

FAIR – in aller Kürze unter die Lupe genommen

Daniel Martini

Kuratorium für Technik und Bauwesen in der Landwirtschaft e.V. (KTBL)

20. Juni 2022



Was bringen FAIR und die NFDI?



FAIRagro

Als Datenwissenschaftler: Wieviel Zeit verbringen Sie damit...

- ...Dokumentationen zu Datensätzen und APIs zu lesen?
- ...Datenstrukturen fremder Datensätze zu verstehen?
- ...Programmcode für Extract-Transform-Load Prozesse zu schreiben?

Als Agrar- und/oder Naturwissenschaftler: Wieviel Zeit verbringen Sie damit...

- ...Vergleichsdatsätze für Ihre Mess-/Versuchs-/sonstige Daten zu suchen?
- ...individuelle Anfragen zu formulieren und Abstimmungen zu organisieren um an Daten zu kommen?
- ...Datensätze zu beschaffen und initial zu analysieren, nur um festzustellen, dass sie für Ihre Fragestellung nicht geeignet sind?

Wie wäre es, wenn man diese Aufwände um Faktor 10 eindampfen könnte?

Findable

- 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

Accessible

- 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

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

Reusable

- 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

<https://go-fair.org/fair-principles>

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. (2016): The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018. <https://doi.org/10.1038/sdata.2016.18>

Was wollten uns die Autoren *wirklich* sagen?

- Begriffe und Definitionen
- Zusammenhänge zwischen den FAIR-Prinzipien
- Erläuterungen

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

Ein Identifier: FR0124175367

Valide mit Blick auf F1?



CC BY-SA 3.0,
<https://commons.wikimedia.org/w/index.php?curid=432414>

ISO 6166:2021

Financial services – International securities identification number (ISIN):
UNEDIC 17/22 MTN Anleihe



 in F1 steht *nicht* “data sets are assigned a globally unique and persistent identifier” sondern “(meta)data”

PIDs, DOIs, URLs, URNs, UUIDs...: Was ergibt in welchem Kontext Sinn?

Wie machen wir existierende Identifiersysteme konform zu F1?

FAIR: Breaking it down...by example: Definitionen

11. (Meta)data use a formal, accessible, shared, and broadly applicable **language for knowledge representation.**

Ziel:

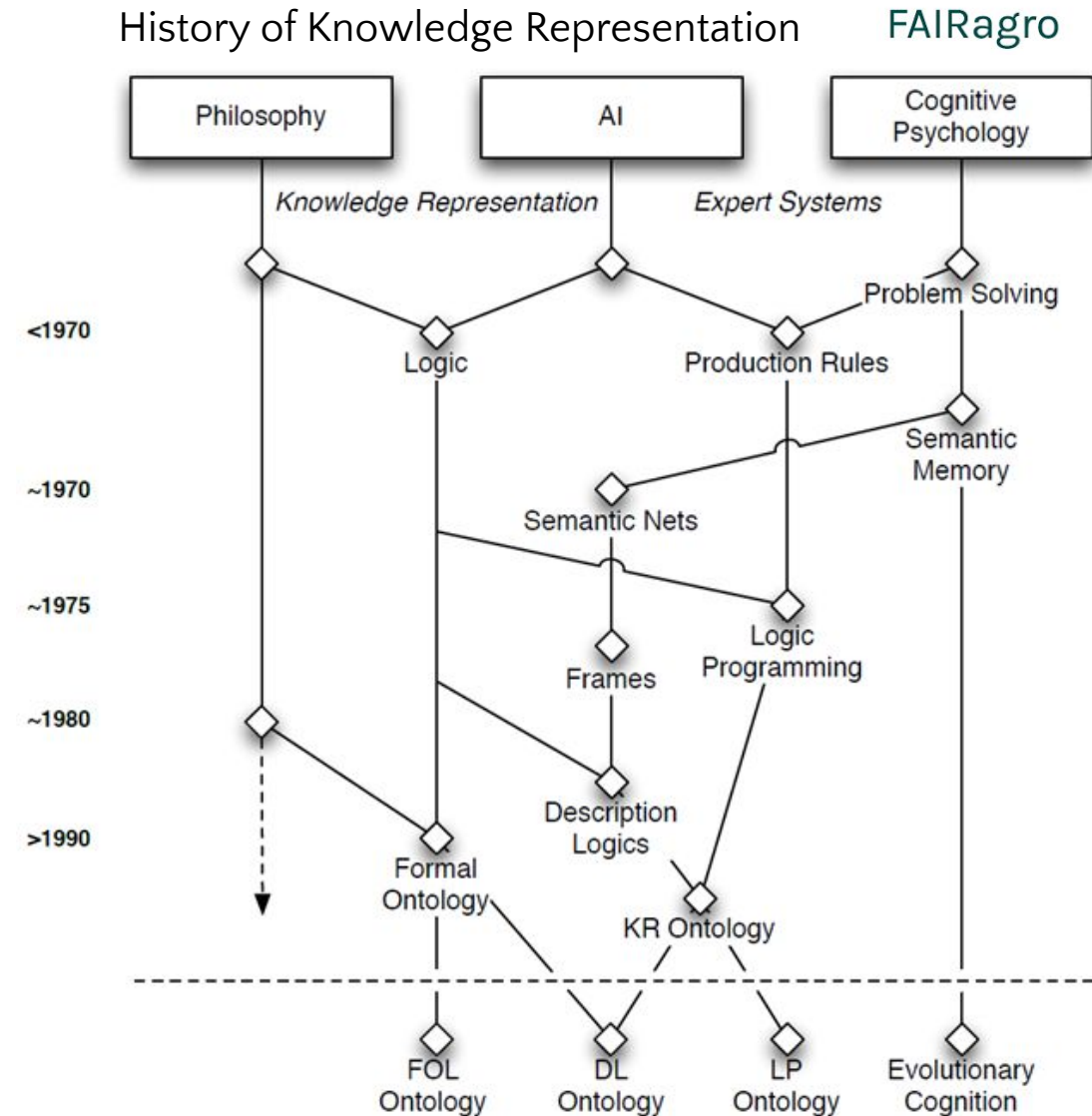
Daten-Integration, Ableiten neuen Wissens (Inferenz):
machine-actionability

