

Datensicherung im Forschungsprozess

Workshop der AG Digitale Romanistik und des FID Romanistik

26.11.-27.11.2020



Einführung:

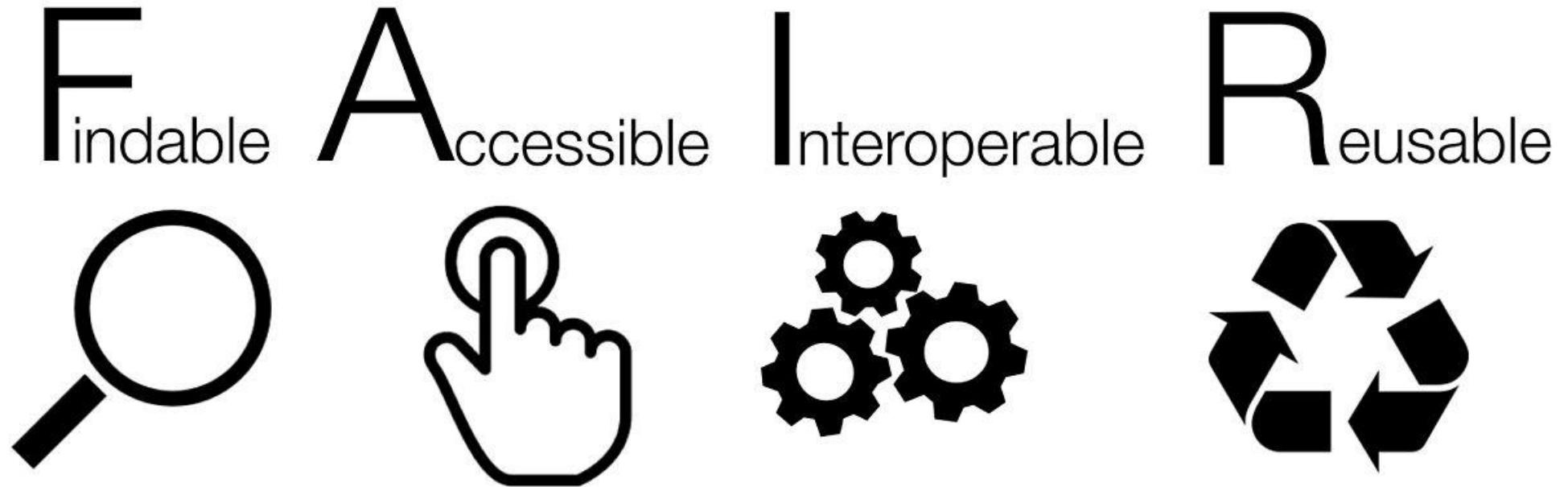
**Die FAIR-Prinzipien, die gute wissenschaftliche
Praxis und die Anforderungen der
Forschungsförderer**

Johannes von Vacano (FID Romanistik / ULB Bonn)

Inhalt

- Die FAIR-Prinzipien
- Die FAIR-Prinzipien & ihre (förder-)institutionelle Verankerung
- Die FAIR-Prinzipien & die gute wissenschaftliche Praxis

Die FAIR-Prinzipien



[SangyaPundir](#), [CC BY-SA 4.0](#), via [Wikimedia Commons](#)

Begriffe: Was sind Forschungsdaten?

- Eine Definition:

»Daten, die im Zuge **wissenschaftlicher Vorhaben** entstehen, z. B. durch Beobachtungen, Experimente, Simulationsrechnungen, Erhebungen, Befragungen, Quellenforschungen, Aufzeichnungen, Digitalisierung, Auswertungen.«

Rat für Informationsinfrastrukturen: Begriffsklärungen (RfII Berichte No. 1),
Göttingen 2016, S.11 – [urn:nbn:de:101:1-201607146410](https://nbn-resolving.org/urn:nbn:de:101:1-201607146410)

Begriffe: Was ist Forschungsdatenmanagement?

