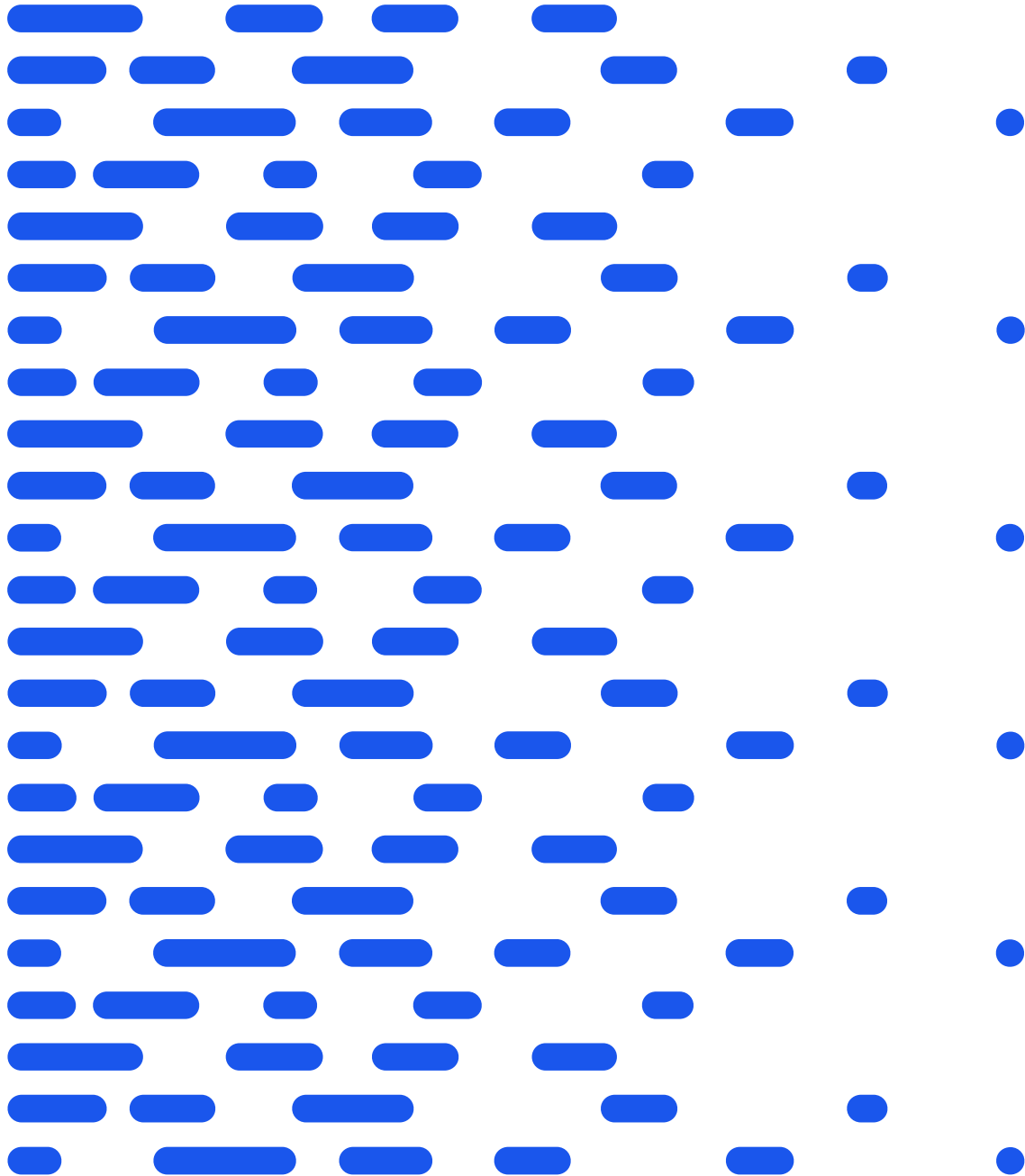


Tiedon laadun yhteistyöverkosto



Ohjelma

- Miika Malin, Oulun yliopisto:
Koneoppiminen ja tekoälyn
opettaminen: tiedon laadun merkitys
- Antti Helin, Valtiovarainministeriö:
Julkisen hallinnon automatisointi ja
lainsäädäntö
- Janika Tarkoma, Tilastokeskus:
Katsaus tiedon laatu näkökulmiin
koneoppimisessa

Tiedon laatukriteerien taustalla olevia periaatteita

- Euroopan yhteentoimivuusperiaatteet (EIF)
- FAIR-periaatteet
- Tilastoalan käytännösäännöt (Code of Practise)
- ISO/IEC 25012



Tiedon laatukriteerit verrattuna ISO 25012 kriteereihin

Laatukriteerien ISO 25012 vastaavuudet		
Tietoaineiston laatukriteeri	ISO/IEC 25012	Vastaavuus
Tarkkuus	Accuracy	Läheinen
Kattavuus	Completeness	Melko läheinen
Oikea-aikaisuus	-	-
Ajantasaisuus	Currentness	Melko läheinen
Metatietojen ymmärrettävyys	Understandability	Läheinen
Suosittelun mukaisuus	Compliance	Melko läheinen
Koneluettavuus	Portability	Melko läheinen
Virheettömyys	-	-
Johdon mukaisuus	Consistency	Erittäin läheinen
Alkuperäisyys	Traceability	Läheinen
Käyttöoikeudet	-	-
Lisätietoja tiedon laatukriteereistä verkkosivuilla		



Koneoppimisessa käytettävän tiedon laatu

- Standardia kehitetään ISO/IEC 25024 pohjalta (ISO/IEC 25012 ”mittaristo”) ja siihen on lisätty näkökulmia, jotka tukevat nimenomaan koneoppimisnäkökulmaa:
 - Mm. tehokkuus, tasapainoisuus ja monimuotoisuus, relevanssi ja samankaltaisuus
- Koneoppimisessa tarkastellaan montaa tietoa-aineistoa, jotka eri käyttötarkoituksessa:
 - opettamiseen ja laadun varmistamiseen liittyviä tietoa-aineistoja varsinaisen tuotannossa tarvittavan tiedon lisäksi



Koneoppimisen laatustandardi (tavoite 1.1.2024)

- Laatutarkastelut osittain samoja kuin ISO 25024
 - Tiedon laatukehikosta tuttuja mittareita mukana: kattavuudesta erityisesti puuttuvuustarkastelut, tarkkuuden, johdonmukaisuuden, metatietojen ymmärrettävyyden, suositustenmukaisuuden ja oikea-aikaisuuden mittareita muistuttavia mittareita
- Tarkastelussa on mukana tiedon relevanssin ja teknistä tarkastelua.
 - Tiedon käsittelyn tehokkuuden varmistamiseksi turhia tietoja tai vääriä formaatteja halutaan välttää
 - Metatietoja käsitellään yksityiskohtaisesti
 - Tarkastellaan tarkemmin myös muuta kuin rakenteista dataa, esim. kuvaformaatit
 - Koneluettavuus korostuu



Tiedon laatukehikko vs koneoppimisen laatuäkökulmat

- Tiedon laatukehikolla tavoitellaan ensisijaisesti julkishallinnon tietojen parempaa hyödynnettävyyttä
- Riittävän selkeä aloitus tiedon laadun tarkastelemiseen ja kehittämiseen
- Koneoppimisen tarpeet selvästi yksityiskohtaisemmat – kohtaavat osittain myös yleisen tiedon laatuarkastelun kanssa



Kiitos!

