



Aalto University
School of Science
and Technology

Suomen historia ontologiana

– CIDOC CRM -perustainen kulttuuridatan harmonisointi

Esko Ikkala

School of Science and Technology
Department of Media Technology

University of Helsinki
Department of Computer Science



UNIVERSITY OF HELSINKI



Suomen historiaontologia (HISTO)

- Perustuu [Agricola](#) –yhteisön verkkosivuston aikajanaan
- Historiaontologian tarkoituksena on:
 1. Määrittää ontologiset tunnisteet (URI) historiallisille tapahtumille
 2. Toimia linkittävänä ontologiana erilaisten kulttuuristen sisältöjen välillä
 3. Tarjota kattava ja hyvin kuvailtu tietokanta Suomen historian tapahtumista (itseisarvo)

[Lähde: Antti-Jussi Nygård: Suomen historiaontologian dokumentaatio]

HISTO - materiaali

- 1173 historiallista tapahtumaa, joihin on linkitetty:
 - 921 henkilöä
 - 440 organisaatiota
- lisäksi tapahtumiin on linkitetty satoja jo muista tietokannoista löytyviä henkilöitä, organisaatioita, paikkoja ja muita instansseja

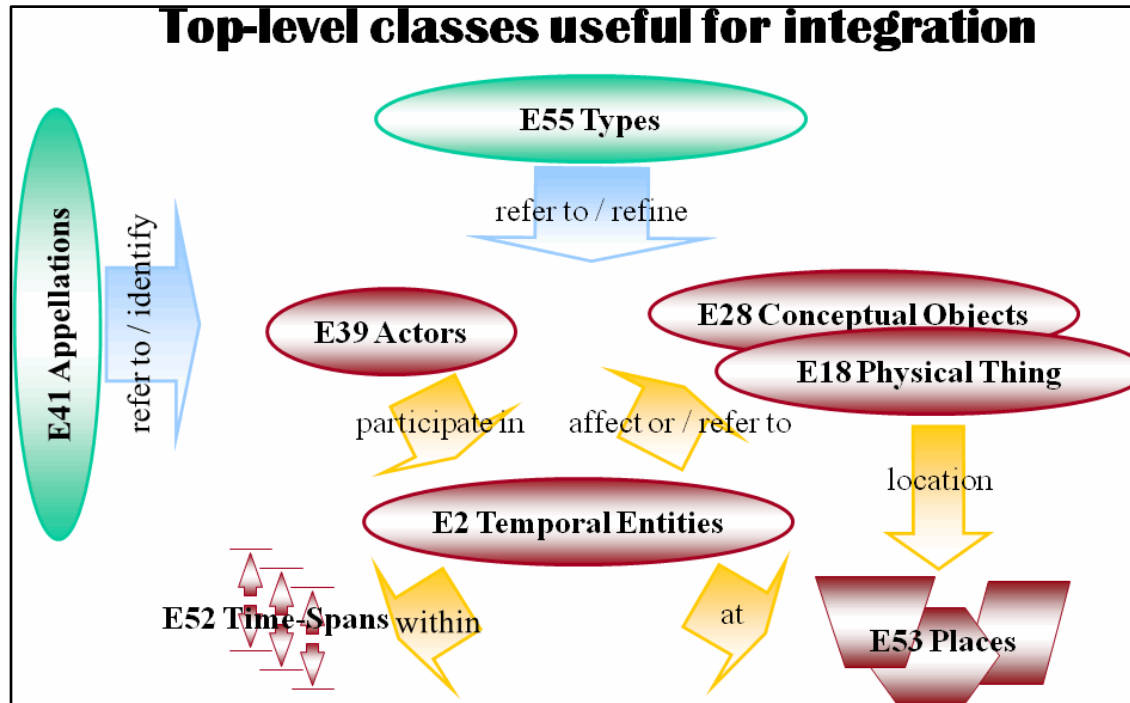
DIPLOMITYÖN TUTKIMUSKYSYMYKSIÄ

- Miten isoissa dataseiteissä oleva tieto saadaan hyödylliseen käyttöön?
- Mitä erityispiirteitä heterogeenisen kulttuuritiedon julkaisemisessa ja käyttöönoton tukemisessa on?

