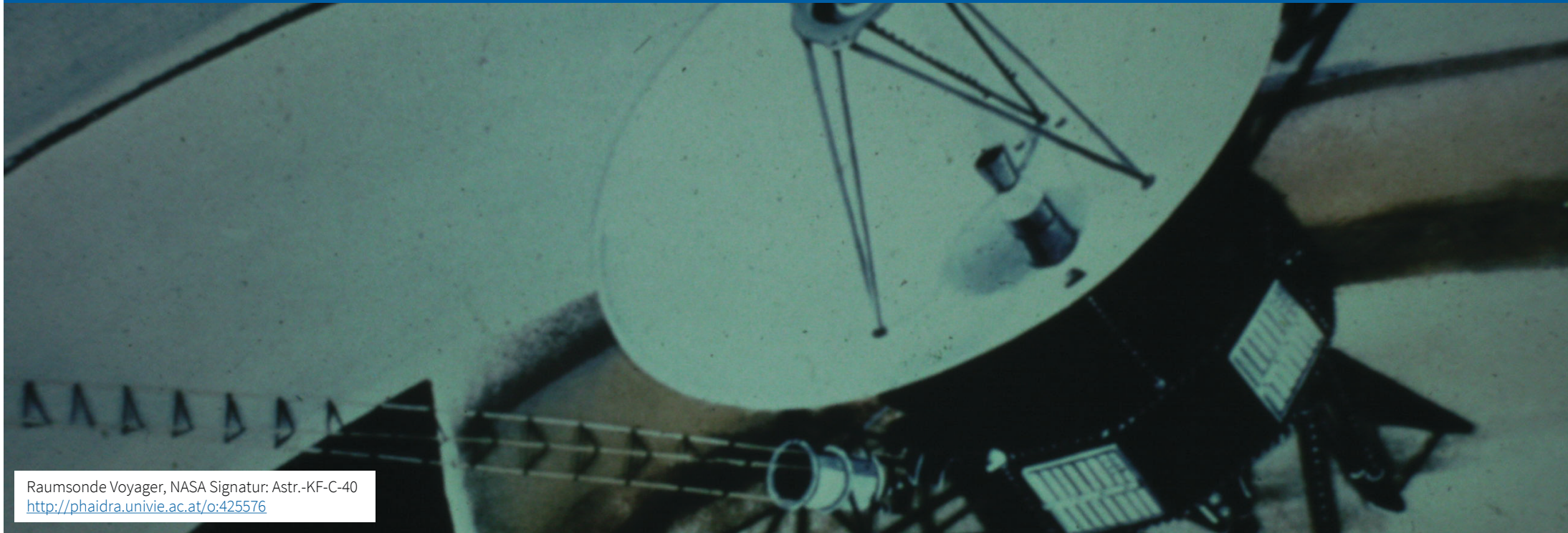


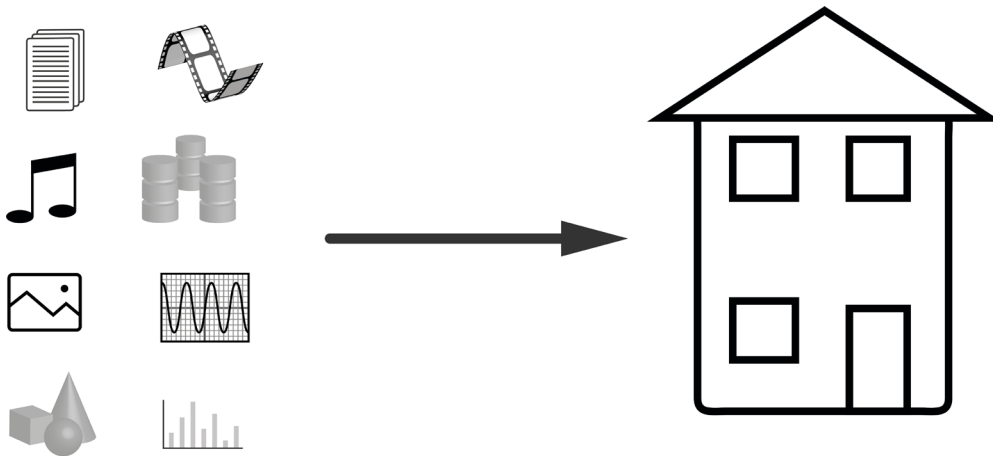
# Datenmanagement und Archivierung von digitalen Daten

