

Dopolnitev repozitorijev nacionalne infrastrukture odprtega dostopa s funkcionalnostmi podatkovnega arhiva za velepodatke

Milan Ojsteršek

Inštitut za računalništvo, Laboratorij za heterogene računalniške sisteme
Fakulteta za elektrotehniko, računalništvo in informatiko, Univerza v Mariboru
milan.ojstersek@um.si



Podatki

- Surovi / Očiščeni
- Nekontrolirano zbrani
- Eksperimentalni
- Analitični

- Strukturirani/
Nestrukturirani

- Odprti/ Omejeno dostopni
- Nepovezani/Povezani
- Metapodatki
- Velepodatki



Vir: <https://www.bitmat.it/blog/news/83536/sviluppare-applicazioni-iot-riducendo-costi-e-risorse>

Viri raziskovalnih podatkov

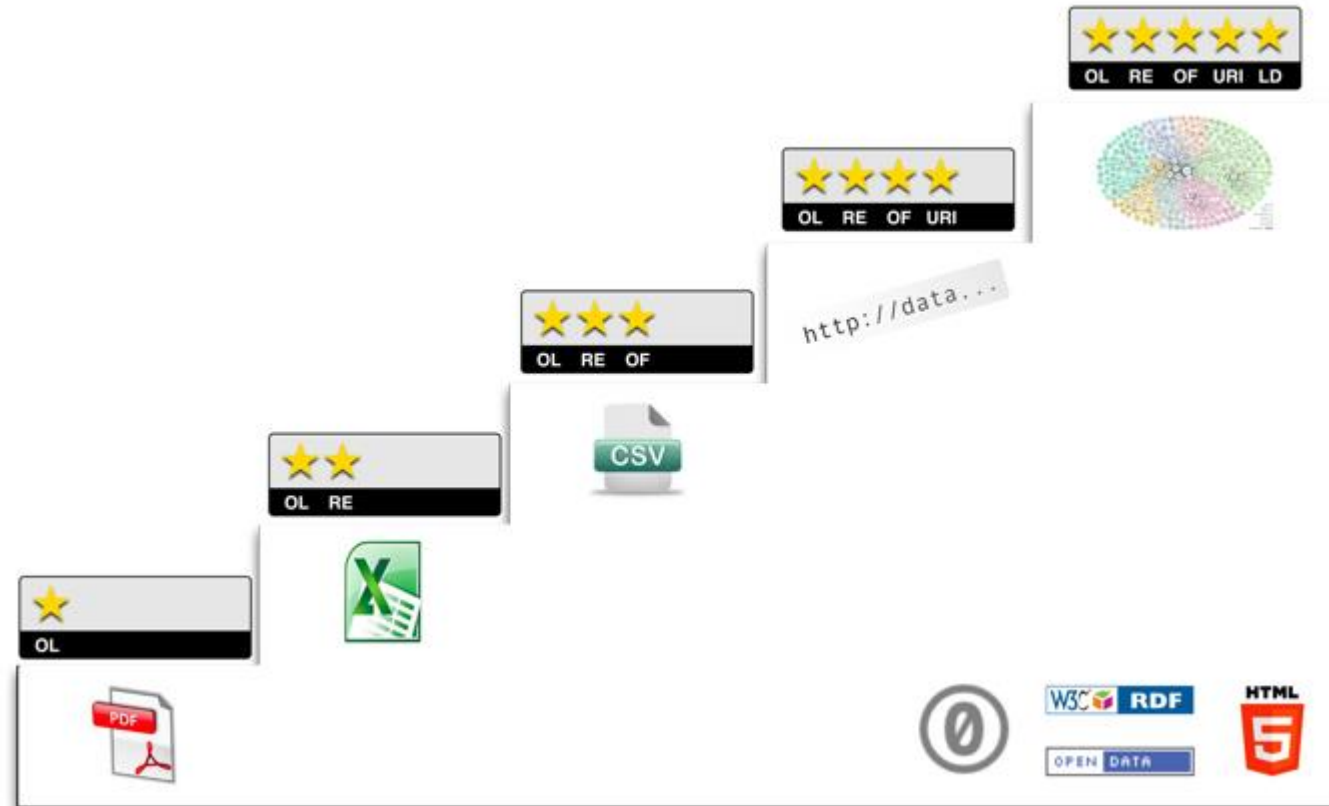
- Naprave
- Instrumenti
- Senzorji
- Programska oprema
- Ljudje



- Opazovanja
- Eksperimenti
- Simulacije
- Emulacije
- Ankete
- Intervjuji
- Analiza besedil
- Tekstovno rudarjenje
-



Pet stopenj odprtih podatkov



Več na <http://5stardata.info/>

Kaj potrebujemo za arhiv odprtih raziskovalnih podatkov?

