

METHODS

How to translate the NFDI generic approach into praxis? FAIR Digital Objects and Canonical Workflows

Claudia Biniossek

Dirk Betz

Virtuelle NFDI-Konferenz der Steuergruppe der Arbeitsgemeinschaft RDMO:
Erstellung von Datenmanagementplänen und Einsatz von DMP Tools in der NFDI

Tuesday 2 Mar 2021, 09:00 → 15:00

METHODS

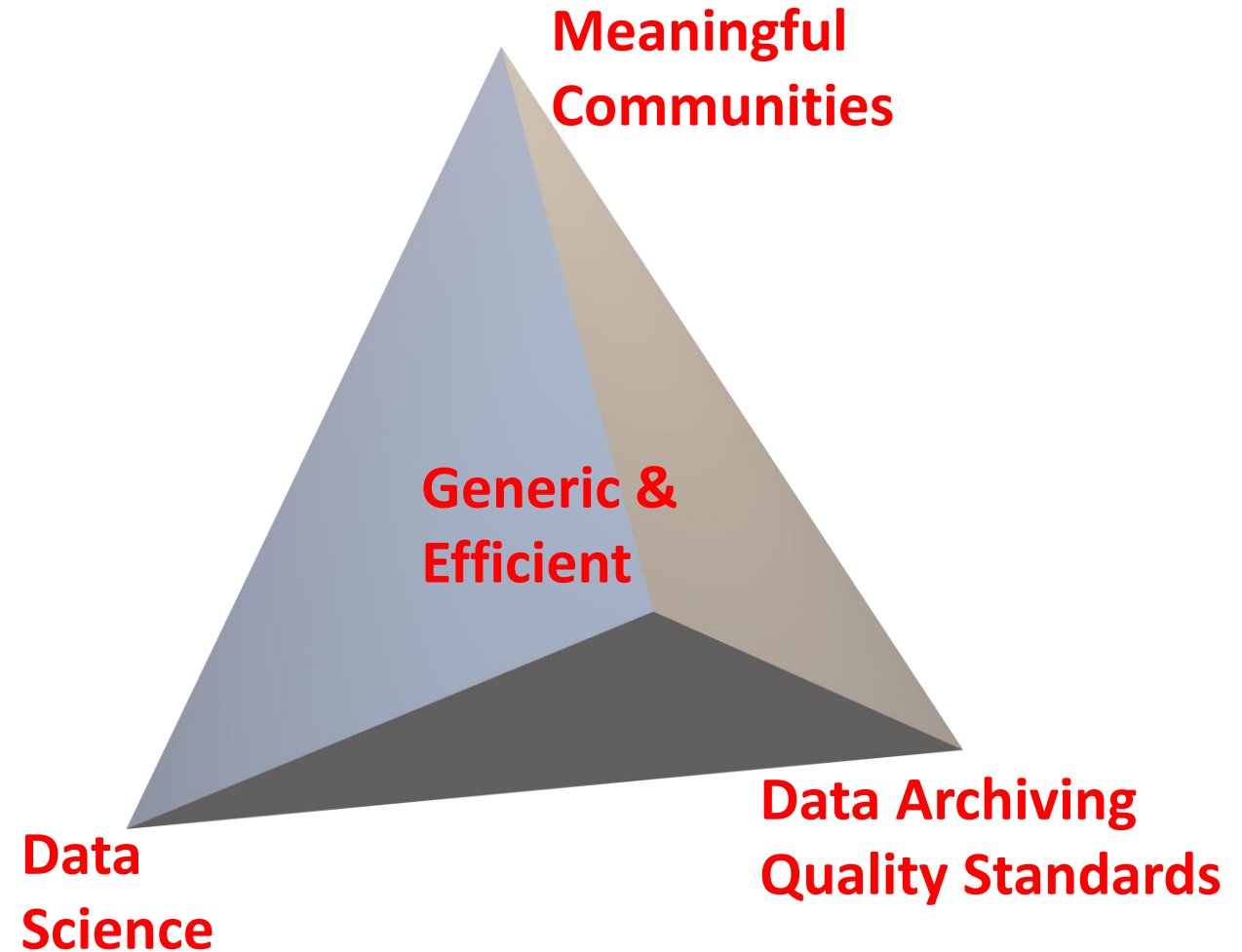
Methodology-oriented Consortium for Generic Research Data Infrastructures for Multiple Disciplines

Motivation


- **NFDI Objective:** Comprehensive coverage of researchers and digital objects
- **Challenge:** NFDI is organized mainly through disciplines
 - Different methods exist within the same discipline
 - Same methods are used within different disciplines
- **METHODS concept:** Identification of “meaningful communities” that share a structural similarity in methodology
 - Therefore, they use similar metadata, tools, DMPs, etc.

