

Forschungsdaten und Langzeitarchivierung

ToC

- Definitionen
- Zwischenfazit
- Internationale Diskussionen, Entwicklungen
- Nationale Diskussionen, Entwicklungen
- Zusammenfassung

Definition Langzeitarchivierung (LZA)

„Langzeit“ ist die Umschreibung **eines nicht näher fixierten Zeitraumes**, währenddessen wesentliche, nicht vorhersehbare technologische und soziokulturelle Veränderungen eintreten; ... Es gilt also, **jeweils geeignete Strategien** für bestimmte digitale Sammlungen zu entwickeln, die je nach Bedarf und zukünftigem Nutzungsszenarium die **langfristige Verfügbarkeit** und **Nachnutzung** der digitalen Objekte sicherstellen.

2007/2009

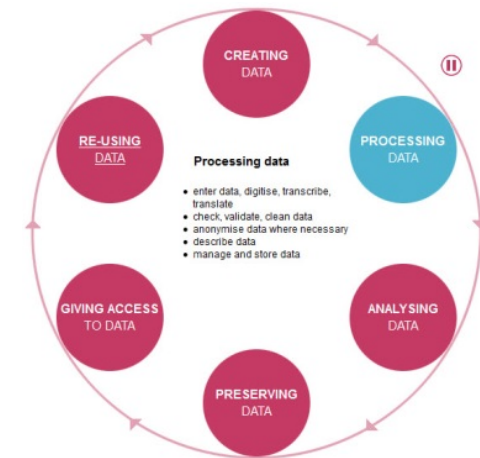
nestor Handbuch - Eine kleine Enzyklopädie der digitalen Langzeitarchivierung

Version 2.0, Juni 2009 (gedruckte Version, online seit 2007)

Definition Forschungsdatenmanagement (FDM)

Forschungsdatenmanagement bezeichnet den **PROZESS**, der alle Methoden und Verfahren umfasst, die zur Sicherung der **langfristigen Nutzbarkeit** von Forschungsdaten angewendet werden:

die **Allokation**,
die **Generierung**,
die **Bearbeitung**,
die **Anreicherung**,
die **Archivierung**,
die **Veröffentlichung**.



2013

„Die digitale Forschungswelt“ als Gegenstand der Forschung

Maxi Kindling M. A / Prof. Dr. Peter Schirmbacher, DOI: <https://doi.org/10.1515/iwp-2013-0017>

UK Data Archive: Research Data Lifecycle
<http://data-archive.ac.uk/create-manage/life-cycle>

Definition Forschungsdaten (FD)

Zu Forschungsdaten zählen u.a.

- Messdaten,
- Laborwerte,
- audiovisuelle Informationen,
- Texte,
- Surveydaten,
- Objekte aus Sammlungen,
- Proben, die in der wissenschaftlichen Arbeit entstehen, entwickelt oder ausgewertet werden,
- methodische Testverfahren (z.B. Fragebögen, Software und Simulationen)



2016

Forschungsdatenmanagement in DFG-Anträgen : Was kann, was soll, was muss beschrieben werden?

Hartig, Katja; Soßna, Volker: DOI: <https://doi.org/10.15488/262>

Zwischenfazit

- Mindestens seit 2002 dynamische Diskussionen gestartet mit Langzeitarchivierung
- Wichtige Keywords: Zeitraum, geeignete Strategien, Prozess ...
- Ziel: Verfügbarkeit (LZA), Nachnutzung (FDM)
- In FDM ist LZA enthalten
- 2019: immer noch keine einheitliche Terminologie und abgestimmte Definitionen
- Seit ca. 2018 international: data steward(ship)

EOSC & FAIR

European Open Science Cloud (EOSC) gelauncht im November 2018,
zwei wichtige Berichte:

[Prompting an EOSC in practice](#)