Princip FAIR (Findable, Accessible, Interoperable, Re-usable) – več na <https://www.force11.org/group/fairgroup/fairprinciples>

Vzpostavitev procesa za podporo življenskemu krogu raziskovalnih podatkov:

- Načrtovanje in iskanje virov podatkov.
- Zbiranje in ustvarjanje.
- Obdelava in analiza.
- Objava in deljenje.
- Dolgotrajno hranjenje.
- Druga raba.

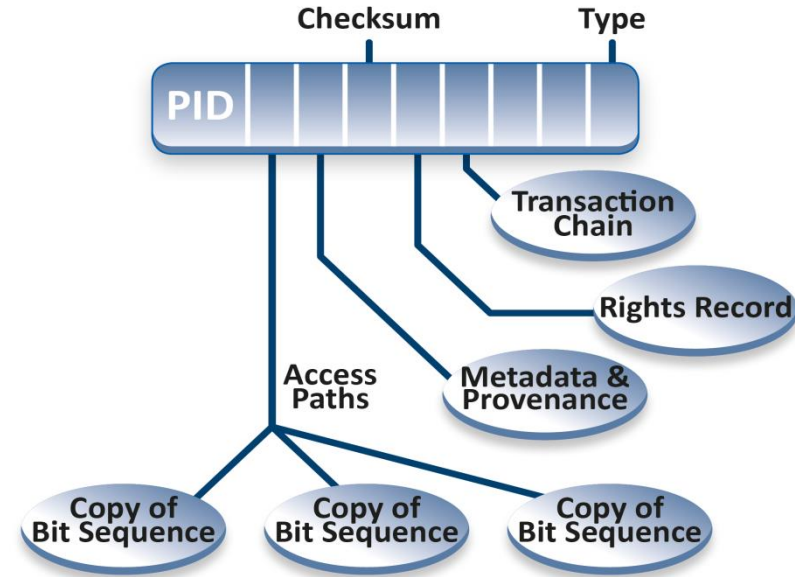
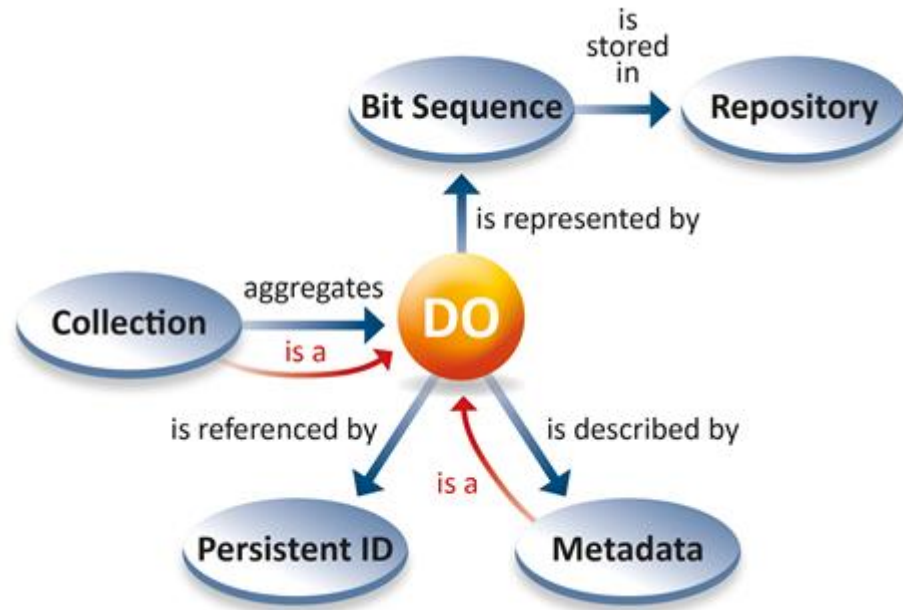
Več na <https://www.adp.fdv.unilj.si/usposobi/ZKG/>