Anhand des Beispiels von Phaidra



Raumsonde Voyager, NASA Signatur: Astr.-KF-C-40  
<http://phaidra.univie.ac.at/o:425576>

## Die Welt der Daten verwalten



- Einen Ort zu finden (wo liegen die Daten)
- Ein System zu finden (wie sind die Daten geordnet)
- Regeln zu definieren (was darf mit den Daten getan werden)
- Dauer zu definieren (wie lange sind die Daten verfügbar)

## FAIR

### Prinzipien für das Datenmanagement

Mark D. Wilkinson and others: The FAIR Guiding Principles for scientific data management and stewardship; in Nature/Scientific Data; DOI: 10.1038/sdata.2016.18; <https://www.nature.com/articles/sdata201618>

**F**indable

Persistent Identifier,  
searchable, useful metadata

**A**ccessible

retrievable by their identifier,  
metadata always accessible

**I**nteroperable

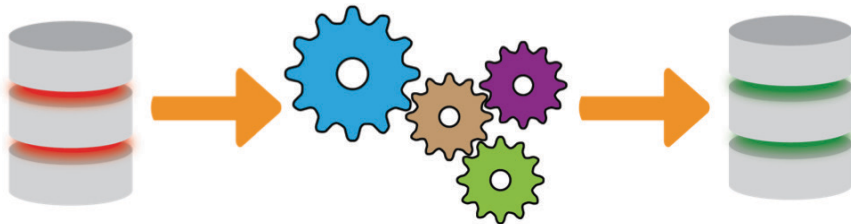
formal and broadly applicable  
language for description

**R**euseable

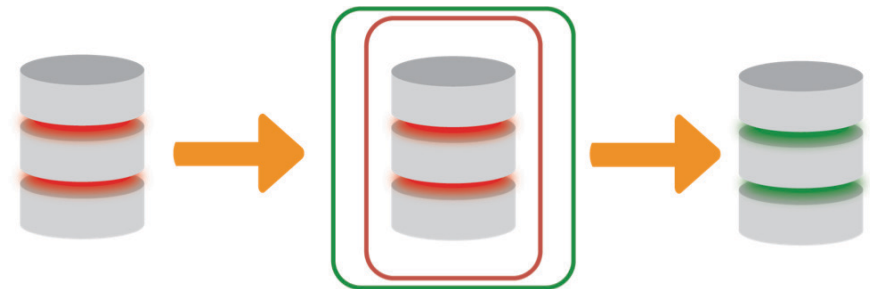
clear and accessible data  
usage license