- Eine Definition:

»Forschungsdatenmanagement umfasst alle [...] Maßnahmen, die getroffen werden müssen, um qualitätsvolle Daten zu gewinnen, um die **gute wissenschaftliche Praxis** im Datenlebenszyklus **einzuhalten**, um Ergebnisse reproduzierbar und Daten zur Nachnutzung verfügbar zu machen und um ggf. bestehenden Dokumentationsverpflichtungen [...] Rechnung zu tragen.«

Rat für Informationsinfrastrukturen: Begriffsklärungen (RfII Berichte No. 1),
Göttingen 2016, S.11 – [urn:nbn:de:101:1-201607146410](https://nbn-resolving.org/urn:nbn:de:101:1-201607146410)

Begriffe: Was ist Forschungsdatenmanagement?

- Eine Definition:

»Das Forschungsdatenmanagement umfasst alle [...] Maßnahmen, die getroffen werden, um qualitativ hochwertige wissenschaftliche Praxis im Datenlebenszyklus einzuhalten, um Ergebnisse reproduzierbar und Daten zur Nutzung verfügbar zu machen und um ggf. bestehenden Dokumentationsverpflichtungen [...] Rechnung zu tragen.«



Rat für Informationsinfrastrukturen: Begriffsklärungen (RfII Berichte No. 1),
S.11 – [urn:nbn:de:101:1-201607146410](https://nbn-resolving.org/urn:nbn:de:101:1-201607146410)

Quelle: fid-romanistik.de

Begriffe: Was sind Metadaten?

- Eine Definition:

» Metadaten sind **Daten über Daten**. Sie dienen der Beschreibung von Daten und unterstützen Datennutzer bei der Recherche nach Daten, bei der Bewertung der Eignung recherchierter Daten für die eigenen Zwecke und bei der Integration gefundener Daten in die eigene Systemumgebung. [...]«

Rat für Informationsinfrastrukturen: Begriffsklärungen (RfII Berichte No. 1),
Göttingen 2016, S.11 – [urn:nbn:de:101:1-201607146410](https://nbn-resolving.org/urn:nbn:de:101:1-201607146410)

FAIR-Prinzipien, was ist das?

- Vollständiger Titel der ursprünglichen Publikation (2016):

Open Access | Published: 15 March 2016

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, [...] Barend Mons 

Scientific Data **3**, Article number: 160018 (2016) | [Cite this article](#)

155k Accesses | **1974** Citations | **1608** Altmetric | [Metrics](#)

[DOI: 10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18)

- Richtlinien für Forschungsdatenmanagement
- Verbesserung der Nachnutzbarkeit: „... a guideline for those wishing to enhance the reusability of their data holdings.“
- Für Mensch und Maschine lesbar

- FAIR-Prinzipien auf Forschungsdaten.info:

»Ziel ist es, dass Forschungsdaten für Menschen und Maschinen optimal aufbereitet und zugänglich sind. Dies impliziert aber nicht, dass jeder Datensatz uneingeschränkt nachnutzbar ist. Vielmehr zielen die FAIR-Prinzipien darauf, im Rahmen des rechtlich und technisch Möglichen, Datenbestände für neue Nutzungsszenarien zu öffnen. Mit der Anwendung der FAIR-Prinzipien soll die Wiederverwendbarkeit von Datenbeständen verbessert werden.«

<https://www.forschungsdaten.info/themen/veroeffentlichen-und-archivieren/faire-daten/>

FAIR-Prinzipien, was ist das NICHT?

- Kein bloßer Zusatz zu („gutem“) Forschungsdatenmanagement
- Kein Gesetz, kein Standard. → Ein Satz allgemeiner Leitgedanken.
- Kein binäres Etikett: FAIR oder Un-FAIR. → Ein positives Kontinuum: möglichst FAIR, FAIRer als zuvor
- Kein Gütesiegel für die Qualität der Daten. → Ein Maß für Zugänglichkeit sowie Umfang und Art der Beschreibung
- Nicht dasselbe wie Open Data. → Begründete Zugangsbeschränkungen sind vorgesehen. FAIRe Daten dürfen offen sein, müssen es aber nicht.