Vir: **Practical guide to building future proof Open Data Portals -**

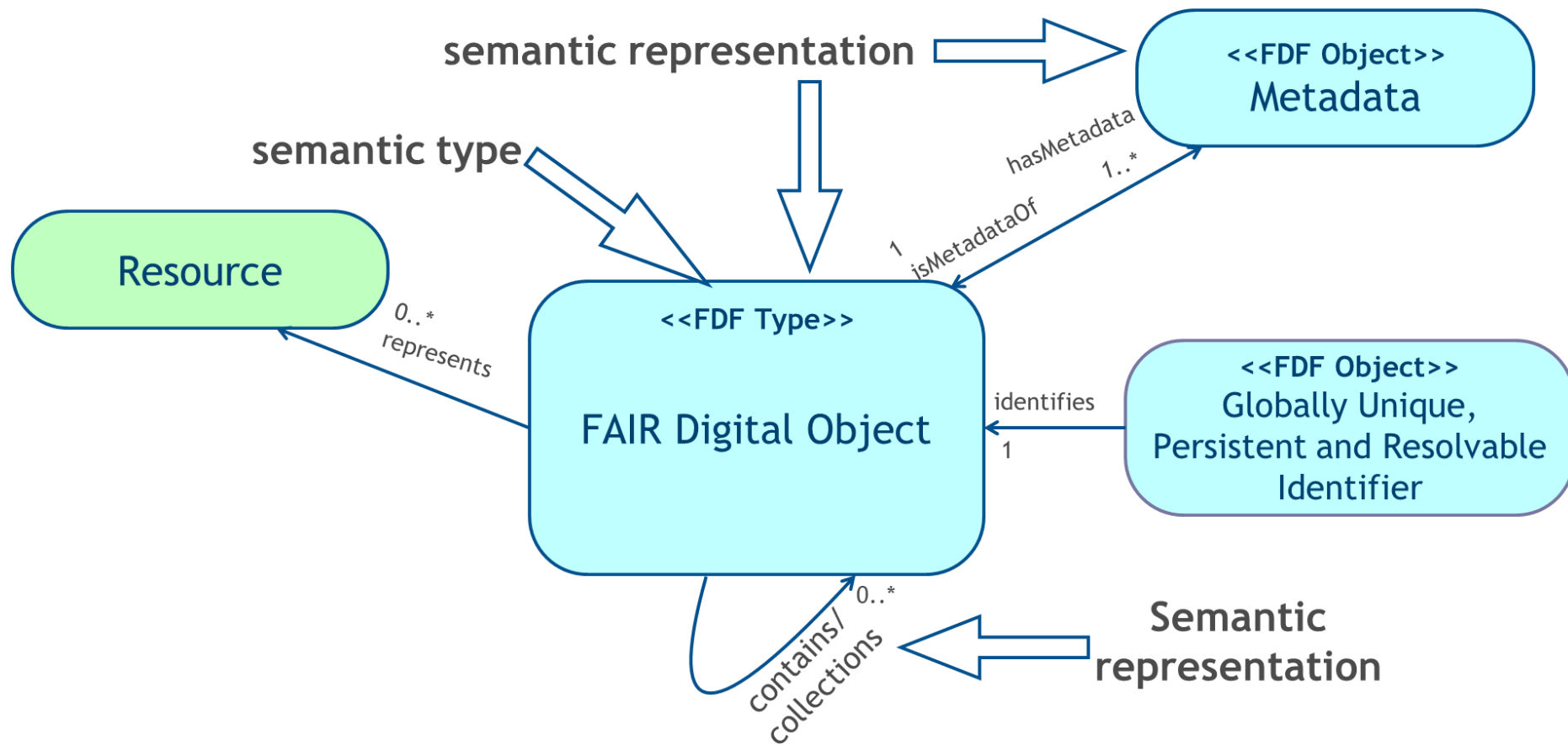
<https://www.europeandataportal.eu/sl/highlights/practical-guide-building-future-proof-open-data-portals>

