

Hallintoa vai hallintaa?

Metatietomallit ja kokonaisarkkitehtuuri 19.3.2013

Antti Auer

Tietohallintopäällikkö, HT

Jyväskylän yliopisto

Strateginen kehittäminen

Hallintoa

- JY:ssa kaksi tutkimusaineistotyöryhmää
- Hallinto on lähtökohtaisesti pahaa = byrokratiaa
 - Tutkijoita ei saa rasittaa hallinnolla
- Ainakin ns. vastuulliset tutkijat ymmärtävät, että jonkin verran hallintoa tarvitaan – säästytään työltä ja sähläykseltä myöhemmin
- Sopimukset heti tutkimushankkeen alkuvaiheessa!
 - Työryhmässä laadittiin seitsemän mallisopimus pohjaa.
- Metatiedon tulisi tietää käyttöoikeudet ja aineiston omistaja sopimusten perusteella
- Myös tietohallinto näyttäytyy hallintona...

TTA-minimimetatietomalli



nro elementin nimi määrä

1. Metatiedon tunniste 1
2. Metatiedon muokkaamisaika 1
3. Aineiston tunniste 1*
4. Muokkaamisaika 1
5. Jakelijan yhteystieto 1*
6. Aineiston nimi 1
7. Käyttöehdot 1*
8. Aihe 1*
9. Toimija 1*
10. Aineiston kieli 0*

Tuottaminen mihin viittaa

- t/k linkki (PID)
- t/k ajankohta
- t/k+y linkki (PID)
- t/y ajankohta
- t/k+y FOAF-kuvaus
- y(t) merkkijono
- y METS Rights -kuvaus
- y(t) käsitelinkki/merkkijono
- t/k+y linkki (PID)
- y ISO 639 -kielikoodi

Hallintaa

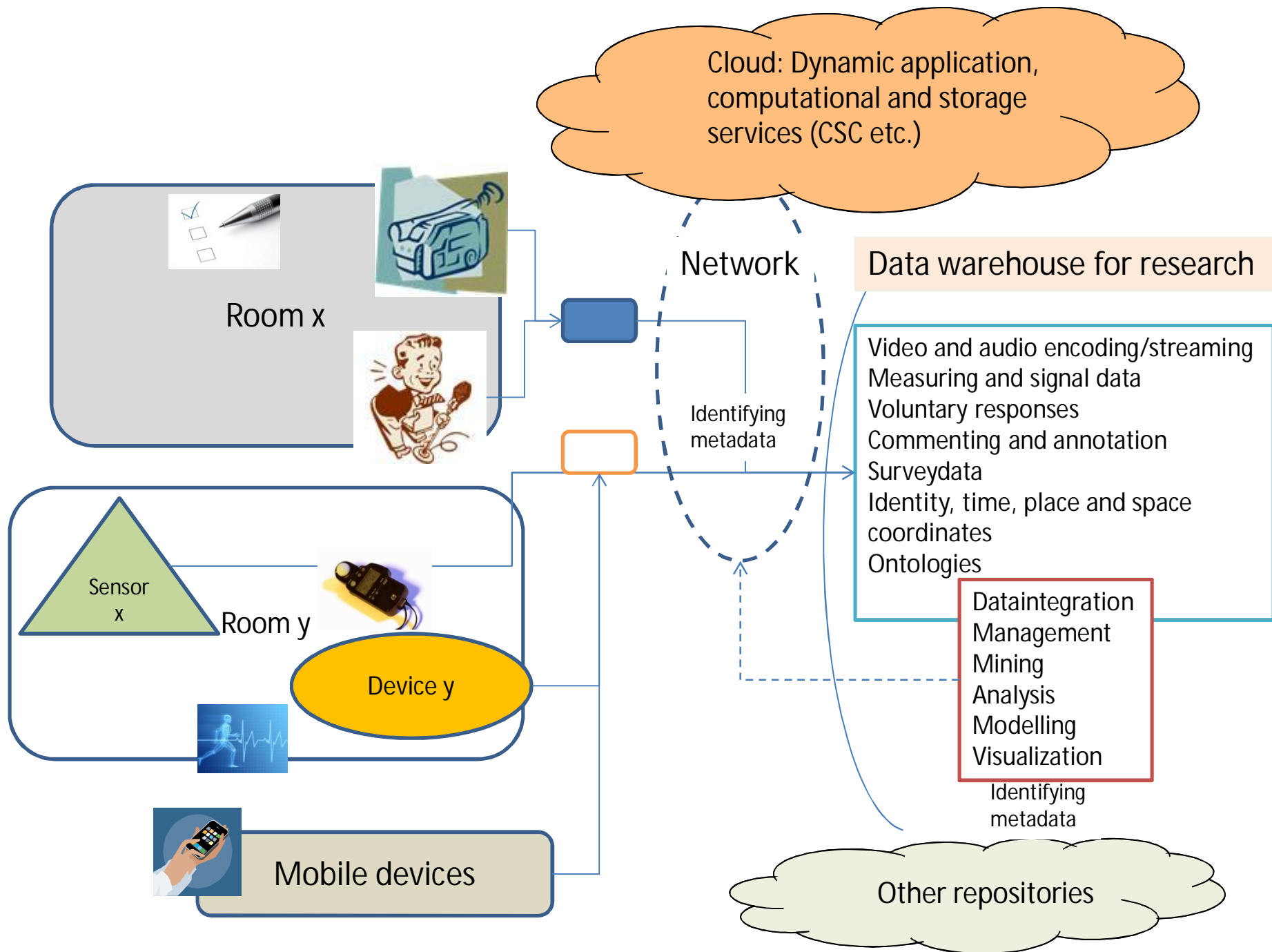
- Tutkijoita kiinnostaa tutkimuksen aikainen aineistojen hallinta – välitön hyöty tutkimushankkeelle
 - Siltä osin aineistojen hallintaa pidetään tärkeänä
- Arkistointi, uudelleenkäyttö ja pitkäaikaissäilytys – ok, jos ei aiheuta lisätyötä
- Jos tutkimuksen aikainen aineistojen hallinta toimii, loppusijoitus hoituu kyllä – jälkikäteen hoidettuna hyvin hankalasti, jos ollenkaan
- Aineisto tulee normaalisti avoimeksi vasta sitten, kun se on hyödynnetty
 - Tärkein uusiokäyttökohde voi olla vertaileva tutkimus tai pitkittäistutkimus – tutkimuksen menetelmät ja käytännöt kuvattava tarkasti (so. ei vain aineisto, vaan miten se kerättiin, miten analyysit tehtiin etc.)