FAIR – Vier Buchstaben & 15 Prinzipien

AUFFINDBAR



FACILE(S) A TROUVER
RINTRACCIABILI
FÁCILES A ENCONTRAR

F1. (Meta)data are assigned a globally unique and persistent identifier

F2. Data are described with rich metadata (defined by R1 below)

F3. Metadata clearly and explicitly include the identifier of the data they describe

F4. (Meta)data are registered or indexed in a searchable resource

Quelle: www.go-fair.org

FAIR – Vier Buchstaben & 15 Prinzipien

ZUGÄNLICH

A_{ccessible}



ACCESSIBLES
ACCESSIBILI
ACCESSIBLES

A1. (Meta)data are retrievable by their identifier using a standardised communications protocol

A1.1 The protocol is open, free, and universally implementable

A1.2 The protocol allows for an authentication and authorisation procedure, where necessary

A2. Metadata are accessible, even when the data are no longer available

Quelle: www.go-fair.org

FAIR – Vier Buchstaben & 15 Prinzipien

INTEROPERABEL

Interoperable

- I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (Meta)data use vocabularies that follow FAIR principles
- I3. (Meta)data include qualified references to other (meta)data

Quelle: www.go-fair.org



INTEROPÉRABLES
INTEROPERABILI
INTEROPERABLES

FAIR – Vier Buchstaben & 15 Prinzipien

NACHNUTZBAR

R1. Meta(data) are richly described with a plurality of accurate and relevant attributes

R1.1. (Meta)data are released with a clear and accessible data usage license

R1.2. (Meta)data are associated with detailed provenance

R1.3. (Meta)data meet domain-relevant community standards

Quelle: www.go-fair.org



RÉUTILISABLES
RIUTILIZZABILI
REUTILIZABLES

FAIR – Vier Buchstaben & 15 Prinzipien

F_{indable}



FACILE(S) A TROUVER
FACILMENTE
RINTRACCIABILI
FÁCILES A ENCONTRAR

A_{ccessible}



ACCESSIBLES
ACCESSIBILI
ACCESSIBLES

I_{nteroperable}



INTEROPÉRABLES
INTEROPERABILI
INTEROPERABLES

R_{eusable}



RÉUTILISABLES
RIUTILIZZABILI
REUTILIZABLES

Beispiel – Zenodo & FAIR

FAIR Principles

FAIR Principles definition as referenced from: *Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3:160018 doi: 10.1038/sdata.2016.18 (2016).*

To be Findable:

- **F1:** (meta)data are assigned a globally unique and persistent identifier
 - A DOI is issued to every published record on Zenodo.
- **F2:** data are described with rich metadata (defined by R1 below)
 - Zenodo's metadata is compliant with [DataCite's Metadata Schema](#) minimum and recommended terms, with a few additional enrichments.
- **F3:** metadata clearly and explicitly include the identifier of the data it describes
 - The DOI is a top-level and a mandatory field in the metadata of each record.
- **F4:** (meta)data are registered or indexed in a searchable resource
 - Metadata of each record is indexed and searchable directly in Zenodo's search engine immediately after publishing.
 - Metadata of each record is sent to DataCite servers during DOI registration and indexed there.

To be Accessible:

- **A1:** (meta)data are retrievable by their identifier using a standardized communications protocol
 - Metadata for individual records as well as record collections are harvestable using the [OAI-PMH](#) protocol by the record identifier and the collection name.
 - Metadata is also retrievable through the public [REST API](#).
- **A1.1:** the protocol is open, free, and universally implementable
 - See point A1. OAI-PMH and REST are open, free and univesal protocols for information retrieval on the web.
- **A1.2:** the protocol allows for an authentication and authorization procedure, where necessary
 - Metadata are publicly accessible and licensed under public domain. No authorization is ever necessary to retrieve it.
- **A2:** metadata are accessible, even when the data are no longer available
 - Data and metadta will be retained for the lifetime of the repository. This is currently the lifetime of the host laboratory CERN, which currently has an experimental programme defined for the next 20 years at least.
 - Metadata are stored in high-availability database servers at CERN, which are separate to the data itself.