FAIR digitalni objekti in trajni identifikatorji (PID)



Vir: RDA's Data Foundation & Terminology Group (DFT) 2014: Core Model

Pomenski opis digitalnega objekta








Vir: Luiz Bonino, Fair digital framework FAIR digital object + LDP

Fizični objekt v muzeju



Vir: Alex Hardisty FAIR Digital Objects as Basic Design Choice and the Need for PIDs

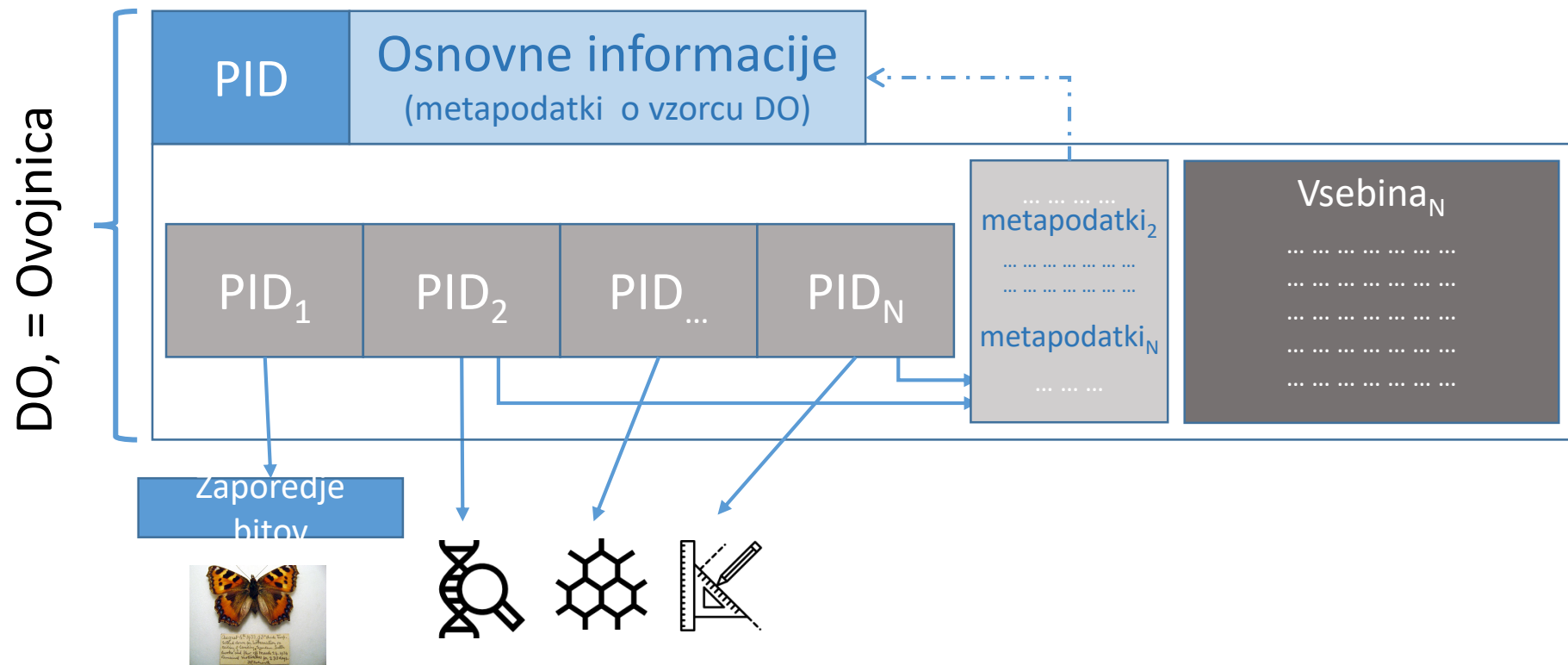
Digitalni nadomestek FAIR Digitalni Objekt

-  Genomic data
-  Biochemical data
-  Morphological data
-  Geographical data
-  Taxonomic Information
-  Species Interactions data
-  Ecological data

