



FAIR principi datu pārvaldībā

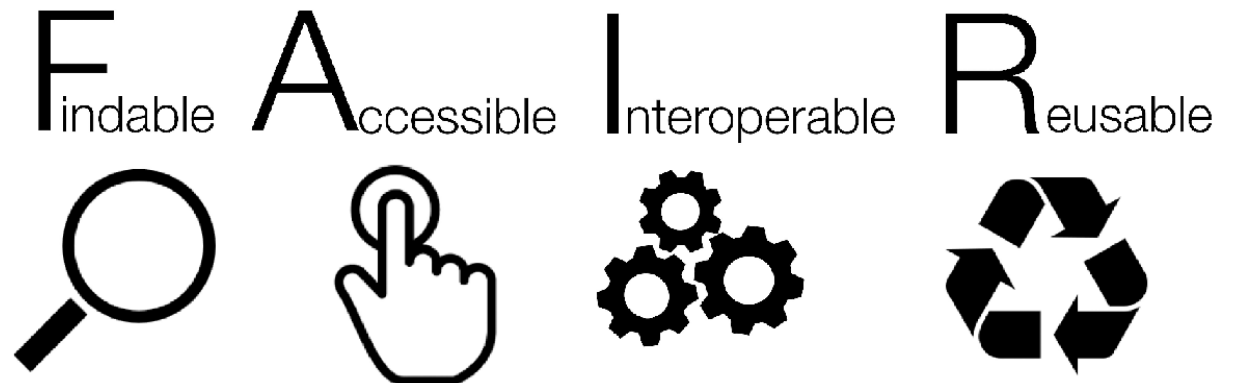
Lauris Cikovskis, Jānis Kampars
EOSC-Nordic RTU grupa

FAIR zinātnisko datu vadlīnijas

- 2016. gadā publicētas zinātnisko datu pārvaldības FAIR vadlīnijas
- Tiek sekmēta datorizēta datu apstrāde un minimizēta nepieciešamība pēc manuāla darba datu atrašanā un apstrādē

- Dati ir:

- Atrodami
- Pieejami
- Savietojami
- Atkāroti izmantojami



Atrodamība

- F1: (Meta) datiem ir piešķirts globāli unikāls un nemainīgs identifikators
 - Arī katram datu kopas elementam ir jābūt identificējamam
 - Datus ir vieglāk atrast, uz tiem atsaukties
 - Jānodrošina, ka cita datu kopai netiek piešķirts identisks identifikators
 - Piemēri - <https://orcid.org/0000-0003-0045-5593>, DOI: 10.1007/s10270-019-00756-7
- F2: Dati ir bagātīgi aprakstīti ar metadatiem
 - Iekļauj informāciju par kontekstu, kvalitāti u.c. datu raksturlielumiem
 - Sekmē datu atrodamību pēc metadatiem
- F3: Metadati ietver saistīto datu identifikatorus
 - Metadati un dati parasti ir atsevišķos failos
- F4: (Meta) dati ir reģistrēti vai indeksēti pārmeklējamā resursā
 - Metadati ir strukturēti un piemēroti datorizētai apstrādei
 - Nepieciešams radīt specializētus meklēšanas dziņus

Pieejamība

- A1: (Meta) dati ir iegūstami pēc identifikatora caur standartizētiem komunikācijas protokoliem
 - Sensitīvu datu gadījumā var norādīt kontaktpersonas e-pastu
- A1.1: Izmantotais protokols ir atvērts, bezmaksas un globāli implementējams
 - Jebkurš ar interneta pieslēgumu var piekļūt datiem vai vismaz metadatiem
 - Piemēri - Http(s), FTP, SMTP (SKYPE neatbilst)
- A1.2: Protokols nodrošina autentifikāciju un autorizāciju nepieciešamības gadījumā
 - FAIR dati nav obligāti atvērti un bezmaksas
 - Ir jābūt skaidriem nosacījumiem datu pieejamībai
- A2: Metadati ir pieejami arī tad, kad dati vairs nav pieejami
 - Datu pieejamības uzturēšana var būt dārga
 - Metadati ir vērtīgi arī tad, ja dati vairs nav pieejami
 - Iesaistītie pētnieki, eksperimentu plāns, eksperimentu atkārtošana utm.