## Langfristige Verfügbarkeit der Daten

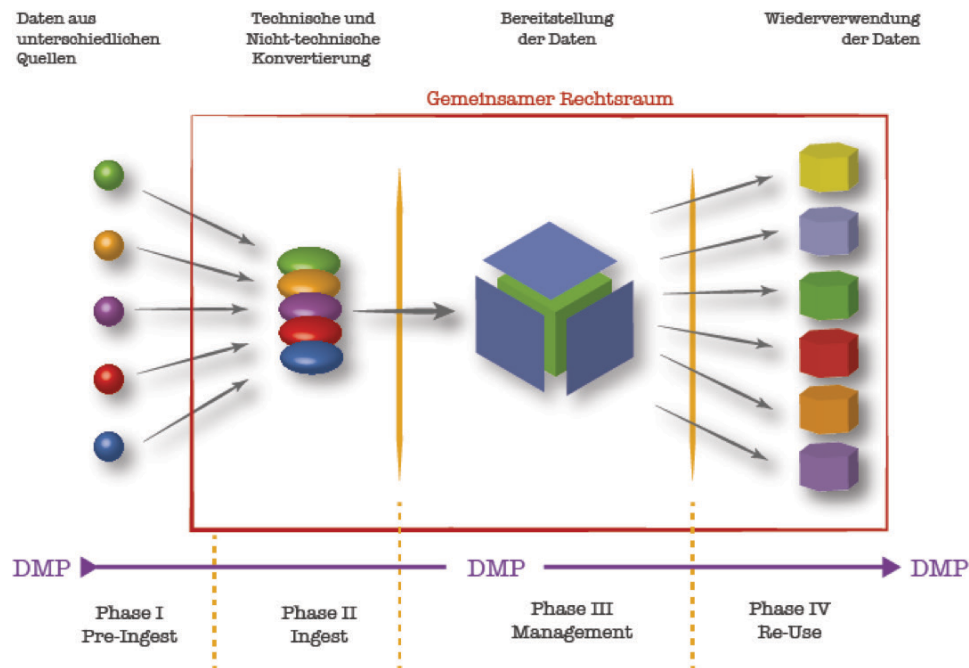
### Migration



### Emulation

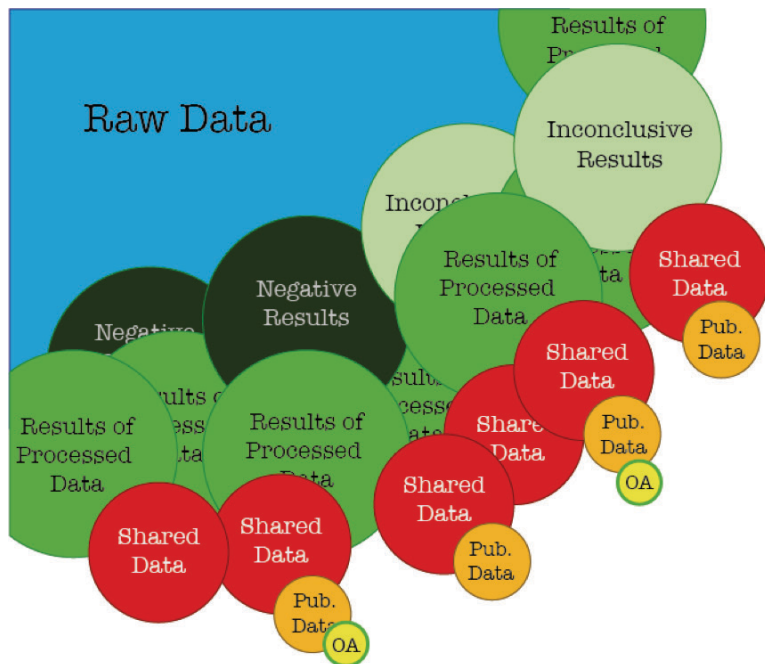


## Workflow aus der Sicht des Datenmanagement



- Pre-Ingest: Daten werden in der bestmöglichen Qualität erstellt
- Ingest: Übergabe der Daten von der Produktion zum Datenmanagement
- Management: Die Daten werden über eine bestimmte Zeit in der übergebenen Qualität den aufbewahrt.
- Re-Use: Die Daten werden für eine bestimmten Zweck an berechtigte BenutzerInnen übergeben

## Vorbereitung für die Archivierung

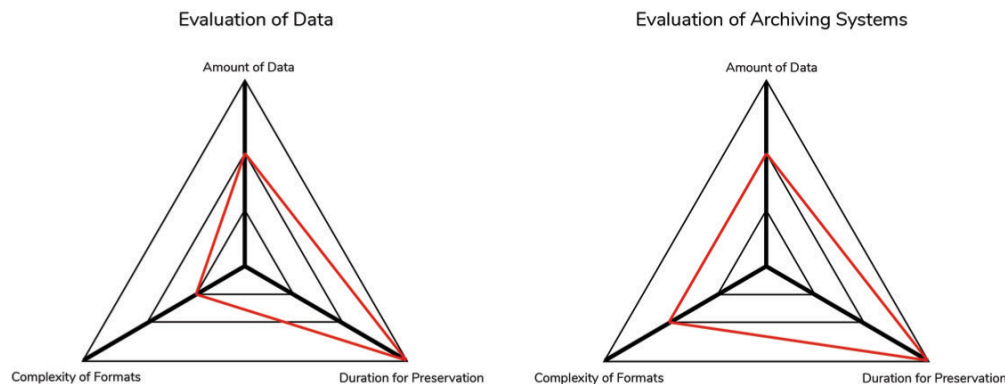


- Welche Daten sollen archiviert werden?
- Wie lange sollen die Daten archiviert werden?
- Wie soll auf die Daten später zugegriffen werden?
- Welche Datenmenge wird archiviert?
- Zu welchem Zweck wird Archiviert?
- Juristische Fragestellung so früh klären: Urheberrecht, Datenschutz und Persönlichkeitsrechte

## Bewertung der Daten aus Sicht des Datenmanagement

Wo können die Daten unterkommen?