Oxfordin yliopiston tutkimusdatahanke

<http://www.ariadne.ac.uk/issue65/wilson-et-al>

Data management sharing plans	Legal & ethical	Best formats & best practice	Secure storage	Metadata	Access & discovery	Computation analysis & visualization	Restricted sharing	Data cleaning	Publication	Assess value	Preservation	Add value
Support												
Infrastructure and tools												
Policy												
Business model												

"Data repositories tend to think of it primarily in terms of preservation and curation, but researchers are more likely to associate the term with structuring data in a database, or the organisation of files and folder. At its broadest, research data management involves all the processes that information from research inputs undergoes as it is manipulated and analysed en route to becoming a research output."



Miksi keskitetty datainfrastruktuuri?

- Työryhmän suositusten pohjalta päätettiin perustaa keskitetty aineistopalvelu
- Tutkimusresurssit kohdistettava tutkimukseen, ei tukipalveluihin
- Tutkimuksen peruspalvelut hankkeille automaattisesti
- Tietoturva ja aineistojen säilyminen – aineistojen varmuuskopiointi automaattisesti
- Tietosuoja – ihmistieteissä käsitellään usein yksilötason dataa
- Tutkimuksen aikainen aineistojen käytettävyys
- Aineistojen ja tutkimuksen tulosten auditoitavuus
- Aineistojen uudelleenkäyttö paikallisesti
- Pääsynhallinta roolien ja oikeuksien mukaan
- Virtuaaliset tutkimusympäristöt ja eScience
 - Aineistojen yhdistäminen tutkimushankkeen sisällä ja tutkimushankkeiden välillä
 - Aineistojen jakaminen yli organisaatorajojen
 - Uudet tallennustavat ja tietokantapalvelut, uudet analyysimenetelmät
 - Pohja keskitetyille laskenta-, mallinnus- ja visualisointipalveluille
 - Pohja ambient –teknologioille ja –älykkyydelle
 - Datamassojen syntyminen ja uudet analyysimenetelmät
- Tallennuskapasiteetin hallittu ja läpinäkyvä ulkoistaminen
- Pitkäaikaissäilytyksen ulkoistaminen
- Aineistojen huoltaminen ja Open Data
- Hyvin hoidettu aineisto kasvattaa impact factoria ja se meritoi tutkijaa
- Science 2.0 julkaisemisessa ja tutkimusverkostojen muodostumisessa
 - Tulosten julkaiseminen ei jatkossa välttämättä tarkoita julkaisua
- **KAIKILLE TUTKIMUSRYHMILLE/-HANKKEILLE peruspalveluna automaattisesti keskitetty levytila + virtuaalipalvelimet**