Core-principles of METHODS

- Involving all relevant actors
 - Meaningful communities
 - Data science & toolmaker
 - Data archives & librarians
- Local bottom-up approach (instead of top-down)
- Generic approach, mainly based on existing solutions → Efficiency




Data Management in METHODS: Heterogeneous



A FREE AND OPEN SOURCE PROJECT MANAGEMENT TOOL THAT SUPPORTS RESEARCHERS THROUGHOUT THEIR ENTIRE PROJECT LIFECYCLE

The Open Science Framework (OSF) is a scholarly commons developed and maintained by the Center for Open Science and used for documentation, files, collaboration, and connecting to services for artifacts of research activity.

Nearly 200,000 users across disciplines use OSF to manage their research.



Year	Number of Registered OSF Users
2012	371
2013	2,322
2014	6,713
2015	15,216
2016	38,151
2017	73,801
2018	120,256
2019	196,581



Yale Application for Research Data

Yale Institution for Social & Policy Studies

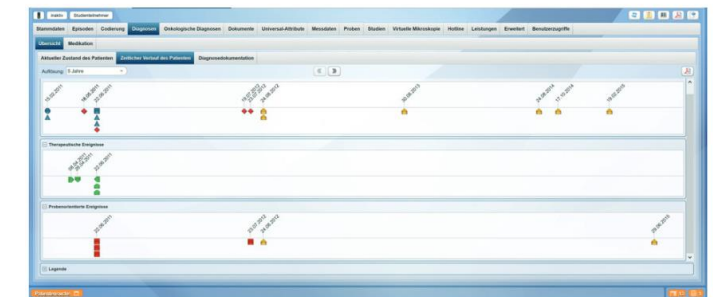


KAIROS
NOW IS THE TIME.
CENTRAXX®

Forschungsakte

Die EPA (Elektronische Patienten Akte) ist die zentrale Speichereinheit in CentraXX. Alle Probeninformationen, die in CentraXX Bio und alle Studiendaten, die in CentraXX Trial zu registrierten CentraXX-Patienten/-Probanden gesammelt werden, werden eingebettet in die EPA. Somit bietet CentraXX die Erfassung longitudinaler Daten eines Patienten und beschränkt sich nicht lediglich auf Momentaufnahmen. Neben Proben- und Studien-

informationen eines Individuums werden weitere strukturierte Daten aus Diagnosen (z. B. ICD10), Prozeduren (z. B. OPS), Episoden (z. B. stationäre Aufenthalte), Medikation, Einverständniserklärungen, virtuelle Mikroskopie, Hotline, in derselben Akte zusammengeführt. Desweiteren können Dokumente (z. B. unterschriebene Einverständniserklärungen) aller Art in die Forschungsakte hochgeladen werden.