Nicht jedes Archivsystem ist für alle Daten geeignet.

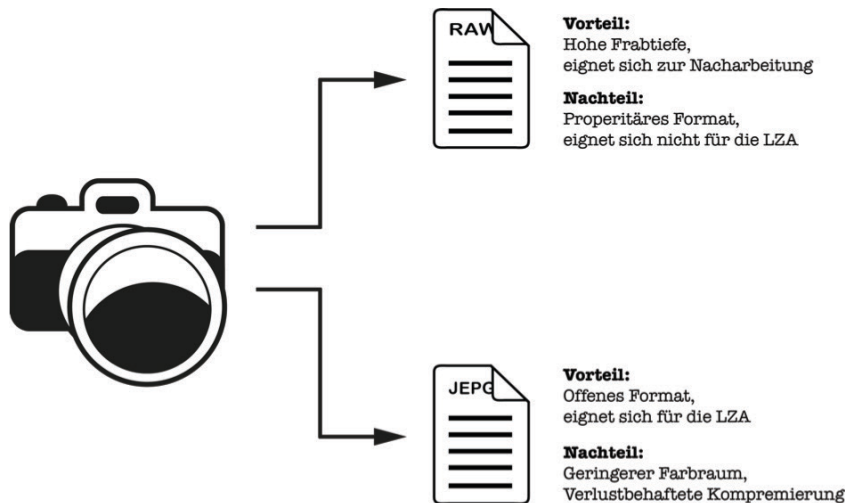


Mit definiten Kriterien und einer Gegenüberstellung der Daten und des Archivsystem kann bewertet werden, ob das System für die Daten geeignet ist.

