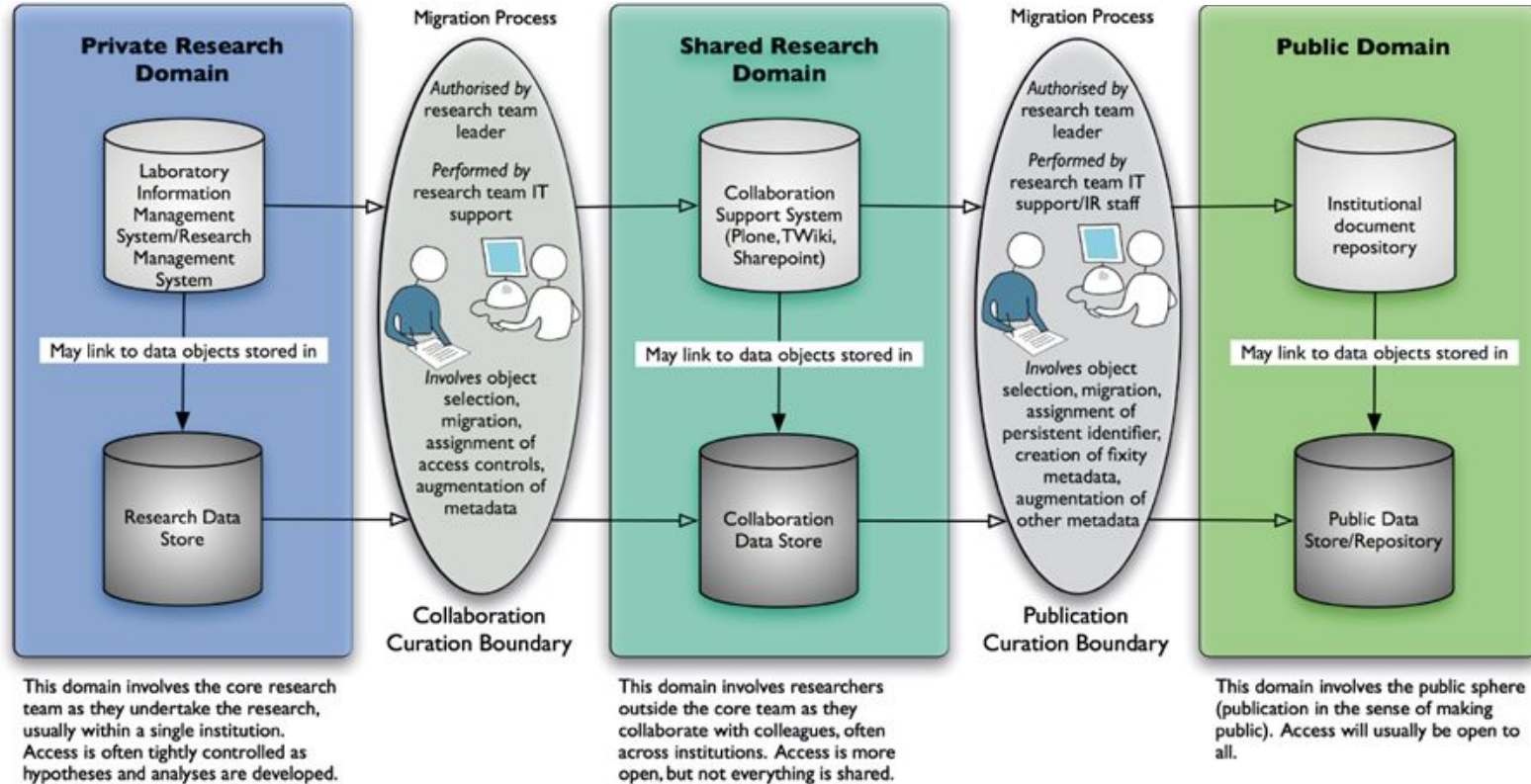


Source: Ferguson et al. (2014)

