

OGD und ORD im Vergleich mit dem Fokus auf Projekte am SDSC

OGD Forum, 21. Oktober 2024

- **SDSC**: kurze Vorstellung
- Mein eigener Weg
- Vergleich **OGD** zu **ORD**
- Vorstellung ausgewählter Projekte des **SDSC**

SDSC: Swiss Data Science Center

SDSC: Swiss Data Science Center von 2017 bis 2024

Zürich



Lausanne



Villigen (PSI)



Kollaboration von ETH & EPFL
Gegründet vom ETH Rat

SDSC:
Nationale Digitale Forschungs-Infrastruktur

2017 - 2024

2025

Forschung

Industrie

Infrastruktur

Biomedizin

Renku (Data Science Platform)





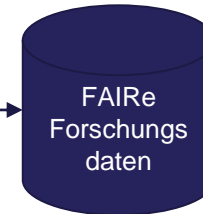
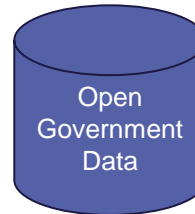
Mein eigener Weg

Diplom
Mathematik:
Uni Hamburg

Software
Ingenieurin

Liip:
Opendata Swiss
DCAT-AP CH

SDSC:
Im ORDES Team:
Open Data Research
Engagement und Services



Vergleich zwischen OGD und ORD

Findability: Auffindbarkeit 

Wie kann ich die Daten finden?

Accessibility: Zugänglichkeit 

Gibt es einen Standard-Weg die Daten zu erhalten?

Interoperability: Interoperabilität 

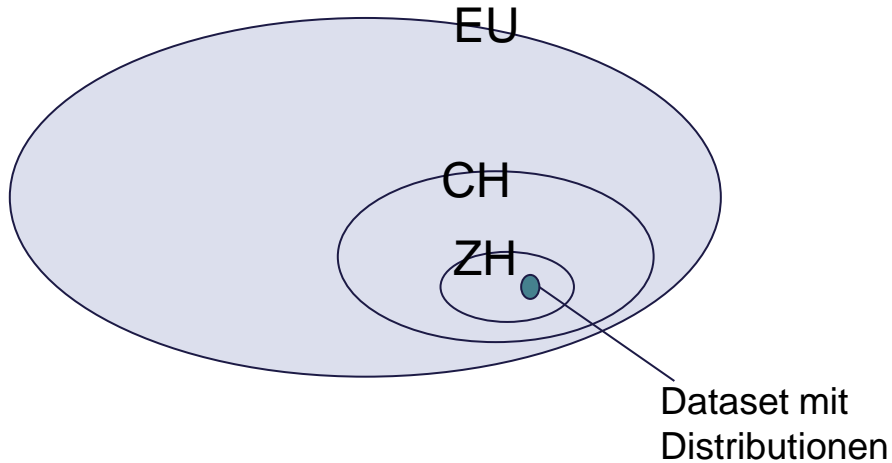
Kann man die Daten verstehen?

Reusability: Wiederverwendbarkeit 

Kann ich die Daten wiederverwenden?



Open Government Data



Portale

- Portale sind bekannt
- Portale haben eine hierarchische Struktur
- Portale harvesten sich gegenseitig

Metadaten

- Es gibt mit DCAT einen gemeinsamen Metadaten Standard

Objekte

- Datasets mit Metadaten und Distributionen (den eigentlichen Daten)
- Daten sind offen und nicht sensitiv

data.europe.eu ist eine zentrale Autorität, die FAIR misst

Country	Name	Findability 100 Points	Accessibility 100 Points	Interoperability 110 Points	Reusability 75 Points	Contextuality 20 Points	Rating 405 Points
---------	------	---------------------------	-----------------------------	--------------------------------	--------------------------	----------------------------	----------------------



opendata.swiss (CHE)

71 / 100

66 / 100

52 / 110

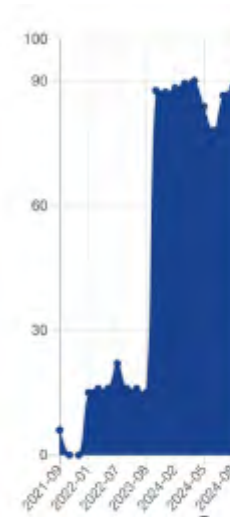
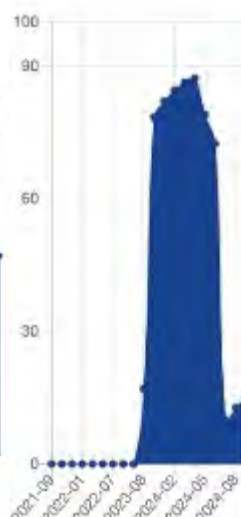
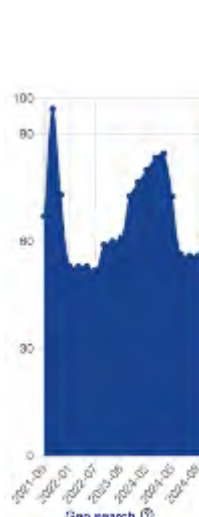
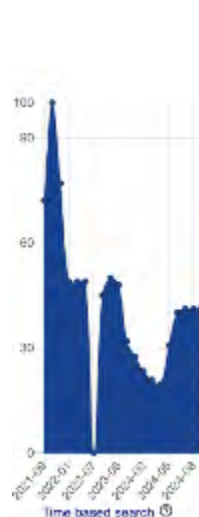
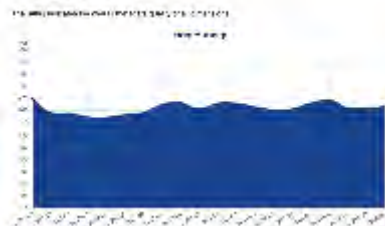
48 / 75