Inhalt

- Die FAIR-Prinzipien
- **Die FAIR-Prinzipien & ihre (förder-)institutionelle Verankerung**
- Die FAIR-Prinzipien & die gute wissenschaftliche Praxis

FAIRRankerung in der Wissenschaftspolitik

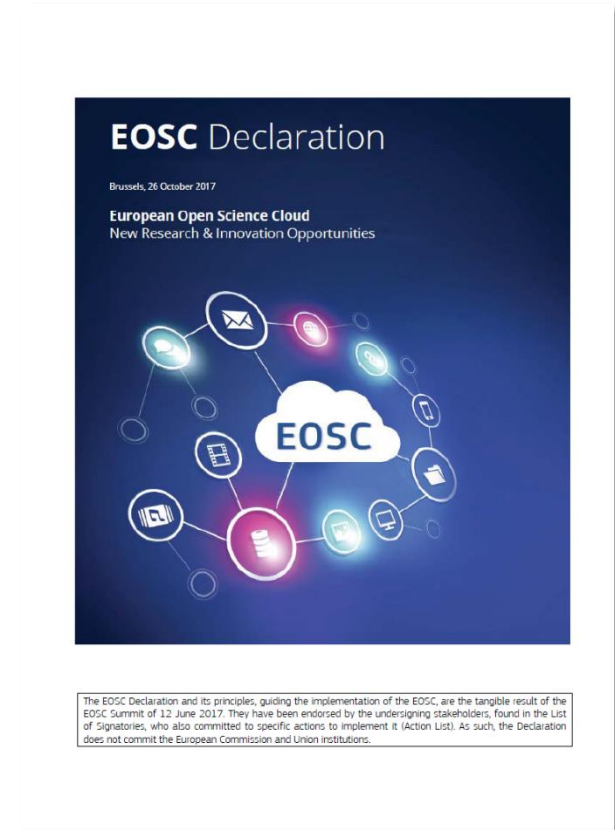
- Auf europäischer Ebene bereits 2016/2017:
 - Horizont 2020 – *FAIR Data Management Plan*
 - »This document helps Horizon 2020 beneficiaries make their research data findable, accessible, interoperable and reusable (FAIR), **to ensure it is soundly managed**. Good research data management is not a goal in itself, but rather the key conduit leading to knowledge discovery and innovation, and to **subsequent data and knowledge integration and reuse**.«

[EUROPEAN COMMISSION H2020 Programme Guidelines on FAIR Data Management in Horizon 2020](#), V3.0, 2016, S. 3.



FAIRRankerung in der Wissenschaftspolitik

- European Open Science Cloud (EOSC)
 - Aufbau einer föderierten Struktur bestehender Datenzentren und Institutionen in Europa
 - Erster Abschnitt der *Declaration: Data culture and FAIR data*
»Implementation of the FAIR principles must be pragmatic and technology-neutral, encompassing all four dimensions: findability, accessibility, interoperability and reusability. [...] The disciplinary sectors must develop their specific notions of FAIR data in a coordinated fashion and determine the desired level of FAIRness.«



FAIRankerung in der Wissenschaftspolitik



www.go-fair.org

- European Open Science Cloud (EOSC)
 - Aufbau einer föderierten Struktur bestehender Datenzentren und Institutionen in Europa
 - Erster Abschnitt der *Declaration: Data culture and FAIR data*
»Implementation of the FAIR principles must be pragmatic and technology-neutral, encompassing all four dimensions: findability, accessibility, interoperability and reusability. [...] The disciplinary sectors must develop their specific notions of FAIR data in a coordinated fashion and determine the desired level of FAIR-ness.«



The EOSC Declaration and its principles, guiding the implementation of the EOSC, are the tangible result of the EOSC Summit of 12 June 2017. They have been endorsed by the undersigning stakeholders, found in the List of Signatories, who also committed to specific actions to implement it (Action List). As such, the Declaration does not commit the European Commission and Union institutions.

[EOSC Declaration](#), 2017, S. 1.

FAIRRankerung in der Wissenschaftspolitik

- Einrichtung einer H2020-[FAIR Data Expertengruppe](#) der Europäischen Kommission
 - Abschlussbericht Turning FAIR into Reality, 2018 - [doi: 10.2777/1524](#)
- [EOSC FAIR Working Group](#) (1 von 6 WG)
 - Aktuellste Publikation: Six Recommendations for implementation of FAIR practice, Oktober 2020 - [doi: 10.2777/986252](#)
- [FAIRsFAIR.eu](#) – Fostering FAIR data practices in Europe (Schwerpunkt Infrastrukturen)



FAIRRankerung in der Wissenschaftspolitik

- EU-Rahmenprogramm: Horizont Europa (ab 2021)
- Nachfolgeprogramm zu Horizont 2020



Open Science across the programme

Open Science

Better dissemination and exploitation of R&I results and support to active engagement of society