*Big data from small data: data-sharing in the 'long tail' of neuroscience.* DOI: 10.1038/nn.3838



# Data Lifecycle And Focus of RADAR



# Data Security

- For each dataset, a checksum is calculated upon ingest
- The checksum is re-calculated after writing to tape ('read after write')
- Three copies at three locations (2x Karlsruhe, 1x Dresden)
- Different hardware, software and administration
- Routine migration of data to new storage media, including fixity check

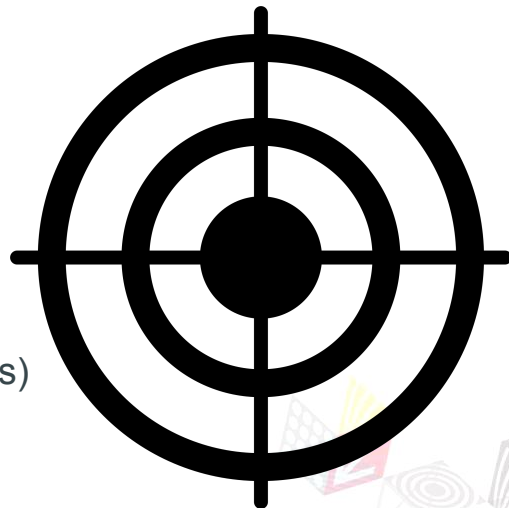


# RADAR4Chem – Workflow: Authorization

- **Contact forms** on NFDI4Chem and RADAR website
- User contacts FIZ Karlsruhe and provides information:
  - Person / institution
  - Subject area
  - Research project
- Research dataset (e.g. size, number of individual files, file format(s), planned publication date, optional embargo)
- FIZ Karlsruhe advises the user (Right choice of repo? How to use RADAR4Chem?)
- FIZ Karlsruhe sets up a **separate workspace for the user in RADAR4Chem**:
  - Authorizes the user as **data provider** in this workspace
  - Storage volume **max. 10 GB** (for highly relevant datasets we can make exceptions)

# Target groups of RADAR

- Universities and research institutions
  - Institutional offering for research data management
  - Integration with existing institutional portals
- Individual researchers
  - Publication and archiving of research data
- Scientific communities
  - GLAM (cultural heritage)
    - Long-term preservation of digitized materials (master scans)
    - Access copies for online access
  - via **NFDI / EOSC**
    - Chemistry, Culture, Earth, Biodiversity...





# Scope

## Within scope

- Data of (completed?) research projects
- Bitstream preservation
- Three copies, two data centers
- 25+ year retention phase for published data
- ‘Long-tail’ data
- Data under German/EU jurisdiction



## Outside of scope

- Management of active data
- (Scientific) data curation
- Functional long-term preservation
- Regular fixity checks
  - (Our data centers do this)
- sensitive data, users outside EU
  - Limitations: Contracts with data centers
  - We are working on data storage with higher security level
- No ‘Big Data’ disciplines
  - We try to change the architecture of RADAR to allow Big Data

## 2-step data transfer

I. Data is  
uploaded  
into  
**Temporary  
Storage**

- Transmission (https) of individual files or containers (e.g. zip, tar)
- Editing, metadata annotation and quality assurance (up to 6 months)
- Data status: "Pending" / "In Review"
- Data integrity: Calculation of checksums
- Data security: Redundant disk system (RAID-6) with daily backup

-> **Curator issues a publication**

II. Data is  
transferred  
to the  
**Permanent  
Storage  
(Archive)**

- Transmission (sftp) of dataset archives (.tar) - 3 different copies, different locations
- Data status: „published" (DOI). No more editing possible
- Dataset archive uses BagIt standard and contains:
- Research data (incl. complete file / folder structure), all metadata, other info (manifest, checksum)
- Data integrity: calculation of checksums
- Data security: 3 copies at 3 locations. Regular migration to new storage media (incl. fixity checks)