14 / 20

250 / 405

Good

Rating evolution

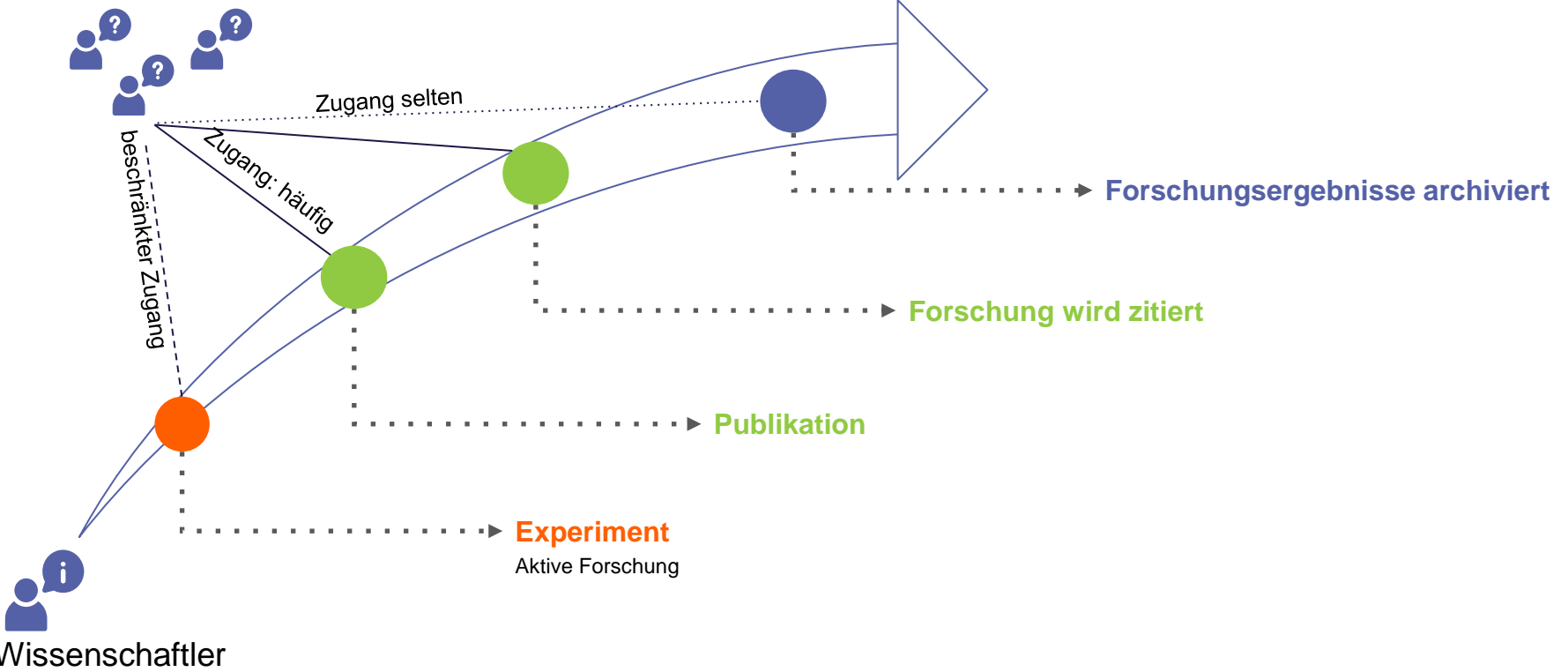




FAIRe Forschung

ORD richtet sich an Wissenschaftler: an jeden einzelnen

Wissenschaftsgemeinschaft

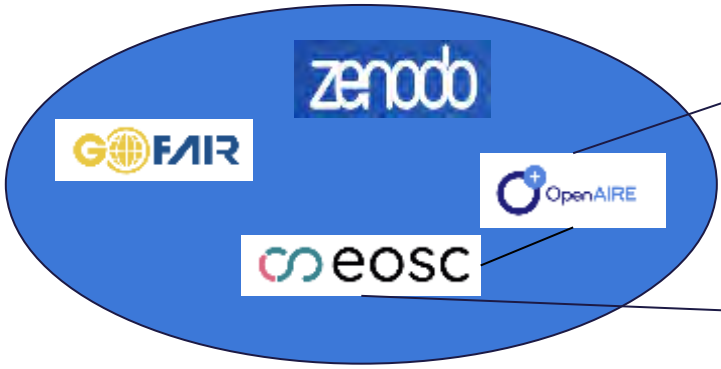


ORD Initiativen bilden Netzwerke: es gibt kein Zentrum

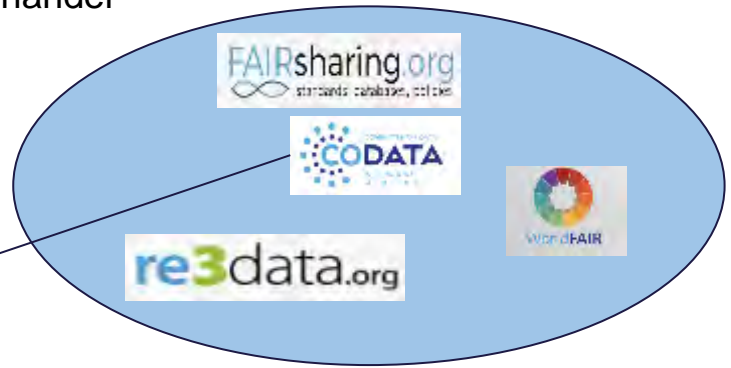


- Einige Initiativen sind zeitlich begrenzt
- Einige hängen zusammen oder gehen auseinander hervor
- Es geht um Publikationen, Daten und Code

EU



International



DE



CH



F: Findability (Auffindbarkeit)

