

Semanttinen yhteentoimivuus

- 1. Taustaa
- 2. Yhteentoimivuuden ohjeiden muodostaminen palveluntarjoajalle, data-asiantuntijalle / kouluttajalle ja tutkijalle
 - 2.1. Ohjeet palveluntarjoajalle
 - 2.2. Ohjeet data-asiantuntijalle / kouluttajalle
 - 2.3. Ohjeet tutkijalle
- 3. Keskusteltavaa/pohdittavaa
- 4. Hyödyllisiä linkkejä*

1. Taustaa

Interoperability (EIF)

Technical interoperability: the ability of different information technology systems and software applications to communicate and exchange data.

Semantic interoperability: the ability of computer systems to exchange data with unambiguous, shared meaning.

Organisational interoperability: the way in which organisations align their business processes, responsibilities and expectations to achieve commonly agreed and mutually beneficial goals.

Legal interoperability: the broader environment of laws, policies, procedures and cooperation agreements needed to allow the seamless exchange of information between different organisations, regions and countries.

In the FAIR principles **Interoperability** contains three principles:

The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.

- **I1.** (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- **I2.** (Meta)data use vocabularies that follow FAIR principles
- **I3.** (Meta)data include qualified references to other (meta)data

2. Yhteentoimivuuden ohjeiden muodostaminen palveluntarjoajalle, data-asiantuntijalle / kouluttajalle ja tutkijalle

2.1. Ohjeet palveluntarjoajalle

Palvelun tarjoajalla tarkoitamme tässä esimerkiksi datakatalogin tai tutkimusdatapalvelun ja sanastojen kehittäjiä ja ylläpitäjiä.

I1: (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.	I2: (Meta)data use vocabularies that follow FAIR principles.	I3: (Meta)data include qualified references to other (meta)data.
---	---	---

- Huomioi käyttäjien tarpeet ja tieteenalan käytännöt, metadatatassa mahdollisimman laajasti ja datassa mahdollisimman tarkasti. Sanastoja voi käyttää rinnan useita: - suomalaisia/suomenkielisiä - kansainvälisiä - yleisiä - tieteenalakohtaisia - Koneluettavuus ja semanttinen yhteentoimivuus on oleellista, sillä käytössä sanastojen on oltava myös ihmisen ymmärrettävää. Riippuvuuksien ja relaatioiden pitää olla ihmiselle näkyviä. - Vastataan käyttäjien tarpeeseen alojen sanastojen käyttöönottoon joustavasti palveluissa esimerkiksi rajapintojen avulla. Käyttäjälle voi myös tarjota mahdollisuutta lisätä tunnisteita termille tai sanastolle. - Palvelun/palveluntarjoajan /taustajärjestelmän pitää pystyä tukemaan tutkijaa tuottamaan kuvailutietoja, joiden perusteella voidaan edelleen tuottaa semanttisesti yhteensopivaa metadattaa - Palveluntarjoaja käyttää jotain linkitetyn datan skeemaa (esim. RDF tai JSON LD) - Sanastojen tulee olla koneluettavia. Tieteenalakohtaiset sanastot tulee mahdollisuuksien mukaan ontologisoida ja ylläpitää kytkettynä ontologiapalveluun	- Käytetään versioituja, ylläpidettyjä, alakohtaisia ja yleisiä sanastoja ja ontologioita - Käytettävien sanastojen on itsessään oltava FAIR - koneluettavia ja dokumentoituja - Sanastojen ja niiden käsitteillä tulisi olla pysyvät tunnisteet ja niiden tulisi olla käytettävissä avoimien rajapintojen kautta - Aina kun mahdollista tulisi hyödyntää olemassa olevia FAIR-periaatteita noudattavia sanastoja. Uusia sanastoja ja tietomalleja rakennettaessa tulee kiinnittää huomiota, että ne julkaistaan (ja kuvataan) FAIR-periaatteita noudattaen. - Sanastot tulee toteuttaa niin, että niiden käyttäjät voivat osallistua niiden ylläpitoon esimerkiksi ehdotus- ja korjaustoimintojen avulla.	- Tunnista kohdeyhteistön tarpeet ja käytännöt ja seuraa niitä. Huolehdi vuorovaikutuksesta ja sisällön ja ratkaisujen kehittämisestä. - Käytä kansallisten ja kansainvälisten keskeisten tietovarantojen tuntemia pysyviä tunnisteita jos mahdollista. Huomioi lisenssi, laatu ja luotettavuus. - Kontrolloituja arvoja (esim. sanaston käsite tai toimijan tunniste) ja mahdollisuuksien mukaan pysyviä tunnisteita mahdollisimman paljon: - Kuvailevassa metadatatassa: lisäään käsitteisiin pysyviin tunnisteisiin perustuvia linkkejä muihin tietojärjestelmiin (esim. finto.fi). - Hyödynnä myös organisaatiolle, tekijöille ja muille toimijoille tarjolla olevia tunnisteita (esim. ORCID). Edistä kansallisten resurssien kehittämistä. - Rakenteellisessa metadatatassa: Metadatan tulisi koneluettavasti kuvata datan rakennetta. Käytä DataCiten relaatioita kuvaamaan esimerkiksi datasettien välisiä suhteita, esimerkiksi versioita. - Hallinnollisessa metadatatassa: Lisen ssit, käyttöoikeuksien rajoitukset ja niiden perusteet - Näitä tulisi tarjota palveluun integroituina ja käyttäjävälisinä toiminnallisuuksina: luokitteluja, sanastoja, PIDit. Linkitysten ja koneluettavuuden pitäisi käyttäjän kannalta syntyä automaattisesti. - Referenssimetadata = sallitut arvot = > koneluettavuus. Laajennuksia ja vapaatekstiä on myös hyvä sallia mutta omissa kentissään.
--	--	--