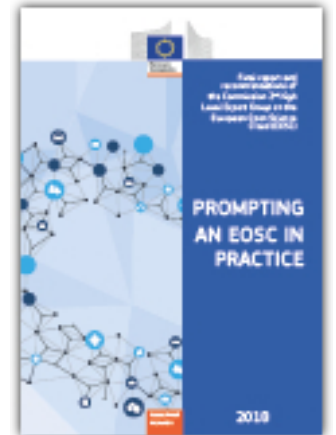
Report of the Commission 2nd High Level Expert Group on the European Open Science Cloud (EOSC 2nd HLEG)

... to establishing the main **Rules of Participation**; also paying attention to issues as **Governance** and possible **Business Models**. The report analyses various aspects of how the EOSC can effectively **interlink People, Data, Services and Training, Publications, Projects and Organisations** and presents a set of detailed practical recommendations ...

[Turning FAIR into reality](#)

Report of the Commission FAIR Data Expert Group (FAIR Data EG)

describes the broad range of changes **required for the implementation of the FAIR data** principles. It offers a survey and analysis of what is needed to implement FAIR and it provides a set of **concrete recommendations** and **actions** for stakeholders in Europe and beyond ...



EOSC & FAIR

ToC-Auszug “*Turning FAIR into reality*”:

- Kap. 1.4: Data science and stewardship skills
- Kap. 1.5: Metrics for FAIR data and assessment frameworks to certify FAIR services
- Kap. 1.6: Sustainable and strategic funding
- Kap. 3.4: Data Management Plans and FAIR
- Kap. 4.3: Data standards, metadata standards, vocabularies and ontologies
- Kap. 4.4: Registries, repositories and certification
- Kap. 5.1: Data science and data stewardship skills for FAIR
- Kap. 5.2: Professionalising roles and curricula
- Kap. 6.1: Metrics / indicators

FAIR

- The FAIR Guiding Principles for **scientific data management and stewardship**
- GO FAIR initiative - practical implementation of EOSC

“... guidelines to improve the **F**indability, **A**ccessibility, **I**nteroperability, and **R**euse of digital assets. The principles emphasise **machine-actionability** (i.e., the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention) because humans increasingly **rely on computational support to deal with data** as a result of the increase in volume, complexity, and creation speed of data.”

2016: Mark D. Wilkinson, Michel Dumontier[...]Barend Mons: *The FAIR Guiding Principles for scientific data management and stewardship*. Scientific Data volume 3, Article number: 160018 sowie <https://www.go-fair.org/>

Data Steward(ship)

“Steps need to be taken to develop two cohorts of professionals to support FAIR data: data scientists embedded in those research projects which need them, and data stewards who will ensure the management and curation of FAIR data.”

Hodson, Simon et al. (June 2018): *Turning FAIR Data into Reality: Interim Report from the European Commission Expert Group on FAIR Data*. <https://doi.org/10.5281/zenodo.1285272>

“New job profiles need to be defined and education programs put in place to train **data professionals** - from **data stewards** to data scientists - who are urgently required to bring FAIR principles to practice and establish the necessary trust and understanding to support data reuse. In order to develop these new professionals, agreed pedagogy and curricula are needed.”

European Commission Expert Group on FAIR Data (Nov 2018): *Turning FAIR into Reality : Final Report and Action Plan from the European Commission Expert Group on FAIR Data*. <https://doi.org/10.2777/1524>

Neue Berufsbilder

Graz University of Technology (Jan 19): Project Researcher in Data Stewardship

https://career.tugraz.at/~career/upload/8bkp4qv/7060_Projektassistent_in%20Data%20Stewardship.pdf

- Conception and development of innovative services and policies for Research Data Management
- Primary research into requirements and tools for machine-actionable data management plans
- Liaison with researchers to assess their data needs and embed research data management into working practices
- Cover all aspects of research data management, including its organisation, description, preservation, and discoverability
- Provide training, guidance and advocacy for research data management and open research

University of Antwerp (Jan 19): Vacancy RDM expert (data steward)

<https://www.uantwerpen.be/en/jobs/vacancies/technical-personnel/2018atpdondex131/>

As expert research data management (or RDM) you will implement the university policy on research data in practice. You will work closely together with the researchers of the UAntwerp and help them to develop a high-quality research data management.