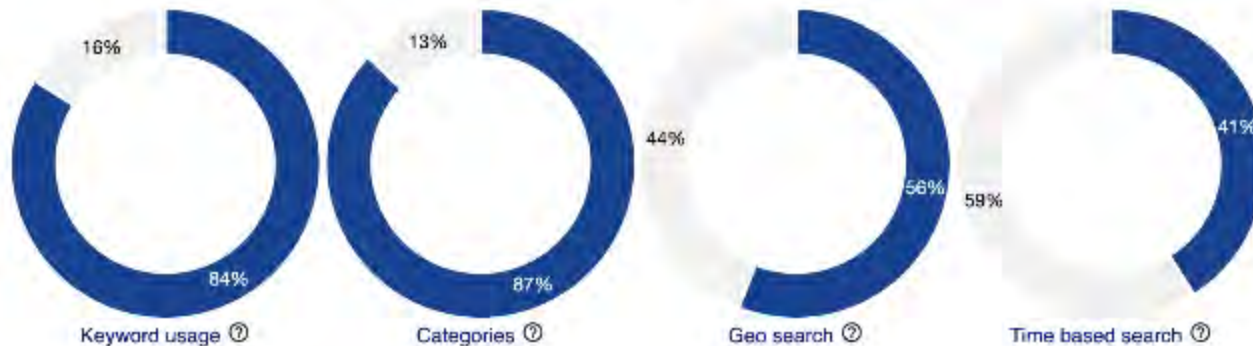


Findability



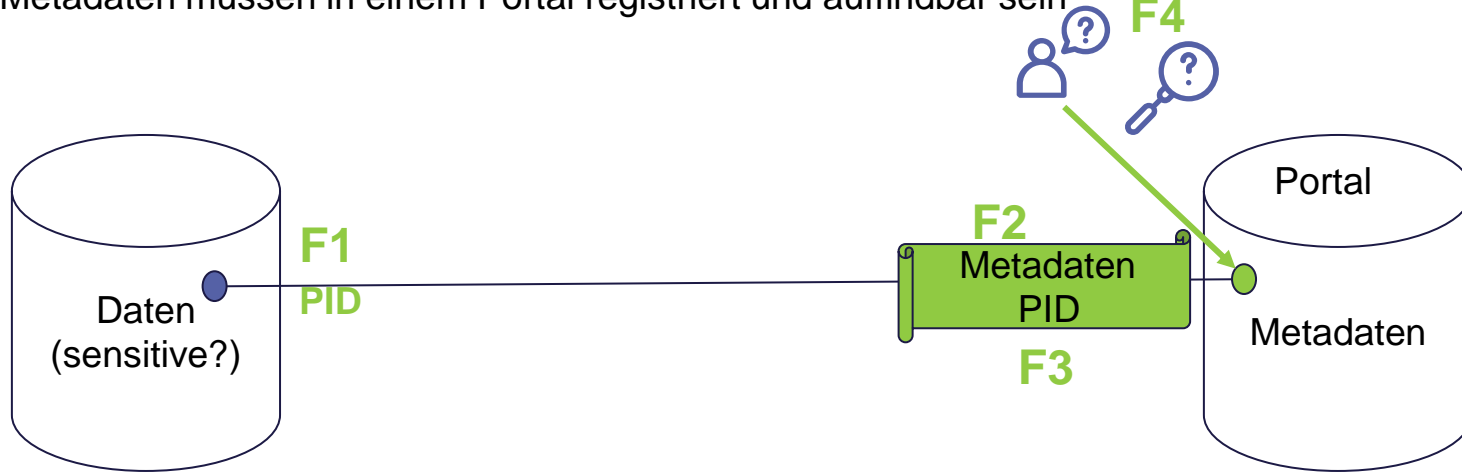
opendata.swiss (CHE)

These are indicators that help people and machines in finding datasets



Auffindbarkeit bedeutet: gute Metadaten für die Auffindbarkeit (Kategorien, Schlagwörter, Raum und Zeitbezug)

- F1** Daten brauchen einen **global eindeutigen persistenten Identifier (PID)**
- F2** Daten brauchen “reiche” Metadaten (hinsichtlich Auffindbarkeit)
- F3** Metadaten müssen den Identifier enthalten
- F4** Metadaten müssen in einem Portal registriert und auffindbar sein





Meistgenutztes Portal für die Publikation von Forschungsergebnissen: gegründet von CERN, finanziert durch die EU



Meinst genutztes Identifizierungssystem für Forschungsdaten, Forschungsportale wie Zenodo können DOIs vergeben



Portal zum Auffinden von Portalen (nicht von Daten)

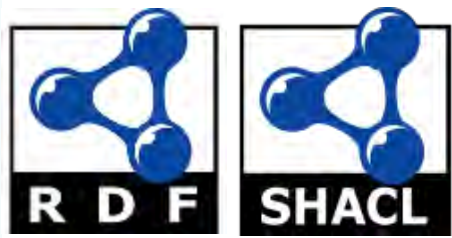


Zur Identifizierung von Forschern (den Personen)



Imaging Plaza ist eine Plattform zum Auffinden von bildverarbeitender Forschungs-Software, die FAIRness misst

Metadaten benutzen RDF und werden mit SHACL validiert



<https://imaging-plaza.epfl.ch/>
<https://github.com/sdsc-ordes/imaging-plaza>

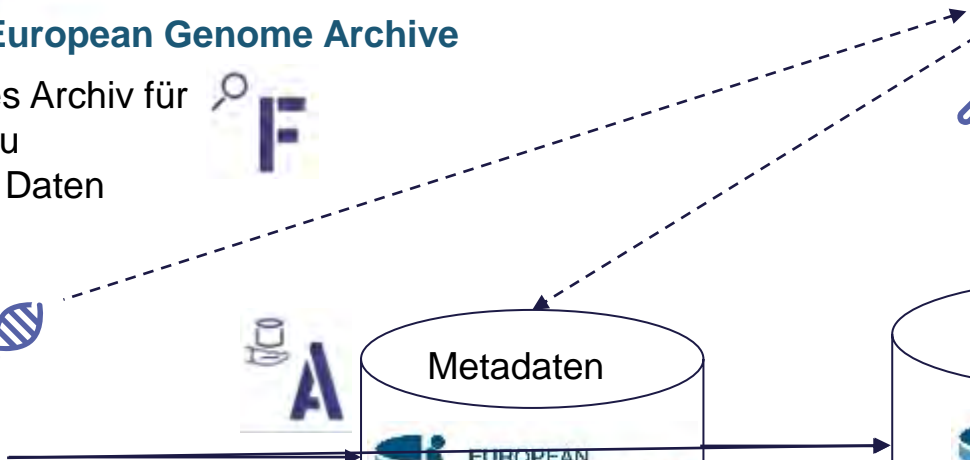
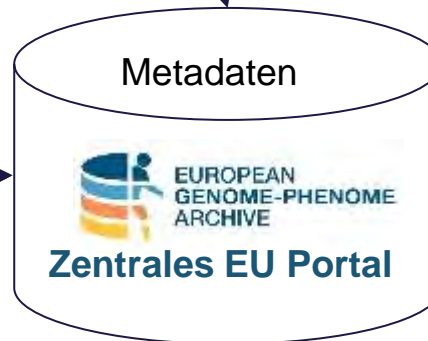


Federated European Genome Archive

Europäisches Archiv für
Metadaten zu
genetischen Daten



Sensitive
Patienten Daten

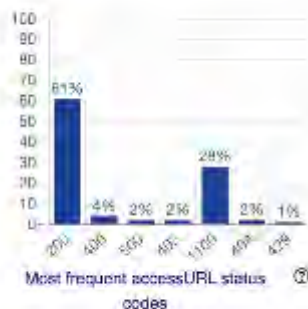


A: Accessibility (Zugänglichkeit)

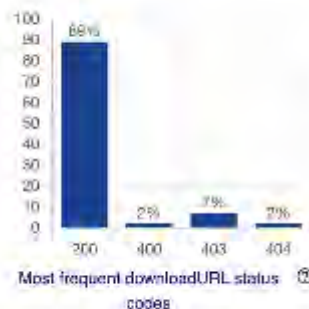


Accessibility

The following diagrams provide information about the accessibility of distributions. For each URL of a distribution, an HTTP HEAD request is performed and the response checked. Please note, that the DownloadURL is optional according to DCAT-AP and must be clearly indicated as such from the source portal. This may only be the case for source portals that provide their metadata in DCAT-AP.