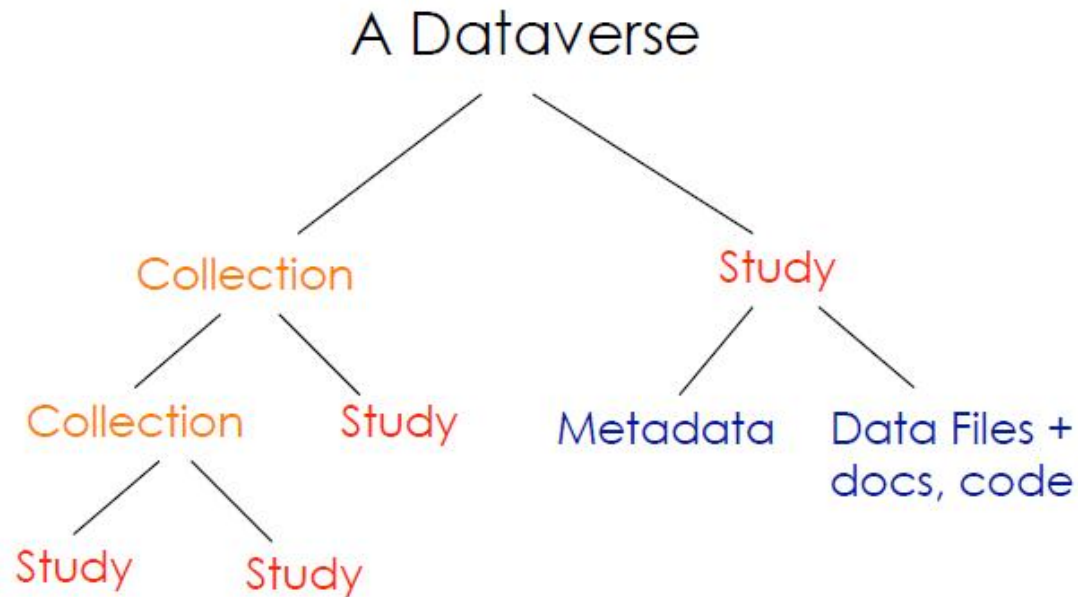
Hallintakerros uutena asiana

- Julkaisuarkisto/DSpace – digitaaliset objektit
- Youdata ja Yousource (IT-tdk)
- iRODS – prosessien automatisointi
- CKAN – avoin julkaiseminen
- Dataverse – tilastolliset aineistot
- Ala-/aineistospesifit hallintaohjelmistot
 - LIMS – laboratorio
 - CDMS – kliininen data
 - etc.
- Laadulliset aineistot
 - Esim. Atlas.ti: "hermeneuttinen yksikkö"
- Metatietomallit: DS, DDI, aineistospesifit....

Esim. Dataverse

- Dataverse network: A centralized software installation and data repository
- " The Dataverse Network Project standardizes the citation of data sets. Until this Project, citations of data were inconsistent or nonexistent in many publications, with future access and scholarly recognition highly uncertain. When you create a study in the Dataverse Network, the citation is generated and presented automatically."
- Persistent URL (Handle) for permanent reference + Universal Numerical Fingerprint (UNF) for verification
- Esim.:
Weisiger, Alex, "Replication data for: Logics of War: Explanations for Limited and Unlimited Conflicts",
<http://hdl.handle.net/1902.1/18738> UNF:5:OJCPMDOPJ96QO9V7fhXJMA== V1 [Version]
- Tulevaisuudessa: Seamless integration between the Open Journal Systems and The Dataverse Network.

Viittaus seuraava rakennetta



Not only does this assure future researchers that they can use the same data set referenced in a years-old journal article, it enables the data set's owner to track each iteration of the owner's research. When an original data set is updated or incorporated into a new, related data set, the algorithm generates a unique UNF each time.

"Kentän näkökulma"

- Kokonaisarkkitehtuuri
 - Koko tutkimusprosessin vaiheiden hallinta
 - Perusmetatiedon + täydentävän (kontekstuaalisen, aineistospesifin...) metatiedon hallinta
 - Minimimetatietoa ei kannata laajentaa – yhteys täydentävään metatietoon luotava
- Prosessien täytyy olla mahdollisimman selkeitä – sekä yliopiston sisällä, että palveluja ulkoistettaessa
 - Aineiston tunniste ja metatieto syntyy (pitäisi syntyä) silloin kuin aineistokin
 - Hallintakerros tukee tutkimuksen tekemistä ja etenemistä
- Eri hallintajärjestelmien tuottama identifioiva/täydentävä (rakenteinen) metatieto otettava huomioon täysimääräisesti (contextual information, tracing provenance)
- Metatiedon tallennusprosessien täytyy olla selkeitä – jälkikäteen täydentäminen hyvin hankalaa
- "Data cleaning" saattaa olla vaikea vaihe – kenen vastuulla?
 - Avoimen datan/uudelleenkäytön kuten myös ulkoistuksen edellytyksenä; ei "digitaalisia pahvilaatikoita"
- Mistä metadatakatalogi saa/hakee tietonsa?
- Erilaisia ulkoistusvaihtoehtoja tarjolla (mm. Certian OpenText)
- Pitkäaikaissäilytys vielä oma ongelma-alueensa (mm. rahoitus).