Mandatory Open Access to publications: beneficiaries shall ensure that they or the authors retain sufficient intellectual property rights to comply with open access requirements

Open Access to research data ensured: in line with the principle "as open as possible, as closed as necessary"; Mandatory Data Management Plan for FAIR (Findable, Accessible, Interoperable, Re-usable) and Open Research Data



European Commission (2019): Horizon Europe - THE NEXT EU RESEARCH & INNOVATION INVESTMENT PROGRAMME (2021 - 2027)
https://ec.europa.eu/info/sites/info/files/research_and_innovation/strategy_on_research_and_innovation/presentations/horizon_europe_en_investing_to_shape_our_future.pdf

FAIRRankerung in der Wissenschaftspolitik

- EU-Rahmenprogramm: Horizont Europa (ab 2021)
- Nachfolgeprogramm zu Horizont 2020



Open Science across the programme



Open science across the programme

Art. 10 - Open Science The approach	Art. 35 - Exploitation and Dissemination The modalities
Open access to scientific publications to be ensured Open access to research data to be ensured in line with principle 'as open as possible, as closed as necessary' Responsible research data management to be ensured in line with FAIR principles Other open science practices to be promoted and encouraged Reciprocity in open science to be promoted and encouraged in all association and cooperation agreements with third countries	Open access to scientific publications obligatory: sufficient IPR to be retained Open access to research data, 'as open as possible, as closed as necessary': exceptions Responsible research data management in line with FAIR principles; Data Management Plan mandatory; possible obligations (in specific WPs) to use the European Open Science Cloud for storing and providing access to data Possible additional incentives or obligations through work programmes for other open science practices



Horizon Europe

THE NEXT EU RESEARCH & INNOVATION INVESTMENT PROGRAMME (2021 – 2027)

#HorizonEU

Based on the Commission Proposal for Horizon Europe, the common understanding between co-legislators and the Partial General Approach, both approved in April 2019

Research and Innovation

European Commission (2019): Horizon Europe - THE NEXT EU RESEARCH & INNOVATION INVESTMENT PROGRAMME (2021 - 2027)
https://ec.europa.eu/info/sites/info/files/research_and_innovation/strategy_on_research_and_innovation/presentations/horizon_europe_en_investing_to_shape_our_future.pdf

FAIRankerung in der Wissenschaftspolitik

- Sorbonne-Erklärung, 27. Januar 2020
 - »In der Sorbonne-Erklärung plädieren die Verbände dafür, Forschungsdaten so offen wie möglich zugänglich und wiederverwertbar zu machen. [...] Für die Daten soll das international etablierte Prinzip FAIR angewendet werden, demzufolge Daten **f**indable, **a**ccessible, **i**nteroperable und **r**eusable (auffindbar, zugänglich, interoperabel und wiederverwendbar) sein sollen.«

[Pressemeldung German U15](#), 30.1.2020



FAIRRankerung auf nationaler Ebene

- [Aktionsplan Forschungsdaten](#) des BMBF (Oktober 2020)
- Bündelung von Aktionen und Maßnahmen
 - u.a. NFDI, EOSC-Teilnahme



5. Oktober 2020

BMBF-AKTIONSPLAN FORSCHUNGSDATEN

IMPULSE FÜR EINE KULTUR DER DATENBEREITSTELLUNG UND
WEITERVERWENDUNG IN BILDUNG, WISSENSCHAFT UND FORSCHUNG

- e) Zur Steigerung der Datenqualität sollen die FAIR-Prinzipien („Findable, Accessible, Interoperable, Re-usable“) bei der Datenerhebung, -sammlung, -nutzung und -aufbewahrung stärker berücksichtigt werden. Im Rahmen der Projektförderung von Datenerhebungen und -sammlungen sollen die Zuwendungsempfänger grundsätzlich verpflichtet werden, ihre im Projekt erhobenen Daten systematisch, möglichst standardbasiert zu verarbeiten und aufzubewahren.
- f) Der Kulturwandel hin zu einem qualitätsbewussten Umgang mit Forschungsdaten soll auch durch die Anerkennung von FAIR-Datenpublikationen als Publikationsleistung gestärkt werden. Datenpublikationen nach den FAIR-Standards sollen als wissenschaftliche Leistung beispielsweise bei der Begutachtung von Förderanträgen berücksichtigt werden.