accessURL



downloadURL



Zugänglichkeit: Daten sind erreichbar und können, wenn möglich, heruntergeladen werden



opendata.swiss (CHE)

opendata.swiss

Formate

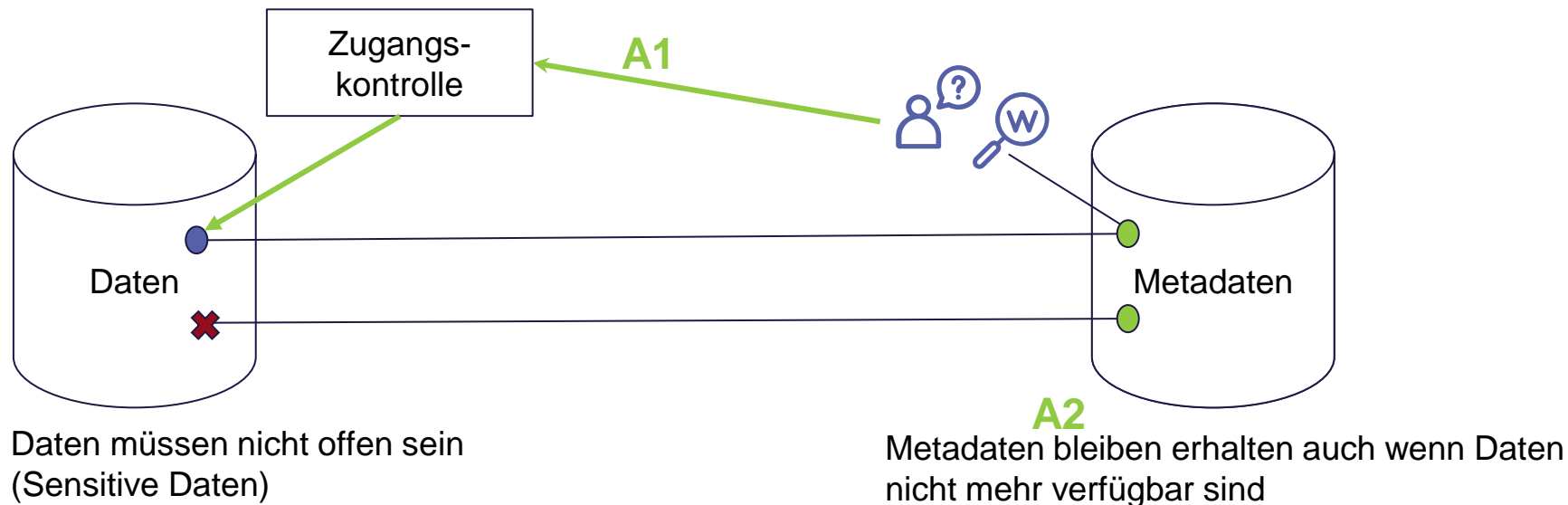
SERVICE 554	HTML 554	CSV 554	
WMS 200	XLS 200	JSON 200	
N/A 102	NFS 102	ZIP 102	
RDFXML 424			
JSON-LD 512	PDF Turtle 311	311 311	
WMS 012	PDF 335	GPX 440	
API 400	DXF 300	TXT 100	SHP 111
KML 100	XML 100	ODS 211	
INDEX 125	SPARQL 86	GeoJSON 77	
TIFF 74	GDB 14	PNG 5	
GeoJSON 5	NMZ 5	YAML 5	

Linked Data

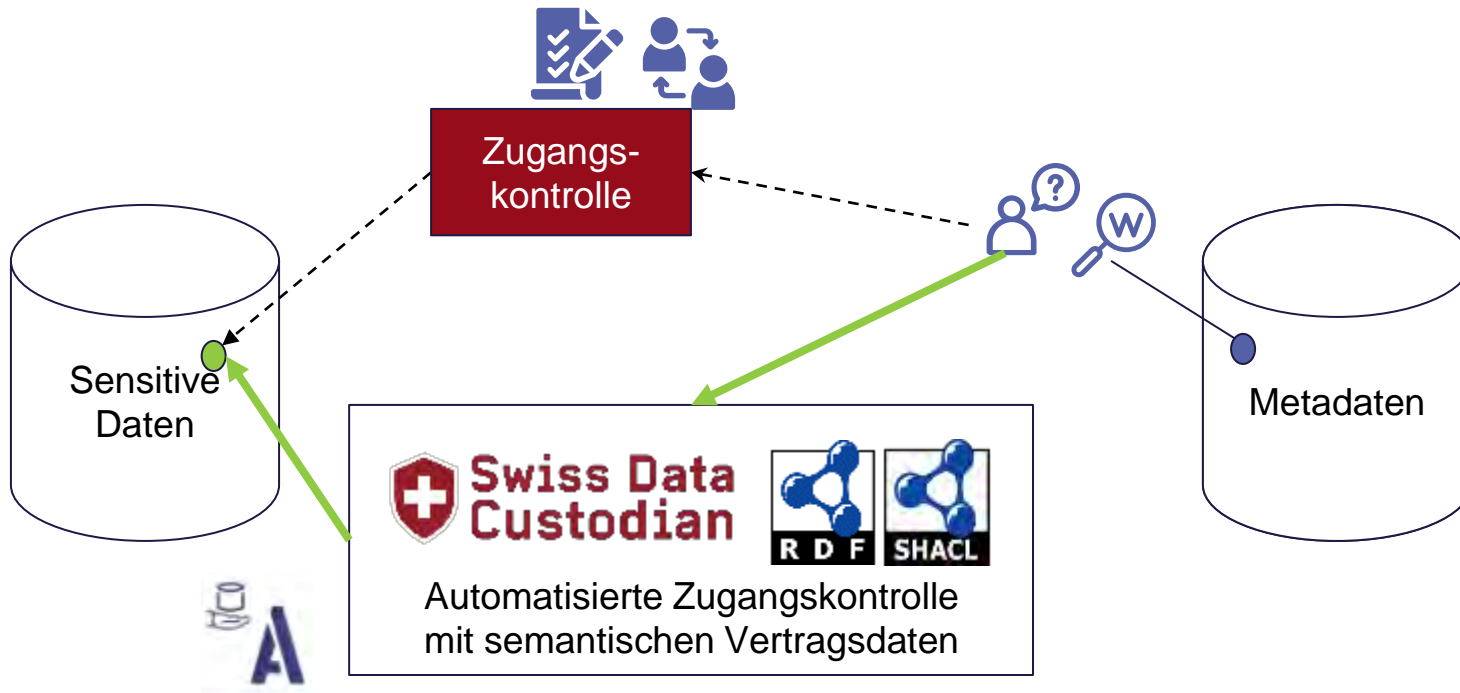
- RDF KML 128
- JSON-LD 813
- N3 811
- RDF Turtle 311
- SPARQL 86

A1. Metadaten müssen durch ein standardisiertes Protokoll über ihren Identifier angefordert werden können. Das Verfahren kann Authentisierung und Genehmigungen beinhalten.