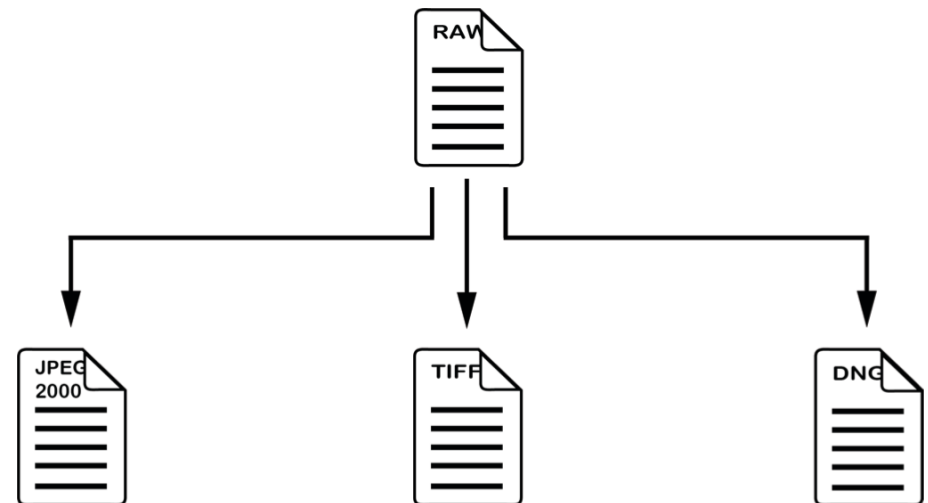


## Bewertung der Datenformate

### Auswahl bei der Produktion

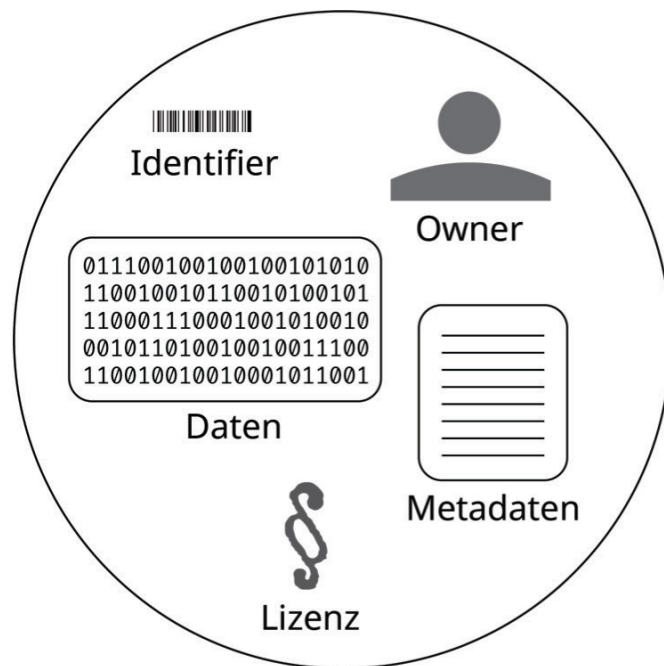


### Auswahl bei der Archivierung





## Digitales Objekt



Merkmale eines digitalen Objekts in einem Archivsystem

- Daten
- Owner
- Lizenz
- Metadaten
- Identifier

## Struktur der Daten

Semantik/Logik

Inhalt: das Wesentliche, das zu Erhalten ist

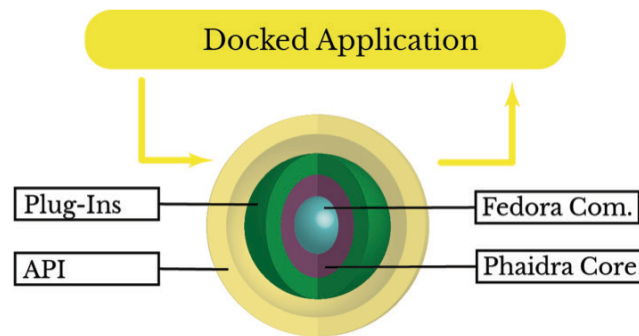
Formate


Wie ist der Inhalt datentechnisch strukturiert

Daten

Bitstream: Abhängig von physikalischen Medien

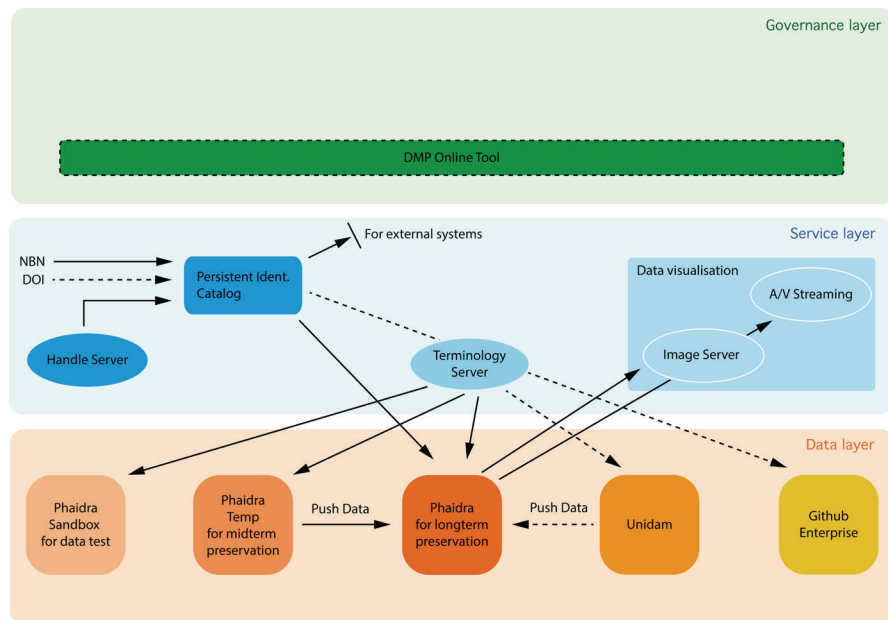
## Struktur von Phaidra Repository



 Schichten von Phaidra  
Raman Ganguly, Universität Wien, 2017

- Konzentration auf Phase 3: Management
- Basiert auf dem Open Source Repository Fedora
- API für Applikationen die vorherigen Phasen abdecken.
- Standardisiert Darstellung der Daten, die über Plug-Ins erweitert werden kann.