⇒ Historiallisten tapahtumien (HISTO) sekä museoesineiden tietojen (MuseoSuomi) muuntaminen CIDOC CRM -muotoon ja julkaisu LDF.fi -portaalissa

CIDOC Conceptual Reference Model (CRM)

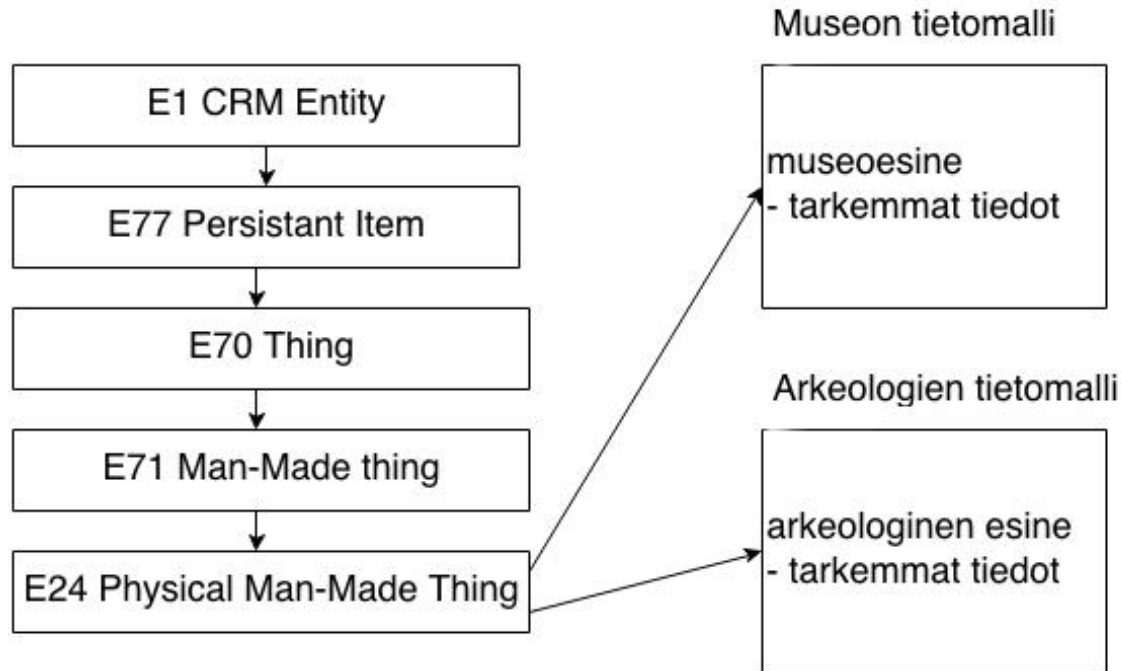
- Käsitelmä muistiorganisaatioiden (museot, kirjastot, arkistot, jne.) tuottaman heterogeenisen kulttuuritiedon yhteensovittamiseen
- Kansainvälinen museoneuvoston (ICOM) alaisen dokumentointikomitean yli 10 vuoden työn tulos
- ISO-standardi (ISO 21227) vuodesta 2006