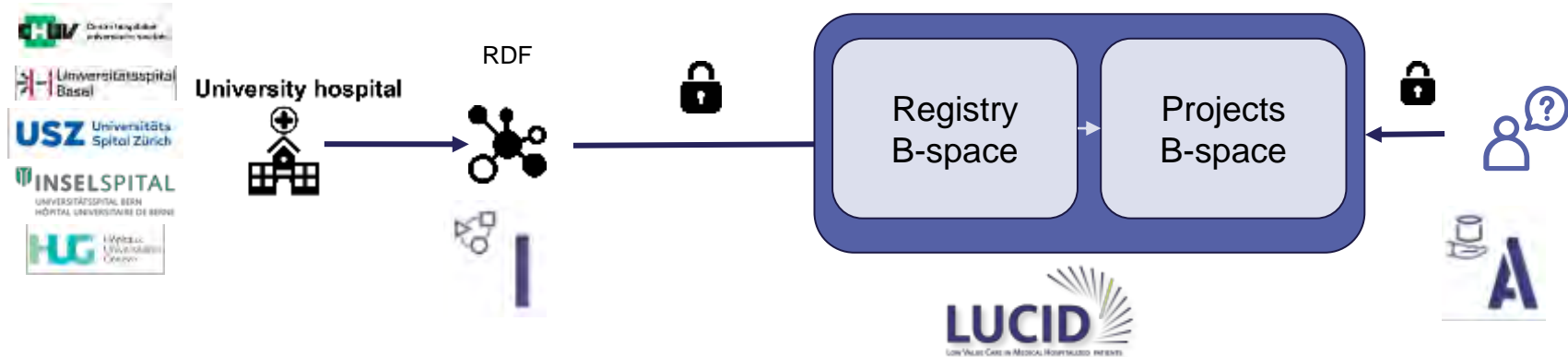
A2 Die Metadaten müssen verfügbar bleiben auch wenn die Daten nicht mehr verfügbar sind.



Anforderung sensibler Forschungsdaten ist oft ein langwieriger Prozess mit vielen Stellen, die zustimmen müssen und vielen Emails, die hin und her gehen



LUCID: nationaler Datenstrom von Universitätshospitälern um unnötige Behandlungen zu verringern (Behandlungen, die keinen Mehrwert für den Patienten bringen)



I: Interoperability Interoperabilität

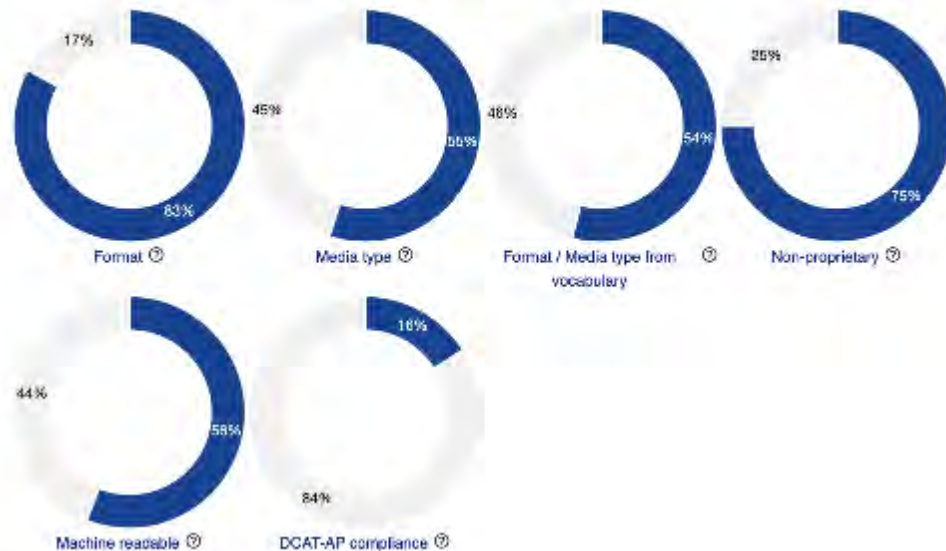
 **F**  **A**  **I**  **R**
R E S E A R C H

Interoperability



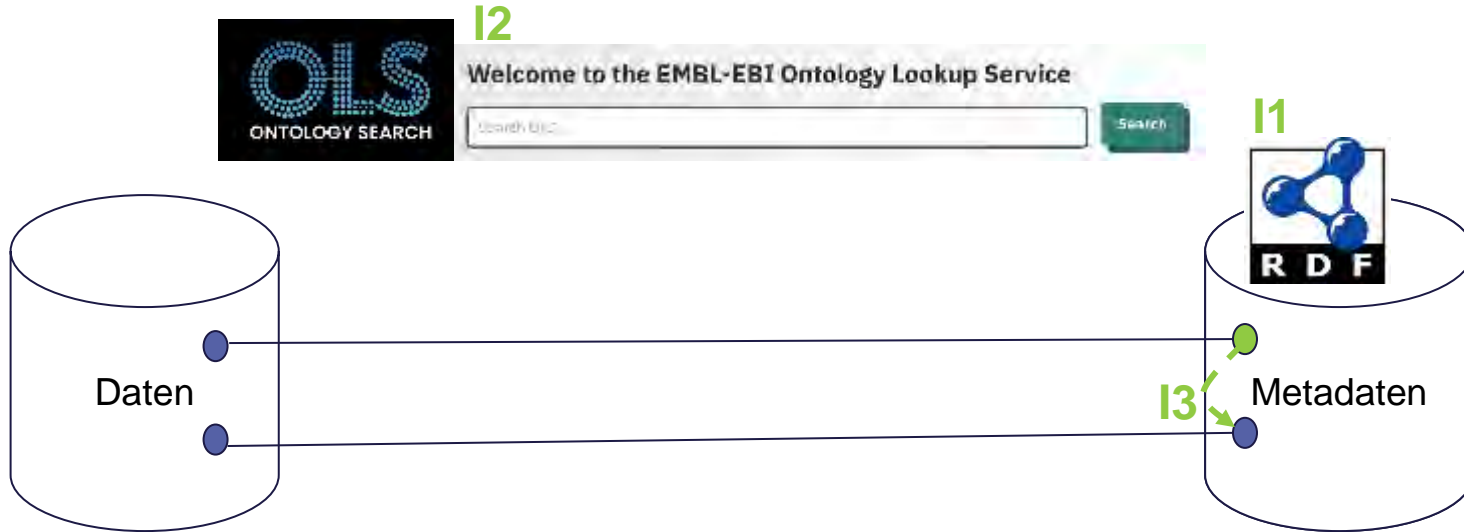
opendata.swiss (CHE)