Wissensrepräsentationsprachen:

- basieren auf "Aussagen" → erweiterbar
- rekursiv-selbstreflektiv
- im Kontext Forschungsdatenmanagement: "open world assumption"
- Negation, Konjunktion, Disjunktion
- ...



FAIRagro



R. J. Hoekstra (2009): *Ontology Representation : design patterns and ontologies that make sense.* Leibniz Centre for Law. IOS Press Amsterdam. <https://hdl.handle.net/11245/1.317612>



FAIR: Breaking it down...by example: Zusammenhänge

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

“Dereferenzierbarkeit”

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

I2. (Meta)data use vocabularies that follow FAIR principles

- Rekursivität: alle FAIR-Prinzipien sollen auch für Metadatenvokabularien gelten
- Metadatenvokabular = “Wortschatz” aus dem sich Metadaten zusammensetzen: Objekte, Datenfelder und Wertebereiche
- “Einträge” des Vokabulars = data:
F1. (Meta)data are assigned a globally unique and persistent identifier
⇒ “Einträge” des Vokabulars are assigned a globally unique and persistent identifier.



machine-readable \neq machine-actionable

"we wish to draw a distinction between data that is machine-actionable as a result of specific investment in software supporting that data-type [...] and data that is machine-actionable exclusively through the utilization of general-purpose, open technologies."

- Existierende Standards
 - Identifier, Protokolle, Formate, Vokabularien, Thesauri, Ontologien
 - MIAPPE, AGROVOC, schema.org, ISO19115...
- Welche sind “FAIR-by-Design”?
 - kritische Betrachtung: FAIR-Metriken
- Wo müssen wir anpassen?
 - Identifier-Sourcing/Minting, Policies
 - Formate
 - Alignmentfähigkeit und Alignment
- Wenn Anpassungen notwendig sind, was muss genau getan werden?
 - Mechanismen

Analyse und Maßnahmenentwicklung im Dialog mit:

- Community
- andere NFDI-Konsortien und -Partner
- internationale und nationale Initiativen

“Schönheit liegt im Auge des Betrachters”

FAIRagro: “fitness-for-use”

Kriterien, Metriken

räumliche/zeitliche Auflösung

Vollständigkeit

Genauigkeit

...

Metadaten

Herkunft

Prozessierung

Nutzungen

...

Werkzeuge

Datenmatrix

Kuration

Nutzerrückmeldung

...

FAIR ≠ open, free

A1.2 The protocol allows for an authentication and authorisation procedure, where necessary
R1.1 (Meta)data are released with a clear and accessible data usage license

- Welche Daten sind per-se lizenzfrei?
- Welche Daten benötigen Lizenzen?
- Entspricht die Lizenz der Intention?
- Welche Rahmenbedingungen gelten: Personenbezug, Datenschutz?
- Für welche Zwecke dürfen Daten genutzt werden?
- ...

- Grundsätze und Leitlinien
- Modellvereinbarungen
- Aktuelle Informationen zu Regularien (Urheberrecht, Datenschutz...)
- machine-actionable legal metadata: z. B. Open Digital Rights Language (ODRL)
- ...

Notwendige Bedingungen für einen Paradigmenwechsel