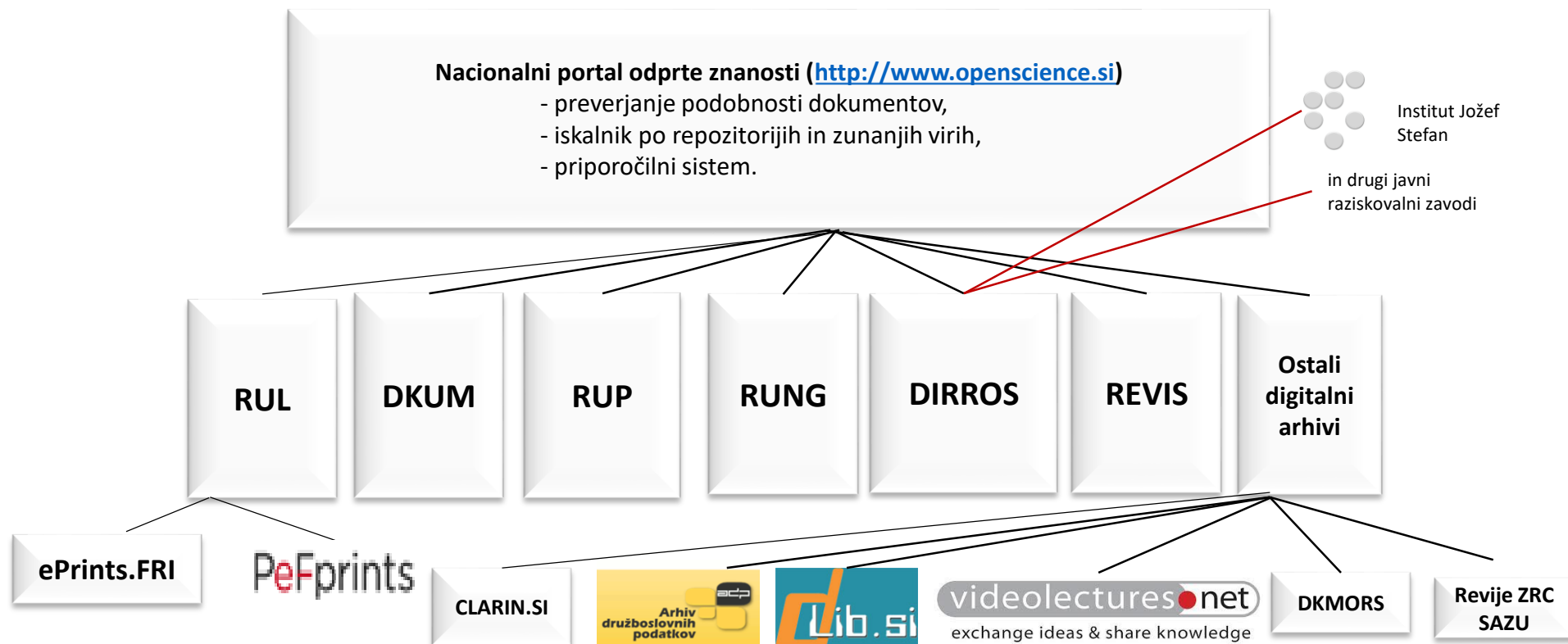


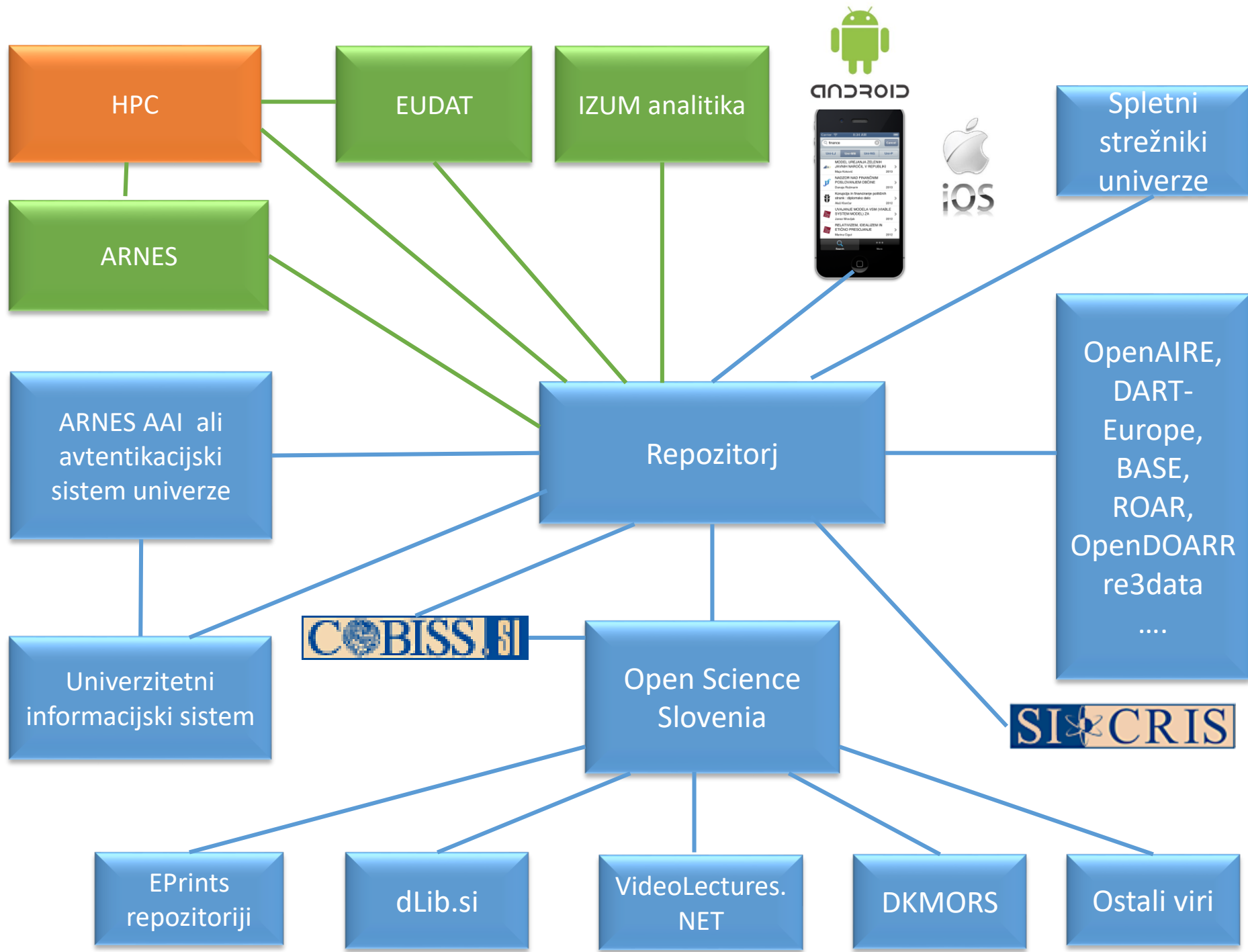
Strojno berljiva enota znanja
(angl. **knowledge unit**)

Zgradba digitalnega vzorca



Nacionalna infrastruktura odprtega dostopa





EUDAT Collaborative Data Infrastructure (CDI)

Dostop in ponovna uporaba podatkov:

- hranjenje podatkov (B2SHARE in B2DROP)),
- iskanje podatkov (B2FIND),
- avtentikacija in avtorizacija (B2ACCESS).

Trajno hranjenje podatkov:

- Implementacija upravljanja s podatki (B2SAFE).
- Omogočanje uporabe trajnih identifikatorjev (B2HANDLE).

Omogočanje procesiranja podatkov:

- Prenos podatkovnih setov med vozlišči EUDAT (B2STAGE).

Upravljanje z občutljivimi podatki.