chlüssel für unsere Zukunft. Die
ft und Forschung ist eine zentrale
im für Bildung und Forschung geht
ngsdaten“ systematisch an. Der
rei Zielen:

[Faktenblatt Aktionsplan Forschungsdaten](#), BMBF, 2020

Inhalt

- Die FAIR-Prinzipien
- Die FAIR-Prinzipien & ihre (förder-)institutionelle Verankerung
- **Die FAIR-Prinzipien & die gute wissenschaftliche Praxis**

FAIR, DFG & die gute wissenschaftliche Praxis

- „[Denkschrift Sicherung guter wissenschaftlicher Praxis](#)“ (1998, ergänzte Fassung 2013) – pre-FAIR.
- „[Leitlinien zum Umgang mit Forschungsdaten](#)“ (2015) ebenfalls pre-FAIR, aber bereits analoge Konzepte
- Fachspezifische *Empfehlungen*:
 - „[Förderkriterien für wissenschaftliche Editionen in der Literaturwissenschaft](#)“ (2015): enthalten verwandte Impulse
 - „[Empfehlungen zu datentechnischen Standards und Tools bei der Erhebung von Sprachkorpora](#)“ (aktualisiert 2019): FAIR im Punkt 2.4 - Nachhaltigkeit, Zitierbarkeit, Nachnutzbarkeit und Langzeitarchivierung

FAIR, DFG & die gute wissenschaftliche Praxis

- 19 Leitlinien
- 11 davon zum wissenschaftlichen Arbeitsprozess
- 7 behandeln den Umgang mit Forschungsdaten
 - FAIR in Leitlinie 13:
Herstellung von öffentlichem Zugang zu Forschungsergebnissen
»Aus Gründen der Nachvollziehbarkeit, Anschlussfähigkeit der Forschung und Nachnutzbarkeit hinterlegen Wissenschaftlerinnen und Wissenschaftler, wann immer möglich, die der Publikation zugrunde liegenden Forschungsdaten und zentralen Materialien – den **FAIR-Prinzipien** („Findable, Accessible, Interoperable, Re-Usable“) folgend – zugänglich in anerkannten Archiven und Repositorien.«



Deutsche Forschungsgemeinschaft, Leitlinien zur Sicherung guter wissenschaftlicher Praxis. Kodex, 2019, doi: [10.5281/zenodo.3923602](https://doi.org/10.5281/zenodo.3923602)

FAIR, DFG & die gute wissenschaftliche Praxis

- 19 Leitlinien
- 11 davon zum wissenschaftlichen Arbeitsprozess
- 7 behandeln den Umgang mit Forschungsdaten:
 - Leitlinie 7: **Phasenübergreifende Qualitätssicherung** (R1, R1.3, R1.2)
 - Leitlinie 10: **Rechtliche und ethische Rahmenbedingungen, Nutzungsrechte** (A1.1, R1.1)
 - Leitlinie 11: **Methoden und Standards** (R1.3)
 - Leitlinie 12: **Dokumentation** (F2, A2, R1.2, R1.3)
 - Leitlinie 13: **Herstellung von öffentlichem Zugang zu Forschungsergebnissen** (A1.1, F2, F4, R1.2)
 - Leitlinie 15: **Publikationsorgan** (F4, R1.3)
 - Leitlinie 17: **Archivierung** (F4, R1.3)



[doi: 10.5281/zenodo.3923602](https://doi.org/10.5281/zenodo.3923602)

FAIR, DFG & die gute wissenschaftliche Praxis

- 19 Leitlinien

- 3. Ebene

wissenschaftliche-integritaet.de

»Die dritte Ebene wird „online“ als dynamisches Dokument auf der Webseite der DFG zur Verfügung gestellt werden. Sie wird **fachspezifische Ausführungen, Fallbeispiele und Frequently Asked Questions** enthalten und ab Herbst 2019 detailliert konzipiert werden. Die Inhalte der dritten Ebene sollen im Zusammenwirken mit den Hochschulen und außerhochschulischen Einrichtungen, den wissenschaftsorganisationen, dem Gremium „Ombudsman für die Wissenschaft“ und weiteren Akteuren kontinuierlich erarbeitet und qualitätsgesichert sowie entsprechend den Veränderungen in den wissenschaftlichen Praxen angepasst werden. Damit soll ein aktuelles Referenzwerk für die deutsche Wissenschaftslandschaft geschaffen werden.«