These are indicators that gain the interoperability of a dataset

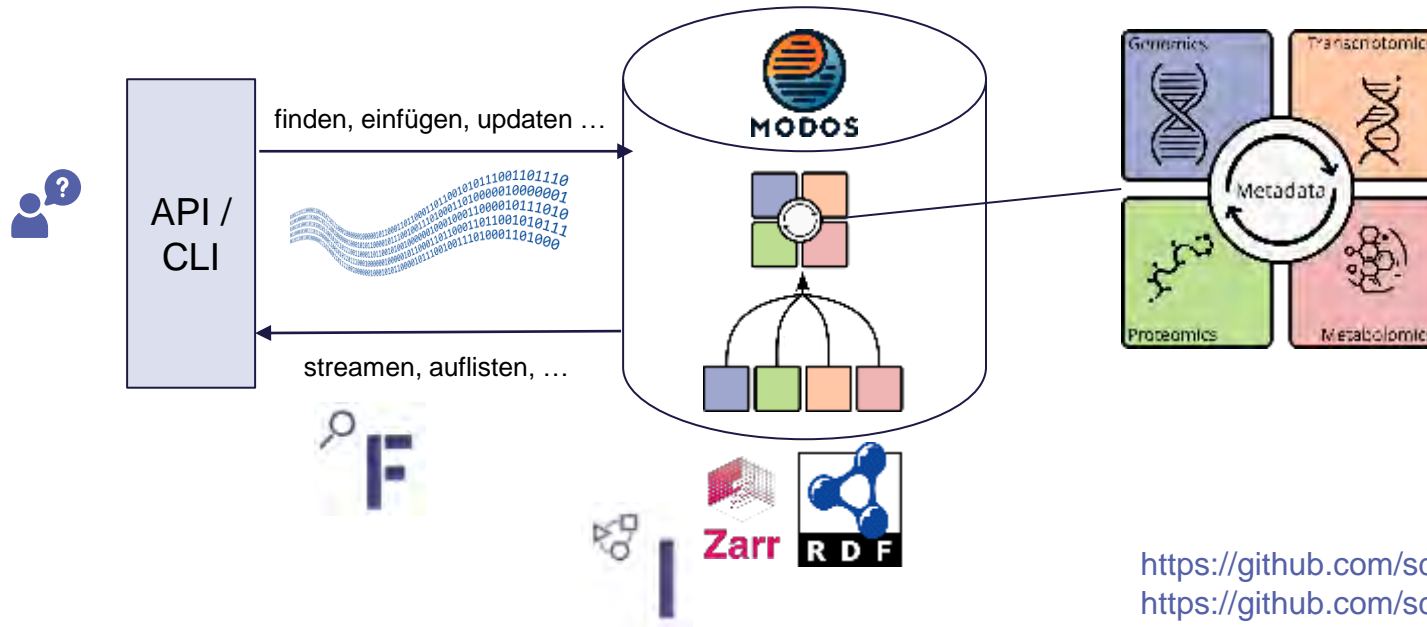


Interoperabilität: Daten sind maschinen lesbar und Daten und Metadaten sind standardisiert

- I1 (Meta)daten müssen in einer formalen, zugänglichen, bekannten Sprache verfasst sein
- I2 (Meta)daten benutzen Vokabulare, die den FAIR Regeln folgen
- I3 Metadaten verlinken mit zu anderen relevanten Metadaten



- Es geht um Multiomics Daten im Gesundheitswesen
- Schaffung einer Forschungsdaten-Infrastruktur für einen modularen Zugang zu diesen Daten

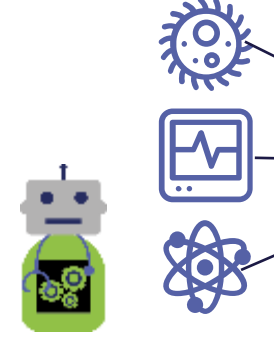


<https://github.com/sdsc-ordes/modos-schema>
<https://github.com/sdsc-ordes/modos-api>
<https://github.com/sdsc-ordes/modos-ontology>

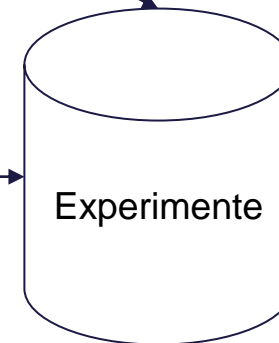
- Automatisiertes Chemie-Labor zur Suche nach Katalysatoren (Prozess-Beschleunigern)
- Zarr Format speichert Daten zusammen mit den Metadaten
- Entwicklung eines Vokabulars und Portals für Daten und Metadaten der Experimente



Automatisiertes Labor in dem Roboter die Experimente durchführen



Chemiker





R: Reusability Wiederverwendbarkeit

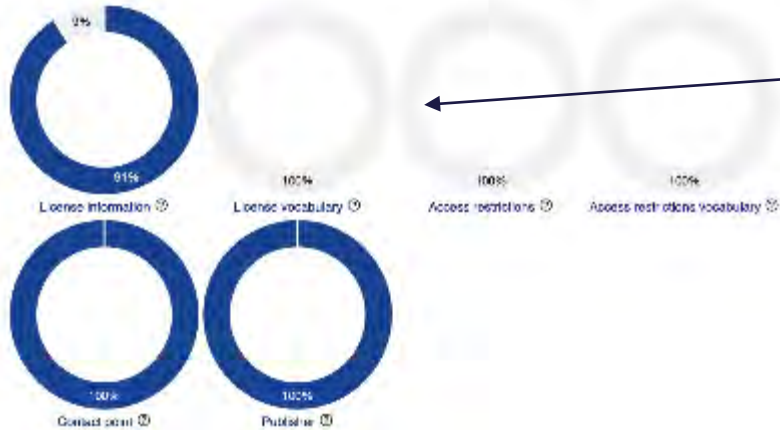


Reusability



opendata.swiss (CHE)

The following table describes which metrics are used to check the reusability of the data. A maximum of 75 points can be scored in this area.