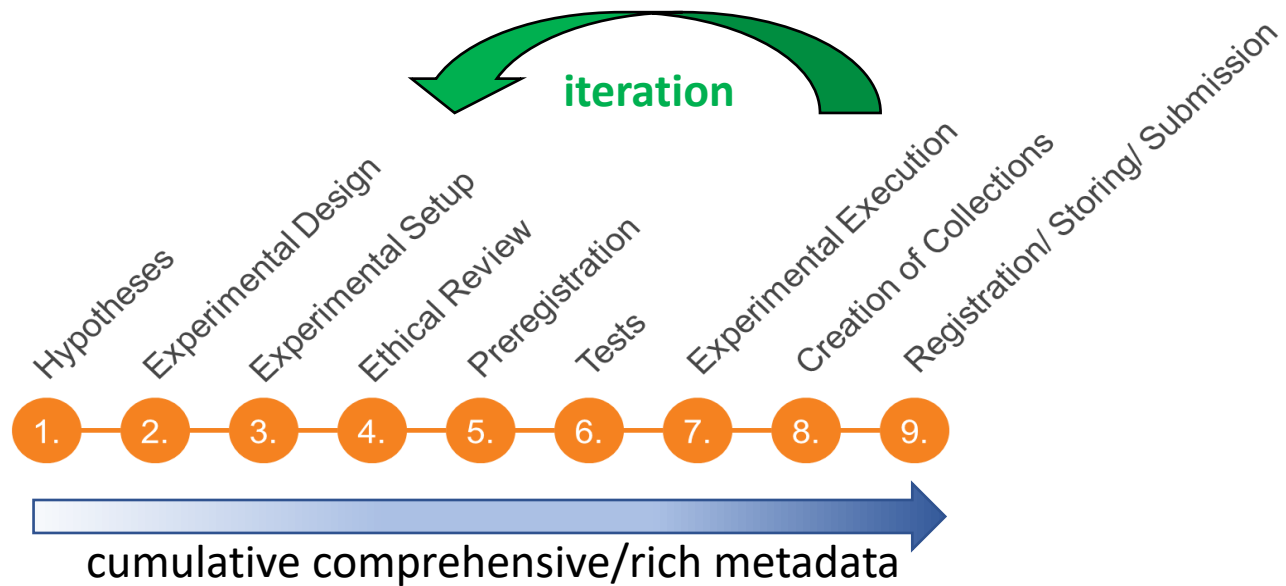
Der zeitliche Verlauf eines Patienten

How to bring together our core principles and existing excellent tools?

- Many excellent tools out there integrating some workflow steps
 - Data generating workflows, e.g., experimental workflows
 - Data management workflows
 - Data analysis workflows
 - Data curation and archiving workflows
 - (Transdisciplinary) data reuse workflows
- Heterogeneous communities and tools, but homogeneous canonical workflows and overarching standards

CWFR Analysis: Experiments

- Identify sufficiently similar coherent and recurring patterns
- Identify recurring canonical actions

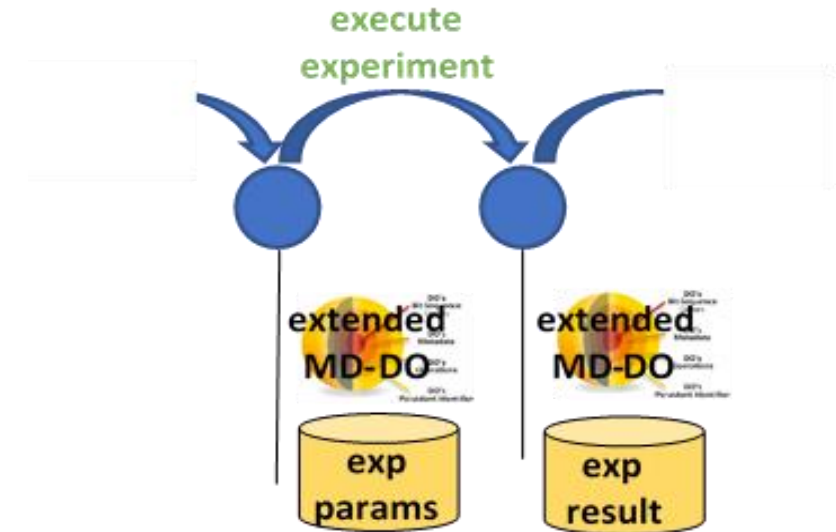
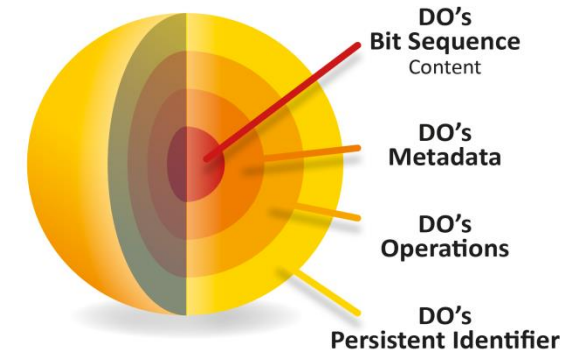


Are the canonical steps all the same across institutes/disciplines? NO

- In physics you don't need an ethical review (normally) – so skip it
- In medicine the regulations for ethics are stricter than in psycholinguistics