(Vorwort, S. 5)



[doi: 10.5281/zenodo.3923602](https://doi.org/10.5281/zenodo.3923602)

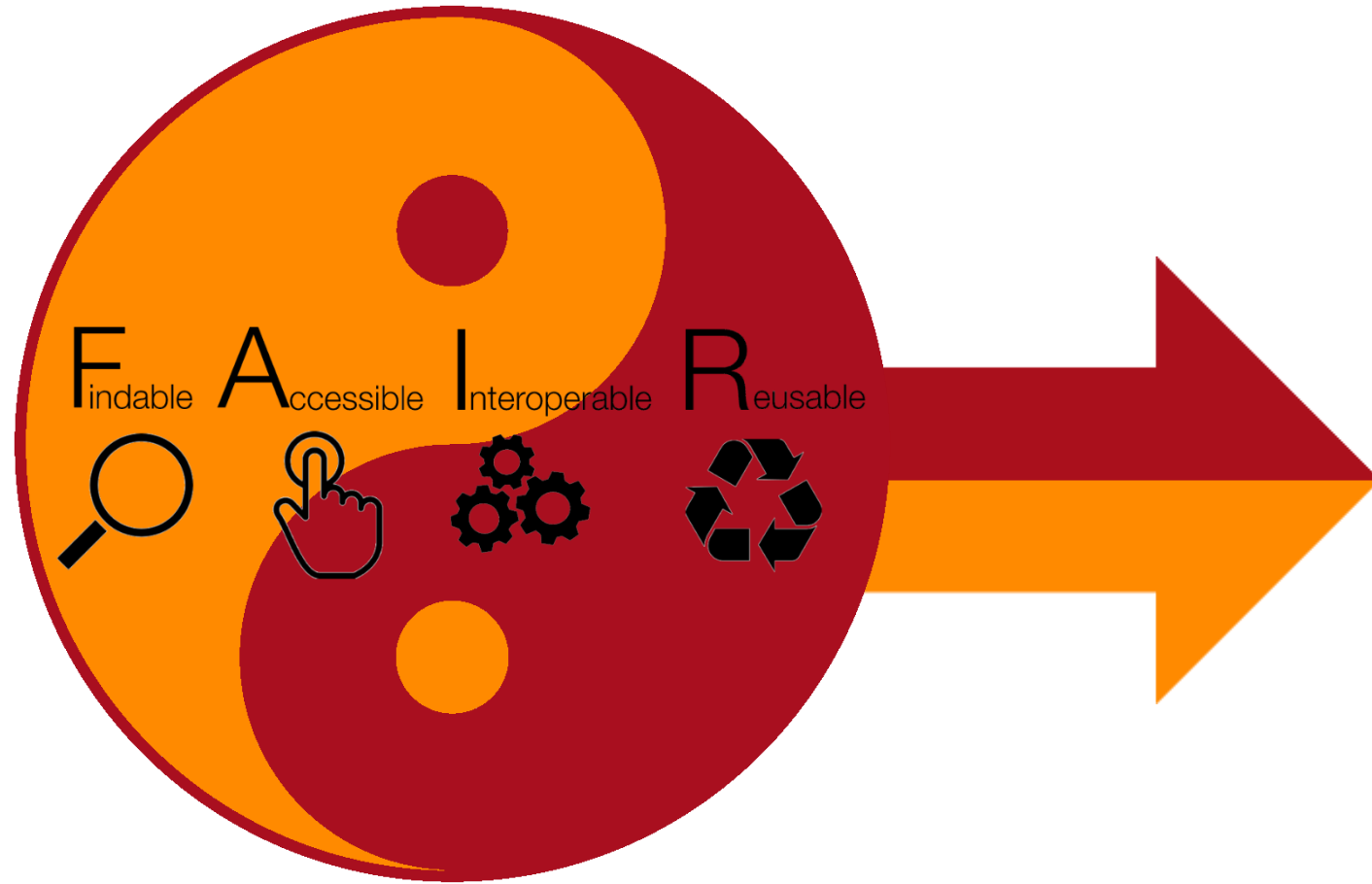
FAIR – in der Praxis?