opendata.swiss

Nutzungsbedingungen



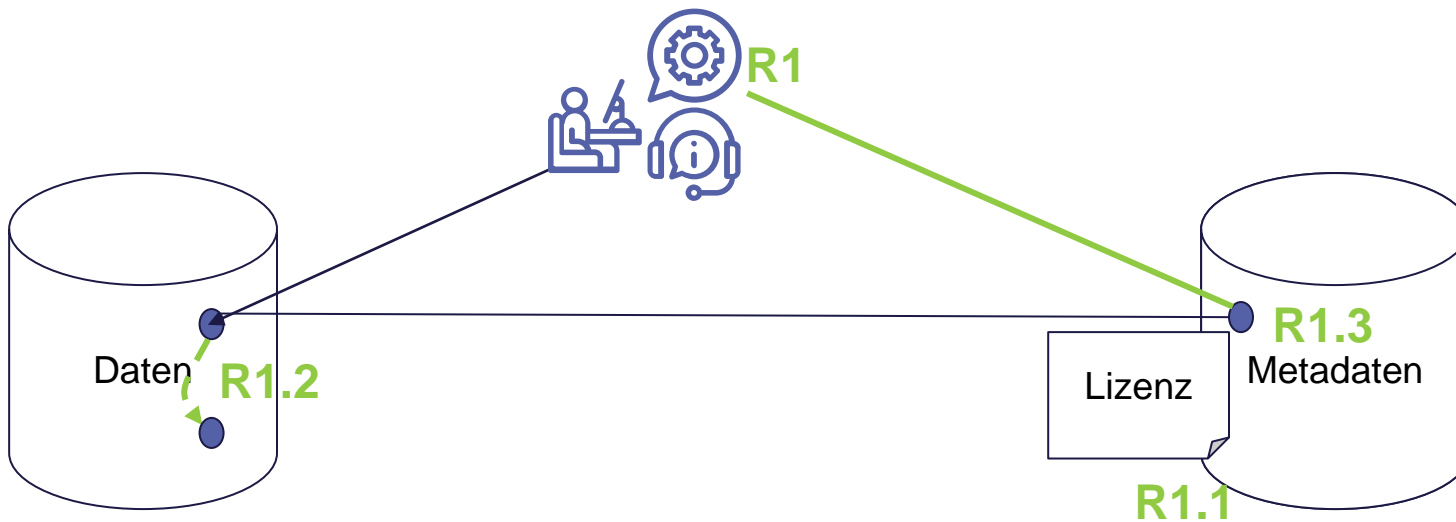
Wiederverwendbarkeit: Daten haben eine Lizenz, Ansprechpartner und eine Publizierende Stelle

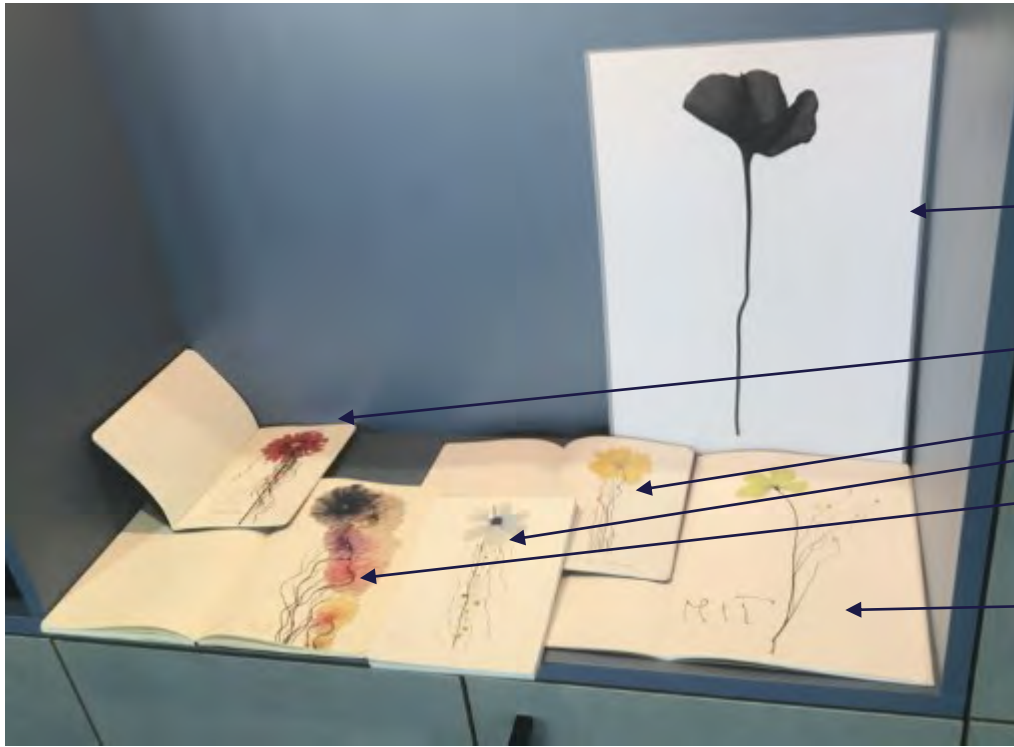
R1 (Meta)daten sind verständlich beschrieben hinsichtlich ihrer Benutzbarkeit