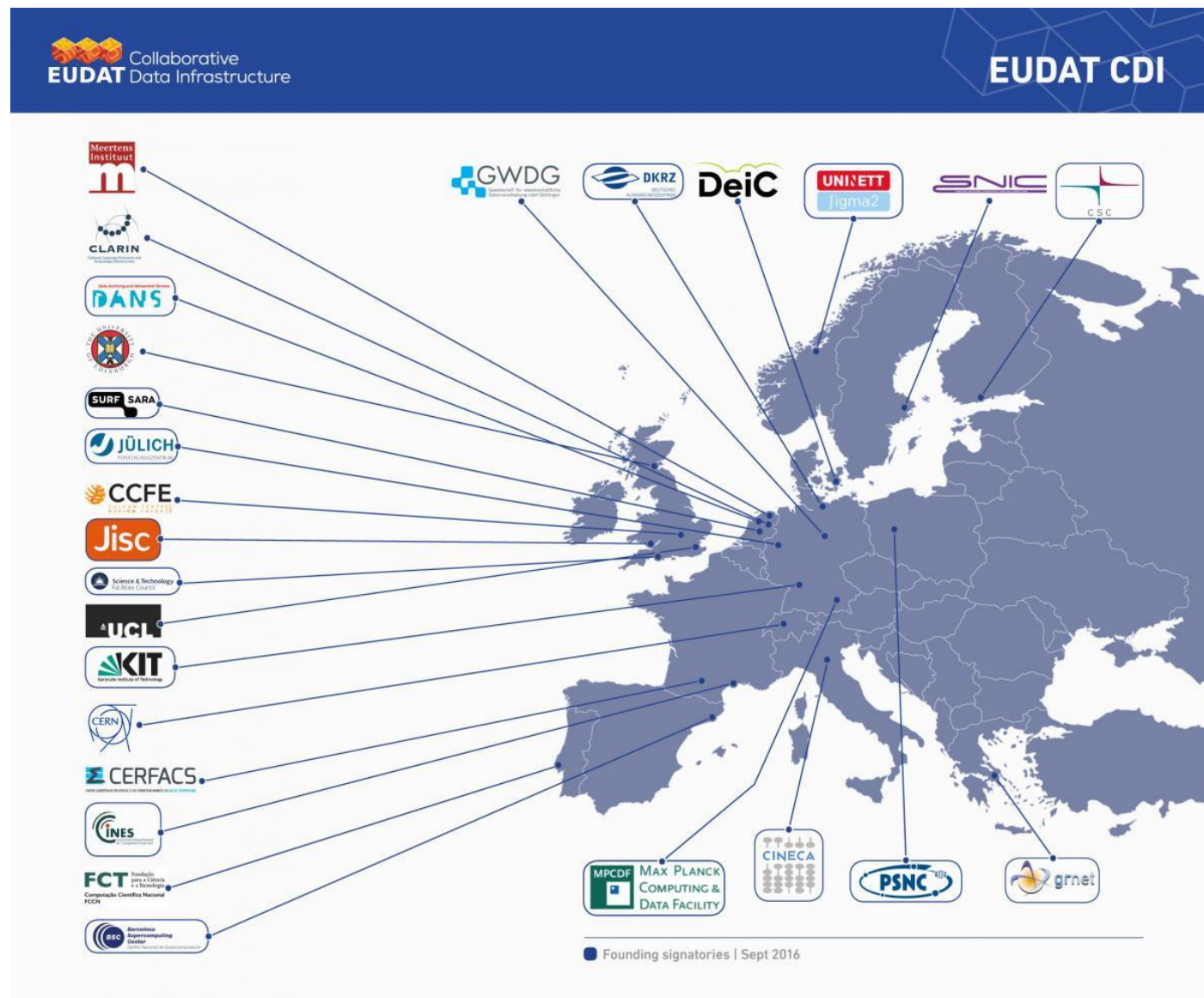
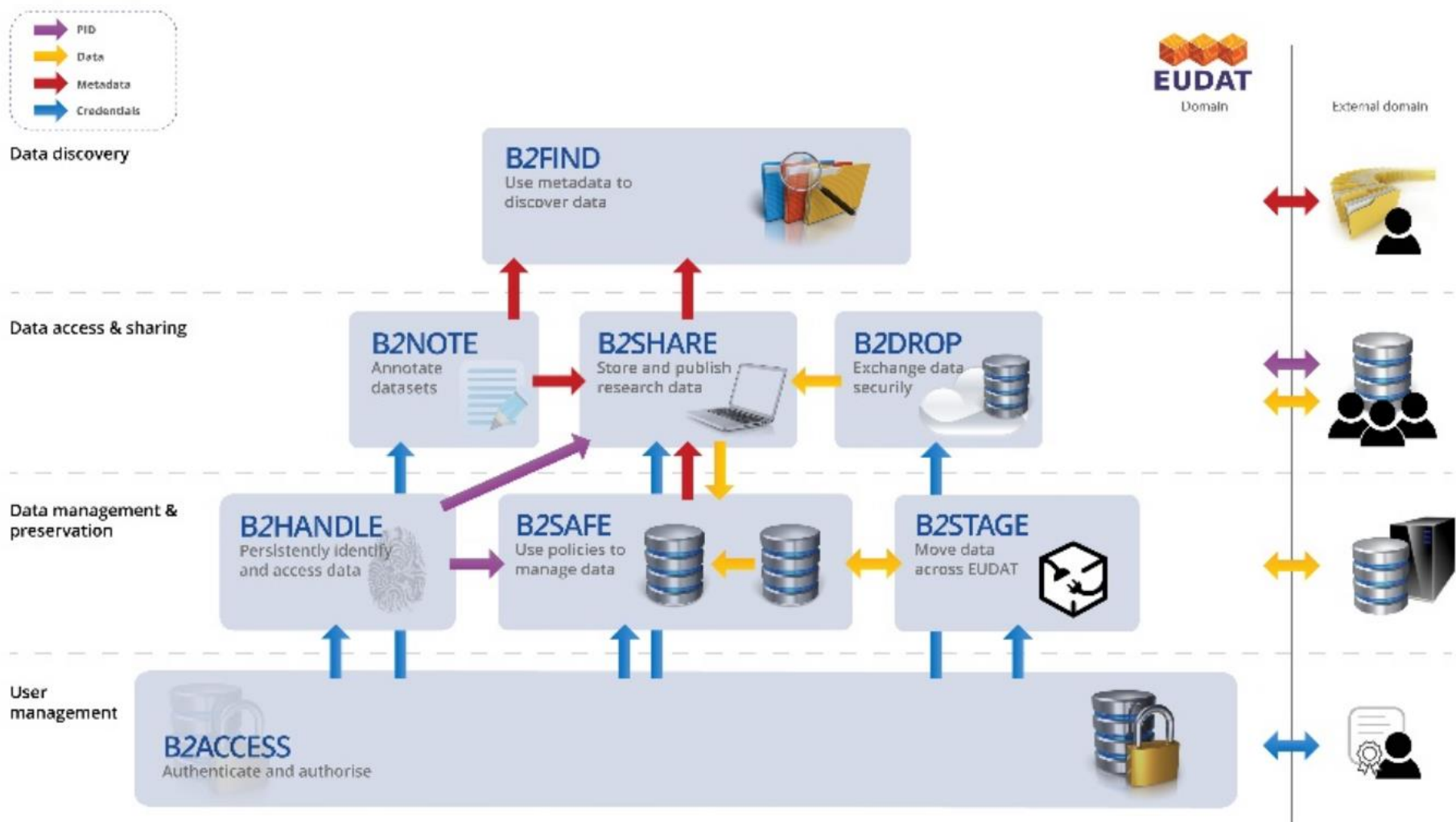


Diagram storitev EUDAT CDI



Aktivnosti dela projekta HPC-RIVR za vzpostavitev podatkovnega arhiva in trajnega hranjena velepodatkov

- Analiza obstoječega stanja raziskovalnih podatkov v Sloveniji.
- Preučitev možnosti trajnega hranjenja velepodatkov na infrastrukturi HPC in ARNES.
- Preučitev storitev EUDAT.
- Preučitev metapodatkovnih formatov za hranjenje raziskovalnih podatkov iz različnih znanstvenih področij.
- Izdelava dodatnih funkcionalnosti repozitorijev za potrebe podatkovnega arhiva, medopravilnosti z EUDAT in trajnega hranjenja velepodatkov.
- Izdelava učnih materialov in predlog za pravne podlage, navodila in procese.
- Izobraževanje uporabnikov.

Hvala za pozornost.

?