Vorgaben



Romanistische
Forschungspraxis

Vorgaben



Romanistische Forschungspraxis

Vorgaben

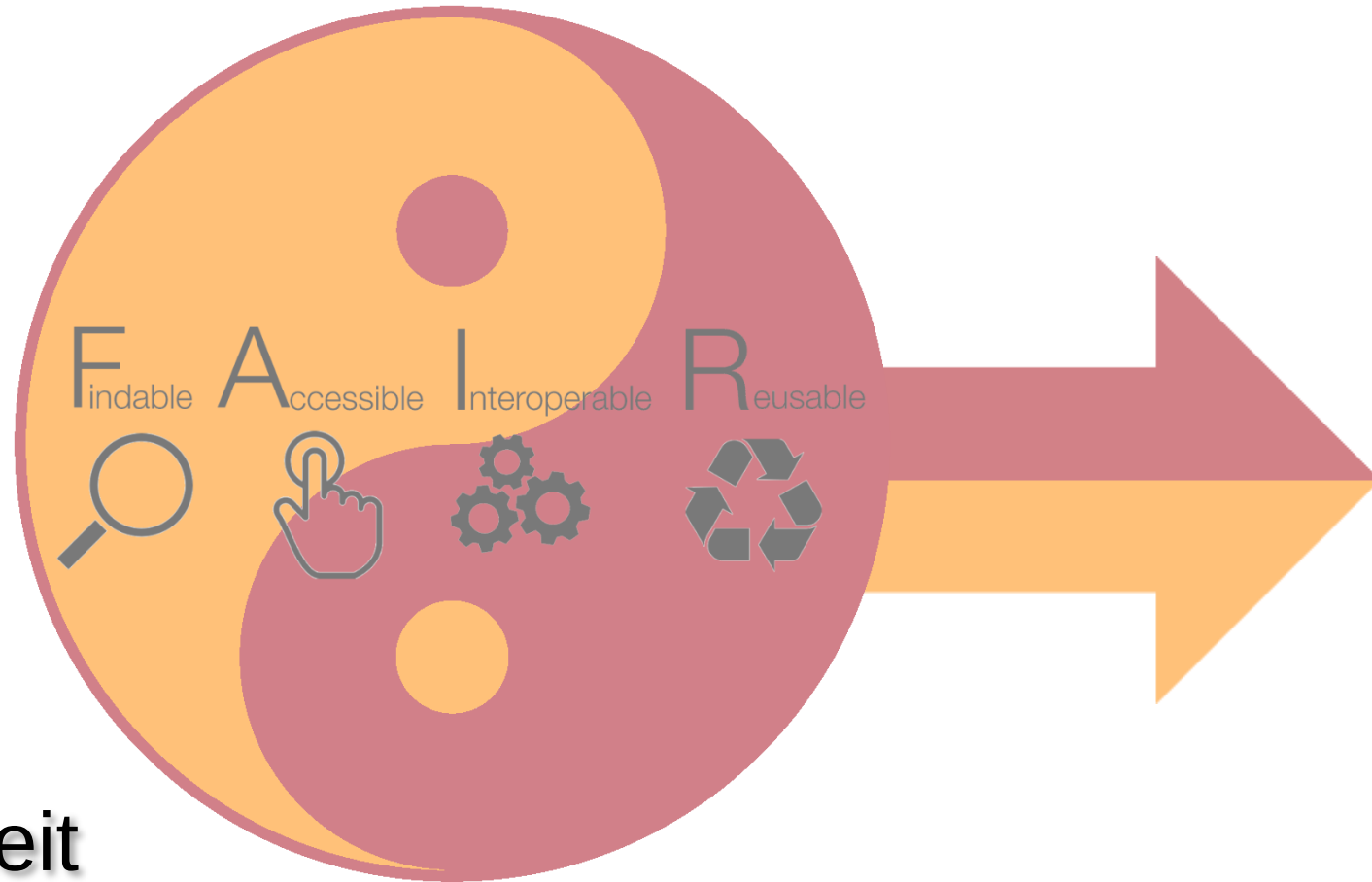
F ielen

D A nk

für

Ihre

Aufme R ksamkeit



Romanistische Forschungspraxis