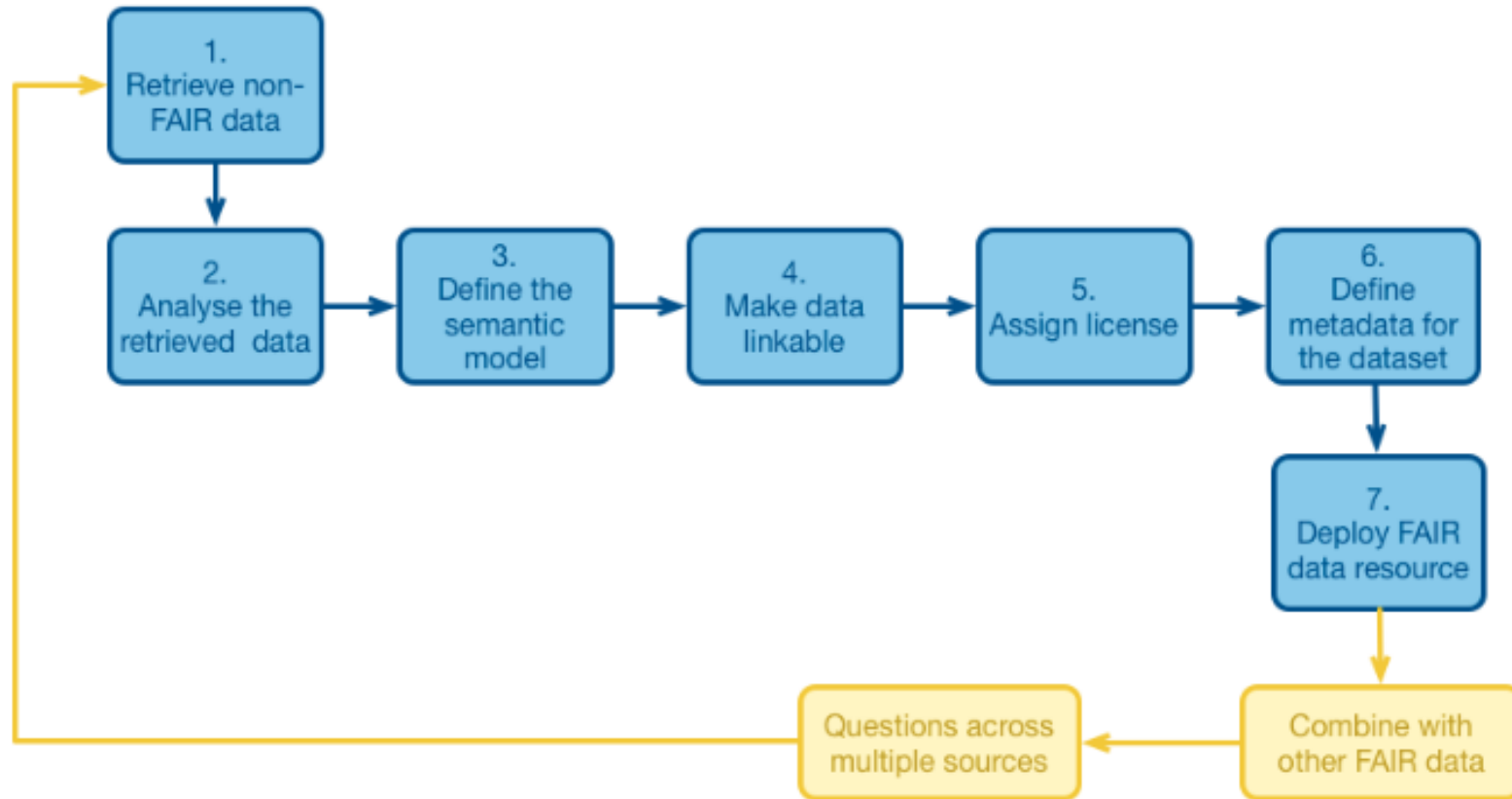
Savietojamība

- I1: (Meta) dati izmanto formālu, pieejamu, koplietojamu un plaši pielietojamu valodu zināšanu attēlošanai
 - Dati ir saprotami gan citiem pētniekiem, gan datorizētām sistēmām
 - Nav nepieciešams radīt specializētus algoritmus, tulkojumus, kartējumus
 - Tiek izmantotas kopīgas vārdnīcas, ontoloģijas, tezauri, labs datu modelis (apraksta datu struktūru)
 - Piemēri - <https://www.dublincore.org/schemas/>, <https://www.w3.org/OWL/>, <https://www.w3.org/TR/daml+oil-reference>, <https://json-ld.org/>
- I2: (Meta) dati izmanto FAIR atbilstošu vārdnīcu
 - Terminoloģija balstās uz vārdnīcu, kas arī atbilst FAIR principiem
- I3: (Meta) satur kvalitatīvas atsauces uz citiem (meta) datiem
 - Datu kopas saites ar citām datu kopām (pētījums izmanto datus no cita pētījuma)
 - Veicina atrodamību