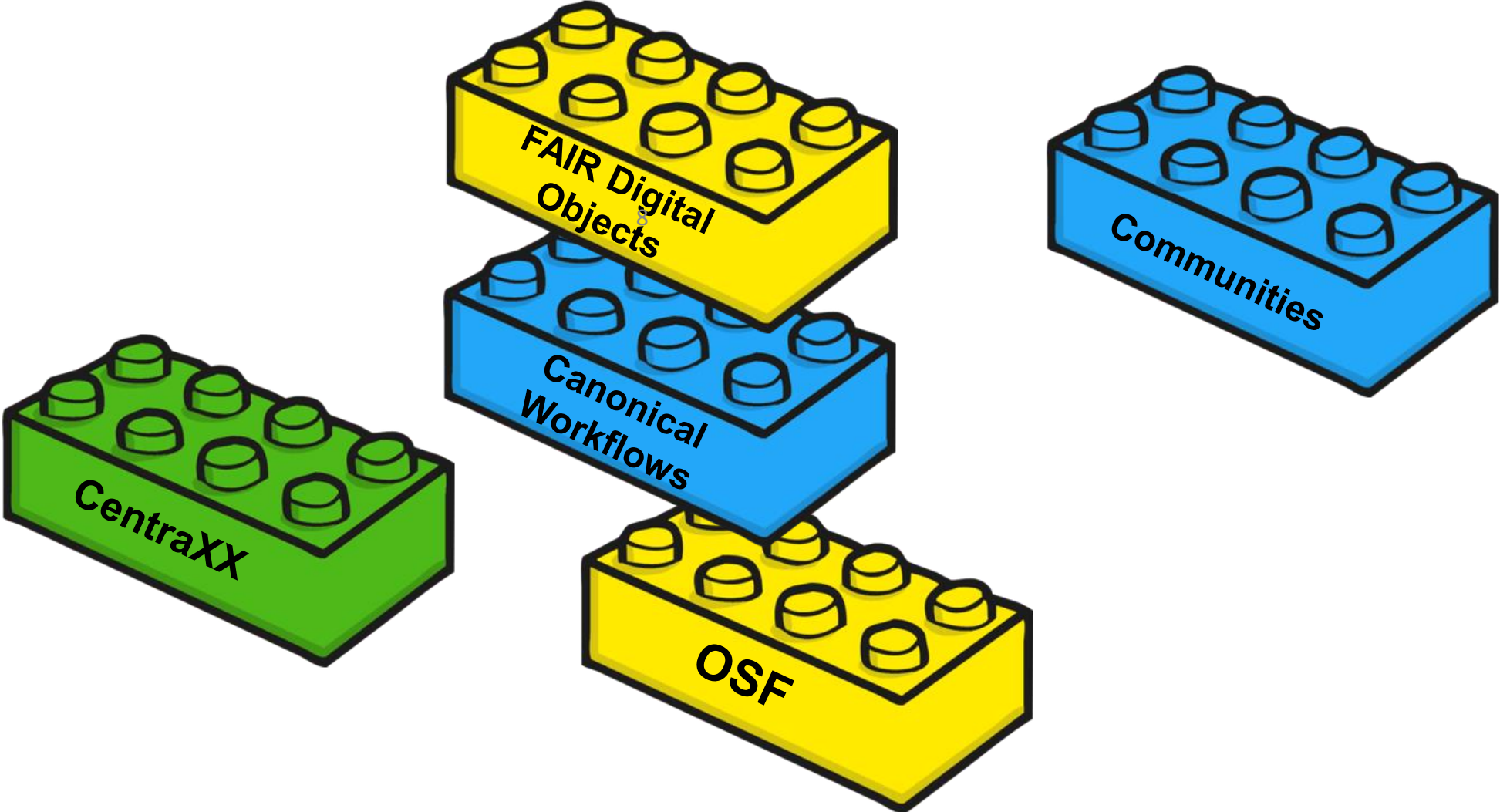
CWFR: FDOs are the glue*

- Create an FDO after each action including all relevant information & references
- Iterations within a project can easily be done
- Exchange between frameworks can also be done via standardized FDO – thus not bound to specific WF technology