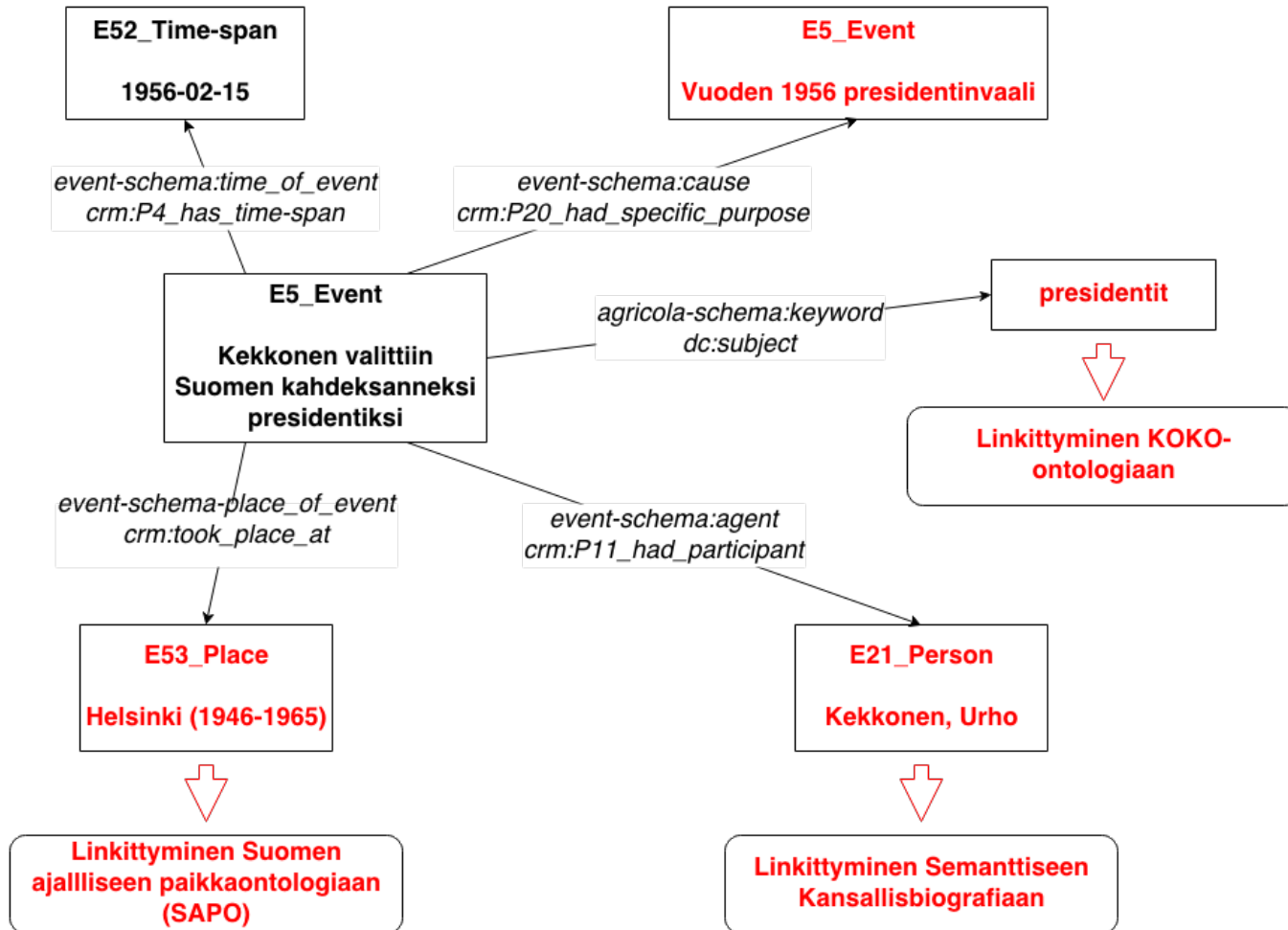
- CIDOC CRM –käsitemallin keskiössä ovat tapahtumat (Temporal Entities)
- Yhteensä 86 luokkaa ja 137 ominaisuutta kulttuuritiedon kuvailuun

CIDOC Conceptual Reference Model (CRM)

- Käsitelmallin tarkoituksena ei ole tyrkyttää museoille ja kirjastoille uutta tapaa tallentaa tietoa, vaan toimia ”semanttisena liimana” eri tietolähteiden välillä



HISTOn muuntaminen CIDOC CRM -muotoon



HISTOn muuntaminen CIDOC CRM -muotoon

- Alkuperäinen HISTO oli valmiiksi RDF-muodossa, joten työkaluina muunnoksessa Apache Jena -kirjasto sekä SPARQL-kyselyt
- Muunnoksen lisäksi historiaontologian sisältämä luokittelu erilaisiin tapahtumatyyppeihin tuotiin paremmin esille
- Haasteita:
 - alkuperäisen datan virheet, skeeman puute
 - kuinka dokumentoida muunnos?
=> taulukko mäppäyksistä tuntuu selkeimmältä

HISTORIA JA HENKILÖHISTORIA YHDESSÄ

- Myös Semanttinen Kansallisbiografia (6000 henkilöä) on muunnettu CIDOC CRM –muotoon SeCo-ryhmän toimesta
- Noin tuhannelle Semanttisen Kansallisbiografian henkilölle saatiin muodostettua yhteys historiaontologian tapahtumiin:
 - *tapahtuma <--> henkilö <--> henkilön historia*
- Lisäksi historiaontologian 921 henkilöä linkitettiin Semanttisen Kansallisbiografian henkilöihin niiltä kohdin kun vastaavuuksia löytyi

HISTO-DATASETIN JULKAISU

- Julkaistu [LDF.fi](http://ldf.fi) –portaalissa yhdessä Semanttisen kansallisbiografian kanssa
- Linked Data –selailu ihmiskäyttäjille:
<http://demo.seco.tkk.fi/ssaha/project/index.shtml?model=history>
- SPARQL-endpoint: <http://ldf.fi/history/sparql>
- Automaattisesti generoitu [dokumentaatio](#) datasetille saatavilla Miika Alosen Vocab.at - palvelussa

TYÖ JATKUU

- Suomen historian tapahtumien ja Kansallisbiografian henkilöiden yhteyksien parantelu sekä näiden yhteyksien visualisointi kartalla
- Museoesineiden CIDOC CRM –muunnoksen viimeistely ja julkaisu LDF.fi –portaalissa
- CIDOC CRM –muunnosten dokumentointi ja evaluointi diplomityössä

Kiitos mielenkiinnosta!