Atkārtotā izmantošana

- R1: (Meta) dati ietvar precīzus un noderīgus atribūtus
 - Palīdz novērtēt, vai dati ir noderīgi noteiktā kontekstā
 - Apraksta, kādos apstākļos dati ir ievākti, eksperimentu plānu, datu kopas trūkumus, veikto priekšapstrādi utm.
 - Nemēģina prognozēt datu izmantošanas apgabalu, iekļauj pēc iespējas vairāk aprakstošo informāciju
- R1.1: (Meta) dati tiek izplatīti ar skaidru un pieejamu izmantošanas licenci
 - Ir jābūt saprotamam, kādos gadījumos datus drīkst izmantot
 - Piemēri - MIT un Creative Commons licences
- R1.2: (Meta) datiem ir skaidri zināma izcelsme
 - Datu ievākšana, priekšapstrāde, autors, saistītie dati – darbplūsma aprakstīta mašīnlasāmā formātā
- R1.3: (Meta) dati atbilst konkrētā problēmapgabala kopienas standartiem
 - Datu kopas tiek standartizētas, izmanto vienus terminus, ir strukturāli līdzīgas
 - Piemērs - <http://www.psidev.info/miape>

FAIR-ifikācija



Attēls <https://www.go-fair.org/fair-principles/fairification-process/>

Demonstrācija

- Pārlūkprogrammā atverot DOI adresi <https://doi.org/10.5281/zenodo.3753193>, tiekam pārvirzīti uz lietotājiem paredzēto HTML versiju
- DOI adresē pieprasot JSON LD, iegūstam datorizētai apstrādei paredzētus standartizētus metadatus
 - `curl -Lv "https://doi.org/10.5281/zenodo.3753193" -H "Accept: application/ld+json"`

```
* Copying HTTP/2 data in stream buffer to connection buffer after upgrade: len=0
* Using Stream ID: 1 (easy handle 0x7f861f810400)
> GET /10.5281%2Fzenodo.3753193 HTTP/2
> Host: data.crosscite.org
> User-Agent: curl/7.64.1
> Accept: application/ld+json
>
* Connection state changed (MAX_CONCURRENT_STREAMS == 128)!
< HTTP/2 200
< date: Thu, 16 Apr 2020 08:28:41 GMT
< content-type: application/vnd.schemaorg.ld+json; charset=utf-8
< status: 200 OK
< cache-control: max-age=0, private, must-revalidate
< vary: Accept-Encoding, Origin
< etag: W/"b0aac7c8e14c709f7323d29028746b9a"
< x-runtime: 0.060220
< x-request-id: b5652ff3-ed47-46a3-894d-230704dc5e35
< x-powered-by: Phusion Passenger 6.0.4
< server: nginx/1.14.0 + Phusion Passenger 6.0.4
<
{
  "@context": "http://schema.org",
  "@type": "MediaObject",
  "@id": "https://doi.org/10.5281/zenodo.3753193",
  "identifier": [
    {
      "@type": "PropertyValue",
      "propertyID": "DOI",
      "value": "https://doi.org/10.5281/zenodo.3753193"
    },
    {
      "@type": "PropertyValue",
      "propertyID": "URL",
      "value": "https://zenodo.org/record/3753193"
    }
  ],
  "url": "https://zenodo.org/record/3753193",
  "name": "Is SARS-CoV-2 transmitted by speaking?",
  "author": [
    {
      "name": "Valentyn Stadnytskyi",
      "givenName": "Valentyn",
      "familyName": "Stadnytskyi",
      "affiliation": {
        "@type": "Organization",
        "name": "Laboratory of Chemical Physics, NIDDK, National Institutes of Health"
      },
      "@type": "Person",
      "@id": "https://orcid.org/0000-0002-7202-6846"
    },
    {
      "name": "Christina E. Bax",
      "givenName": "Christina E.",
      "familyName": "Bax",

```