2.2. Ohjeet data-asiantuntijalle / kouluttajalle

I1: (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.	I2: (Meta)data use vocabularies that follow FAIR principles.	I3: (Meta)data include qualified references to other (meta)data.
- tue virallisten ja <i>de facto</i> standardien löytämistä, käyttämistä ja noudattamista - tue kontrolloitujen sanastojen ja ontologioiden käyttämistä, täydentämistä ja kehittämistä	- tunne sanastojen merkitys ja käyttö datojen kuvailu ja sisällön ymmärrettävyys datojen linkittyminen toisiinsa kuvailutermin välityksellä datojen haku tekoälyn käyttö johtopäätösten teossa - tunne keskeisimmät yleiset sanastot sekä semanttiset hakupalvelut ja osaa ohjaa kunkin erikoisanalan sanastopalveluiden äärelle - Lähituessa tieteenalasanastojen tuntijat	- Kouluttajan on tunnettava perusteet linkitetyn datan vaatimuksista ja vaikutuksista sekä ymmärrettävä tieteenalakohtaiset näkökulmat. - Datametriikan näkökulma - Linkitetty data lisää tutkijan ja tutkimuksen näkyvyyttä ja vaikuttavuutta sekä datan uudelleenkäyttöä - Linkitetty metadata: eri vaihtoehtojen, seurausten sekä oman alan metadastandardin tunteminen - Yhdessä tutkijan kanssa asioiden selvittäminen: voisiko tämä sopia teille, että tiedosta tulisi yhteentoimivampaa - Ontologioissa on implisiittistä oletusta tieteenaloista, mikä on haastavaa - Metadatan kuraointi (tarkistus ja täydennys)

2.3. Ohjeet tutkijalle

I1: (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.	I2: (Meta)data use vocabularies that follow FAIR principles.	I3: (Meta)data include qualified references to other (meta)data.
--	--	--

• Harvemmin tutkijan käsissä, usein kiinni tarjolla olevista palveluista • Huomioi datan keräämisessä, luomisessa ja julkaisemisessa määrämuotoiset muuttujat • Suosi avoimia metadataformaatteja ja tiedostomuotoja kun mahdollista	• Pyydä tarvittaessa tukea sopivan sanaston löytämisessä datatuesta tai palveluista tai muilta tutkijoilta. Jaa tietoa. • Tutkijan olisi hyvä antaa palautetta, jos sanasto ei ole riittävä tai kaipaa muutoksia. • Mitä rikkaampi kuvailu niin sen parempi /kuvailun rikastaminen pitkin matkaa • Huomioi, että (meta)tietojen säilyttämispaikka voi vaikuttaa sanastojen käyttämisen mahdollisuuteen • Oman alan sanaston hallinta osa tutkijan ammattiosaamista • Kerrottava, mihin sanasto perustuu. Dokumentaatiosta ja metadatasta tulisi käydä ilmi käytetty sanasto.	• Tietää mistä saa apua ja osaa valita tallennuspaikan datalle, joka tarjoaa linkitetyn datan palveluita • Osa vastuullista tiedettä, koskee kaikkia tieteenaloja • On tietoinen/kiinnostunut asiasta • Osaavat vaatia käyttäjäystävällistä ja ymmärrettävää palveluja • Tutkimuksen tekemisen dokumentointi ◦ Metadatasta viittauksia sekä eri data-aineistojen välille sekä muuhun metadataan • Versionhallinta
--	---	---

3. Keskusteltavaa/pohdittavaa

- Mitä semanttinen yhteentoimivuus edellyttää?
- Millä tavoilla sitä voi edistää?
- Mitä hyviä esimerkkejä ja toimintamalleja on olemassa?
- Mitä hyvä resursseja on käytettävissä?
- Eri aloilla on erilaisia odotuksia - puhuttava myös riskeistä ja hyvä sisällyttää ne ohjeistuksiin
- Minimivaatimuksen täyttäminen tärkeintä ns. nyrkkisääntöjen luominen
- Tarvitaan ohjeistusta, mitä kannattaa tehdä ja mitä ei
- Suomenkielinen sana semantic artefactille?
- Mitkä ovat valintakriteerit tutkijan näkökulmasta sanaston valinnalle? Onko organisaatiolla resursseja/kapasiteettia antaa neuvontaa tässä tapauksessa?
- Ohjeiden kohtien katsominen pitkäaikaissäilytyksen näkökulmasta ennen kuin laitetaan eteenpäin
 - Data voi olla todellā linkitettyä mutta PAS-näkökulmasta käyttökeltvotonta

4. Hyödyllisiä linkkejä*

- FAIRsFAIR D2.5 : 2nd set of Recommendations for FAIR Semantics <https://doi.org/10.5281/zenodo.4314321>
- PID architecture for the EOSC, DOI: 10.2777/525581, see <https://hdl.handle.net/10.2777/525581>
- Sanastorekistereitā, esim. BARTOC (<https://bartoc.org/>) ja LOV (<https://lov.linkeddata.es/dataset/lov/>)
- Esimerkkilista ontologioista: https://www.w3.org/wiki/Lists_of_ontologies
- BioPortal, repository of biomedical ontologies: <https://bioportal.bioontology.org/ontologies>
- The Ontology Lookup Service (OLS), repository for biomedical ontologies: <https://www.ebi.ac.uk/ols/ontologies/uo>
- Linkitettyā dataa tehdāān FAIR Data Pointilla: <https://www.fairdatapoint.org/>, keskeiset formaatit JSON-LD ja rdf, ontologiana DCAT2

* Avoimen tieteen kevāttyöpajassa 18.3. sovittu, ettā tehdāān lista ontologioista/niitā kokoavista palveluista