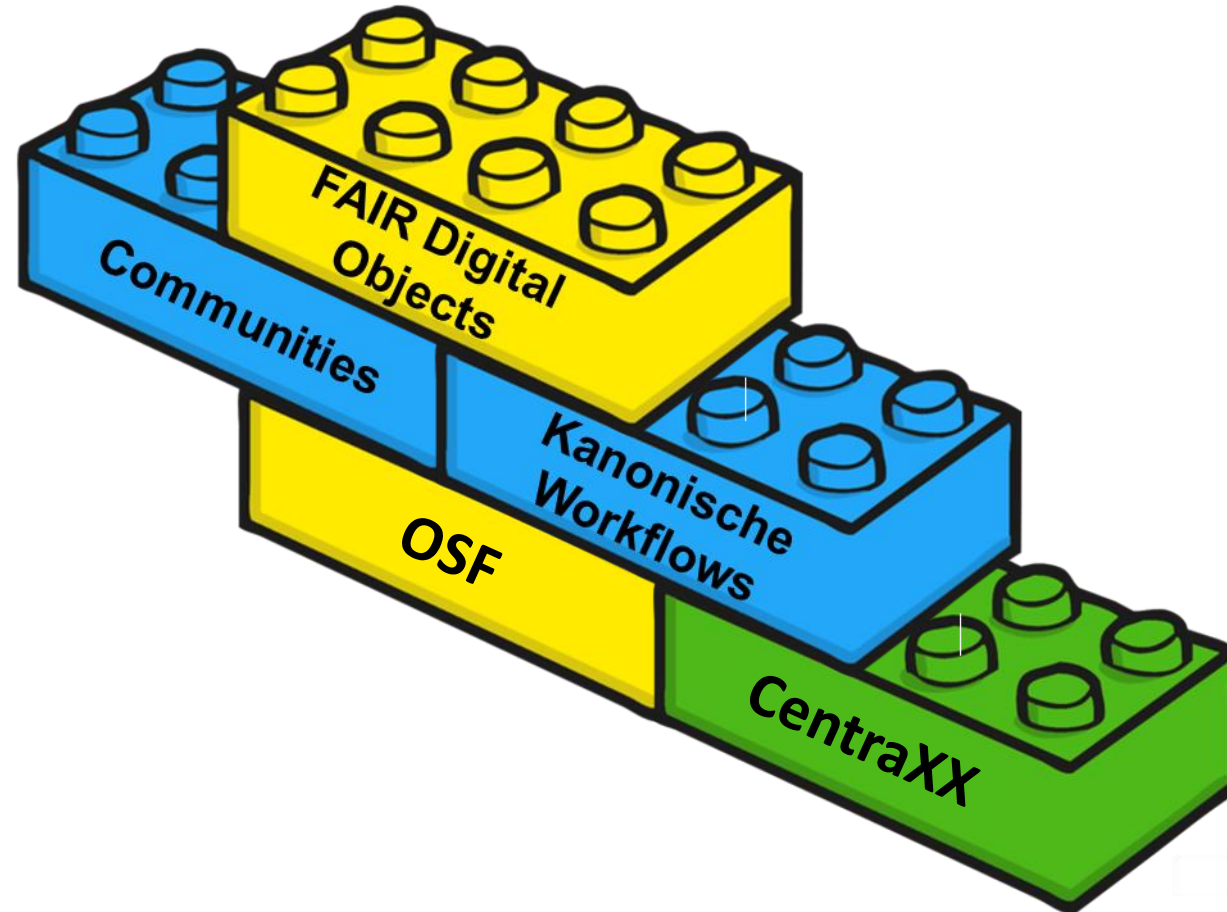


*Jeffery, K., Wittenburg, P., Lannom, L., Strawn, G., Biniossek, C., Betz, D., & Blanchi, C. (2021). Not Ready for Convergence in Data Infrastructures. Data Intelligence 2021, 3:1, 116-135.

METHODS Modular System

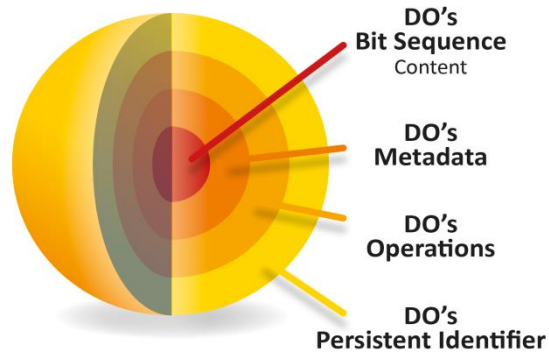


METHODS Modular System



THANK YOU

FDO in a Nutshell



- DOs
 - have bit-sequence encoding content
 - have associated rich metadata
 - have assigned a PID (GUPRI)
- PIDs
 - are persistently resolving to „state“ information based on typed attributes
 - including items such as Type, CKSM, Version, etc. and many references to other FDO such as MD, Rights, BC, etc.
- Types
 - registered in open registries
 - allowing to define operations
- FAIR DO
 - attributes are machine actionable
 - metadata is community defined – not yet machine actionable