## Ecosystem von Repositorien



- Repositories
  - Phaidra
    - LZA
    - Temp
    - Sandbox
  - Unidam
  - Github
- Services
  - PI (Handle, DOI, URN)
  - Streaming/Image Server
  - Terminology Server

## Datenportale

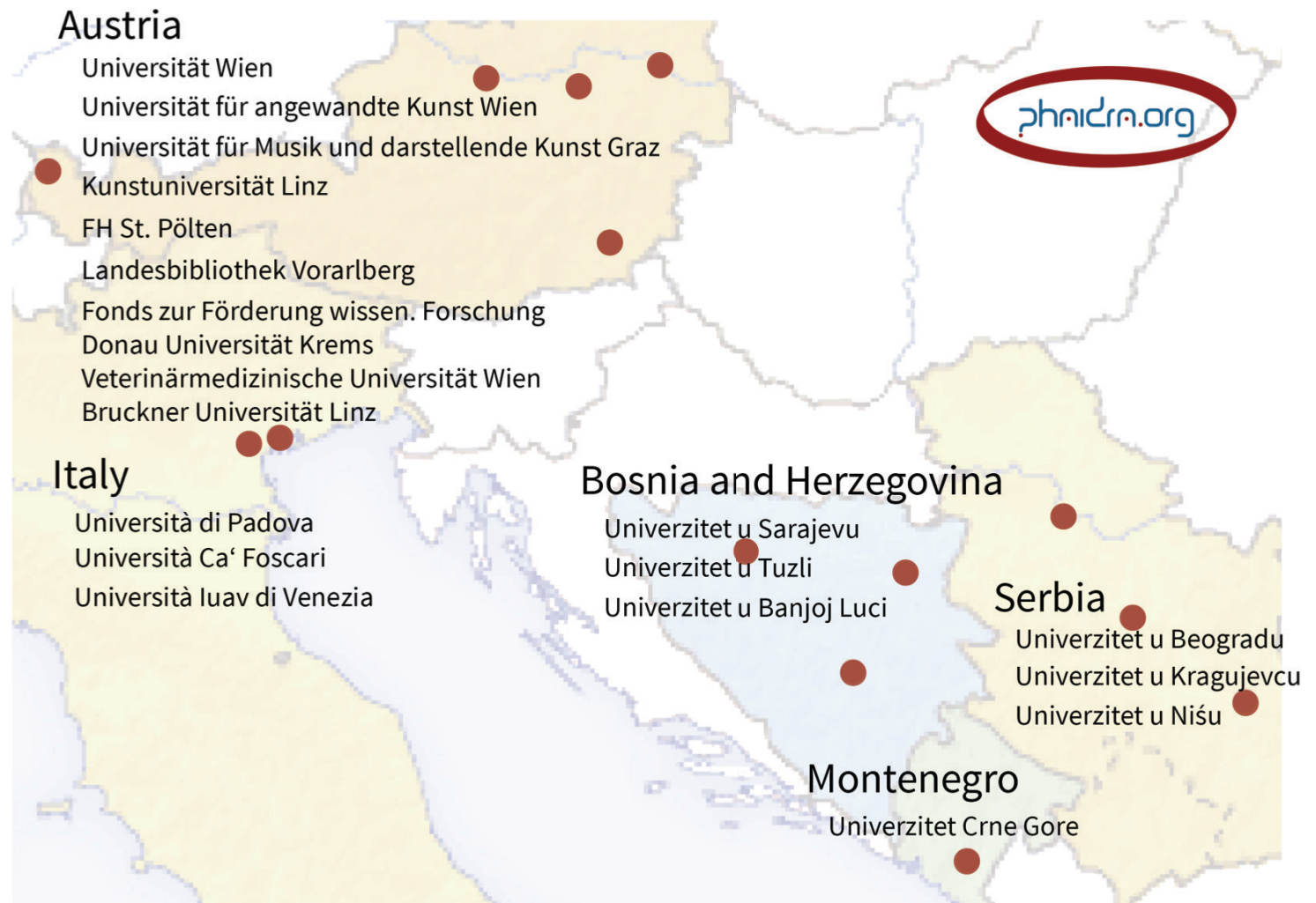


- Metadaten aus unterschiedlichen Repositorien zusammen führen und durchsuchbar zu machen.
- Prototyp im Projekt Open Education Austria
- Möglichkeit um Maschine Learning zu testen

## Netzwerke

e-infrastructures  
austria







## Danke

Autor: Raman Ganguly

Datum: 13.3.2018

Kontakt: [raman.ganguly@univie.ac.at](mailto:raman.ganguly@univie.ac.at)



[CC BY 3.0 AT - Creative Commons Namensnennung  
3.0 Österreich Lizenz.](https://creativecommons.org/licenses/by/3.0/at/)