R1.1 Es gibt eine klare und zugängliche Lizenz für die Wiederverwendung

R1.2 Der Ursprung der Daten ist geklärt und beschrieben

R1.3 Die Metadaten und Daten entsprechen den Community Richtlinien





SDSC KI Kunstprojekt

No Licence

Apache

BSD

CC

GPL

MIT

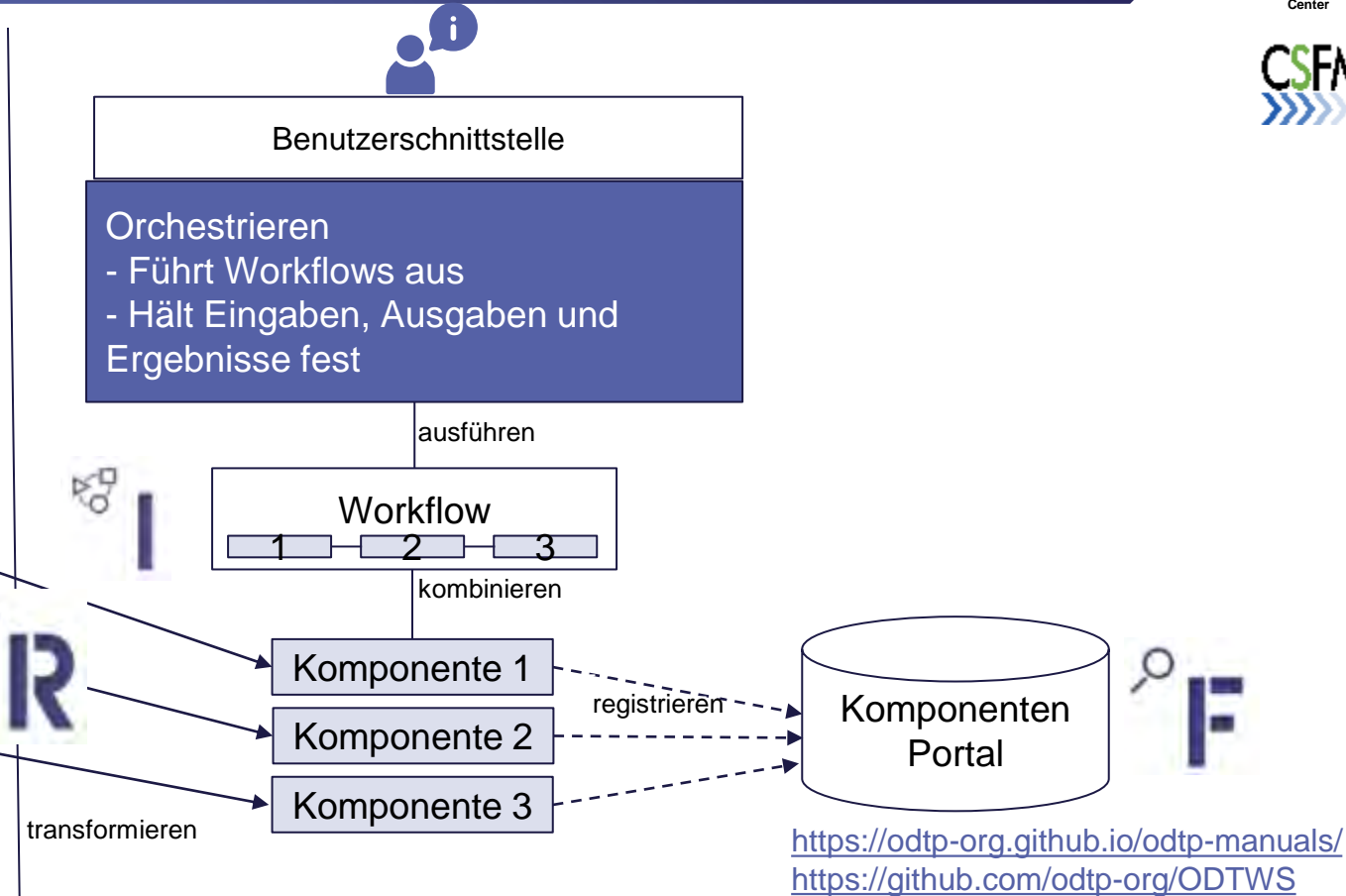
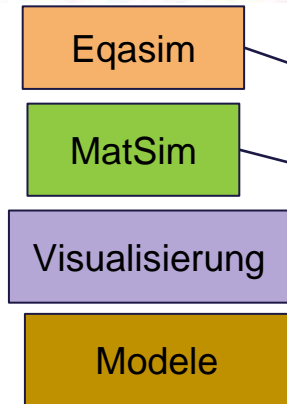
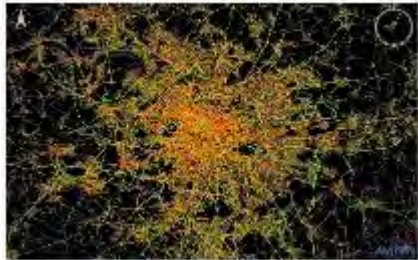


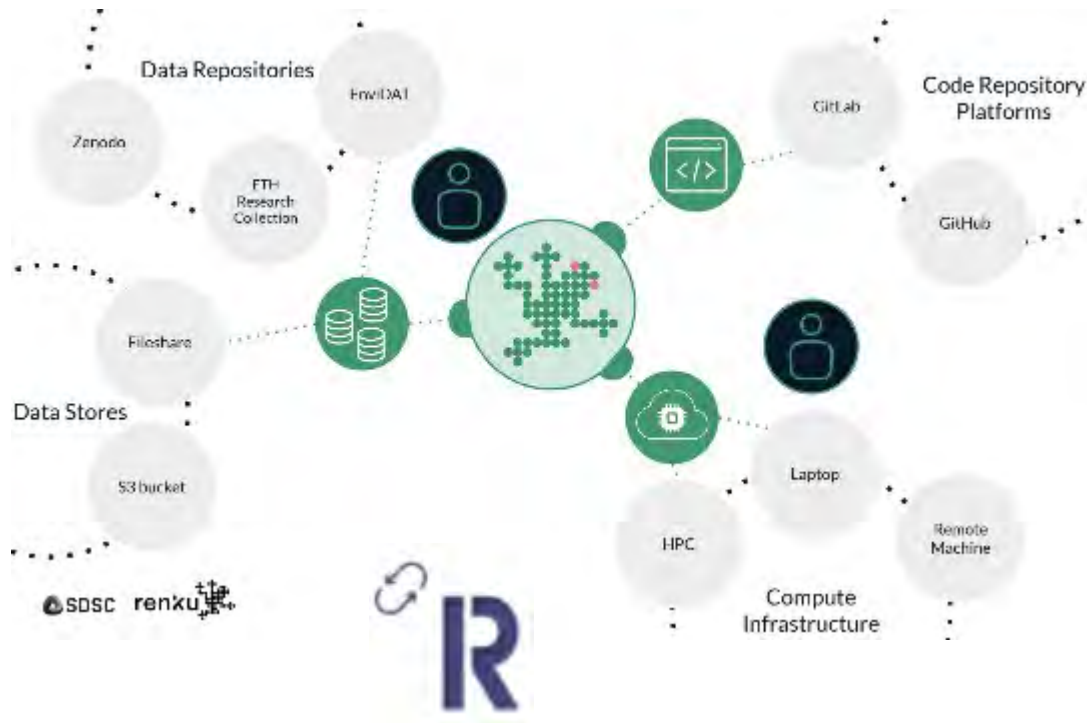
Quelle: Papers with Code 

<https://zenodo.org/records/10012457>

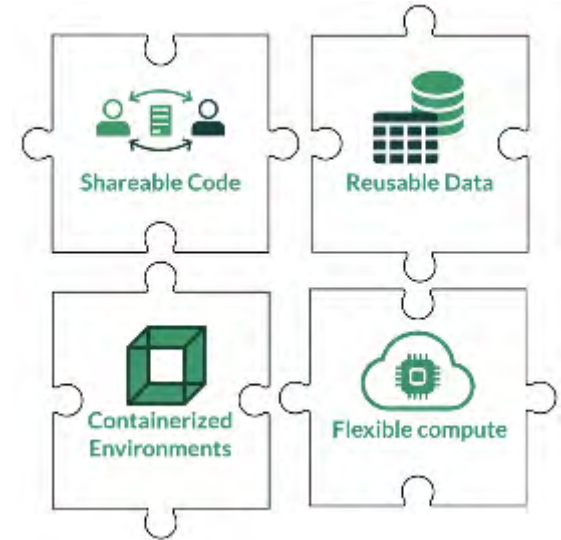
<https://github.com/sdsc-ordes/license-collector>

Bereich:
Mobilitäts-
Simulationen





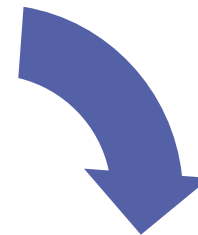
Bei Renku geht es um die Reproduzierbarkeit von Forschungsergebnissen





100 interdisziplinäre Teilnehmer fokussiert auf
Research Daten Infrastruktur (RDI)

Schnelle
Prototypen:
Ausprobieren
von RDI
Konzepten



Aufbau von
scalierbaren,
integrierten und
nachhaltigen
Forschungsdaten-
Infrastrukturen





Fragen?