Nationale Forschungsdateninfrastruktur (NFDI)

- Ca. 20 Konsortien „am Start“ (Schätzung)
- RDA-DE Tagung am 19./20. Feb 2019 in Potsdam mit Schwerpunkt NFDI
<https://www.rda-deutschland.de/events/tagung-2019>
- Ausschreibung steht kurz bevor
- Forschungsdaten, -management, -managementpläne sind integraler Bestandteil
- NFDI ist nationaler Part von EOSC
- FAIR ist „verpflichtend“, sicherlich nicht 100% - aber Weg dorthin

Fazit

- FAIR ist/wird **DAS** beherrschende Thema
- Praktische Implementierung der Kriterien v.a. von „re-use“ noch weitestgehend offen
- Großes und breites Themenfeld für Informationswissenschaften und Infrastruktur-Einrichtungen
- Nächste Themen (z.B. BMBF, RfII):
 - Datenqualität
 - Implementierung FAIR, FAIRification Prozess
 - Bewertung FAIR
 - Neue Aus- und Weiterbildungsangebote, Curricula
 - Ethik, Rechtssicherheit
 - ...

29. Jan 2019: Nds. Ministerium für Wissenschaft und Kultur (MWK)

Fördert das Land Niedersachsen in den kommenden Jahren die Einrichtung von **bis zu 50 unbefristeten Digitalisierungsprofessuren** an den staatlichen Universitäten und Fachhochschulen

Zitate von Björn Thümler (Nds. MWK):

- *„Die Nachfrage in Wirtschaft und Verwaltung nach Fach- und Führungskräften mit **informationstechnischen und informationswissenschaftlichen Fachkompetenzen** steigt stetig an – entsprechende attraktive Studienangebote sind unabdingbar, wenn Niedersachsen wettbewerbsfähig bleiben will“.*
- *„Wenn die Konzepte überzeugend sind, ist auch die **Einrichtung neuer Studienangebote** denkbar.“*

<https://idw-online.de/de/news709669>

Aspekte rund um Daten (nicht final): FD & LZA

1. **Daten-Portale/-Qualität:** Wo finde ich die für meine Fragestellung relevanten Daten? Darf ich diese nutzen und weiterverarbeiten? Wie beurteile ich die Qualität von Daten und Datenquellen?
2. **Daten-Corpus:** Wie stelle ich mir einen Daten-Corpus aus heterogenen Quellen zusammen und was muss ich dabei beachten?
3. **Digitale Infrastruktur, Werkzeuge & Dienste:** Welche technologische Infrastruktur unterstützt mich dabei? Was brauche ich alles an digitalen Diensten und Werkzeugen?
4. **Daten-Aufbereitung:** Wie verarbeite ich die Daten mit Hilfe geeigneter Algorithmen und (halb)automatischer Verfahren? Welche (open source) Tools unterstützen mich dabei?
5. **Daten-Analyse:** Wie werte ich die aufbereitete Datenmenge mit geeigneten statistischen und anderen Methoden aus? Welche (open source) Tools unterstützen mich dabei?
6. **Daten-Dokumentation:** Wie dokumentiere ich meine Arbeitsschritte im Sinne einer guten wissenschaftlichen Praxis und im Sinne von “open data”?
7. **Daten-Veröffentlichung/-Visualisierung:** Wie veröffentliche ich meine Daten-Ergebnisse bzw. visualisiere sie in geeigneter Form? Welche Lizenzbedingungen (z.B. Creative Commons) nutze ich dabei?
8. **Daten-Recht/-Ethik:** Welche rechtlichen und ethischen Aspekte muss ich berücksichtigen?
9. **Daten-Speicherung/-Nachnutzung:** Wie archiviere ich die Daten und Ergebnisse so, dass sie nachgenutzt und weiterverwendet werden können?

Torture the data, and it will confess to anything

Ronald Coase, winner of the Nobel Prize in Economics

Herzlichen Glückwunsch zu 10 Jahren MALIS!

