



Suomalaiset julkaisuarkistot vuonna 2010

Kirjastoverkkopäivät, DSpace-sessio, 29.10.2010
Jyrki Ilva (jyrki.ilva@helsinki.fi)

Julkaisuarkistot, rajanvetoa

- OpenDoar (<http://www.opendoar.org>) listaa 15 suomalaista julkaisuarkistoa
- Päivitin ja muokkasin listaa jonkin verran, ja jäljelle jäivät seuraavat arkistot: Doria (ml. E-thesis), Hanken, Helda, Jyx, Oulu, TaY (3 arkistoa), Theseus, TKK, TTY, UEF (+ VTT, joka on enemmän julkaisurekisteri)
- Samantyyppistä sisältöä (lähinnä opinnäytteitä) lisäksi muutamassa muussakin paikassa: HKKK, TuKK, SibA, VY/Tritonia
 - Sisältävät vapaasti käytettäviä kokotekstejä + kuvailutietoja
 - Ei (ilmeisesti) OAI-PMH:ta, ei brändätty julkaisuarkistoiksi, joten rajasin pois
- Muita rajatapauksia: TSV:n Tieteelliset lehdet –palvelu

Organisaatiot ja ohjelmistoratkaisut

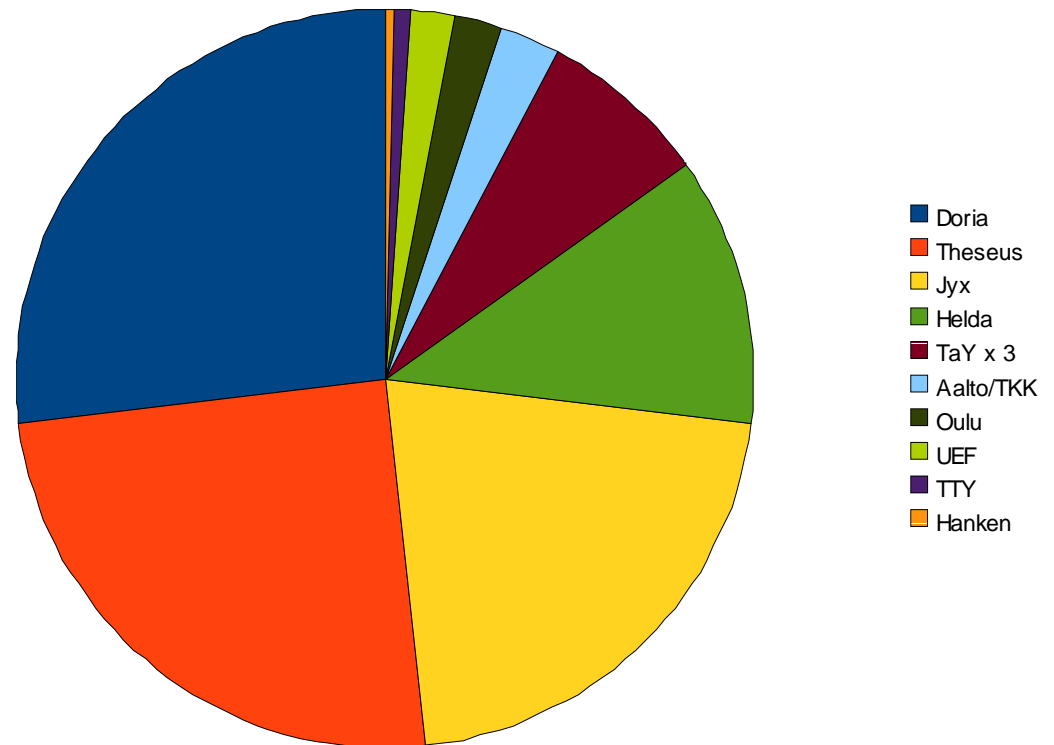
- Suomessa kaikkiaan yli 40 organisaatiota, jolla julkaisuarkisto
- Em. listassa kolme monen organisaation yhteistä arkistoa
 - Doria: HY, LY, LTY, SibA, Stadia, SLS, Tamk, TY, ÅA
 - Theseus: Kaikki amk:t (25 kpl)
 - Helda: HY, Evira, Ilmatieteen laitos, Kela, Suomen metsätieteellinen seura
- Julkaisuarkisto-ohjelmistot: nykytilanne
 - 6 x DSpace (Doria, Hanken, Helda, Jyx, Theseus, TTY)
 - 6 x itsekehitetty ratkaisu (Oulu, Tay x 3, TKK, UEF)
- Kehitys vienyt/viemässä kohti kansainvälisten julkaisuarkisto-ohjelmistojen käyttöönottoa
 - Lähitulevaisuudessa 7 x DSpace, 1 x Fedora, 4 x oma
 - Aalto/TKK rakentaa DSpace-arkistoa, Oulu Fedoraa

Julkaisuarkistojen sisältö

- Kävin läpi edellä listatut kymmenen julkaisuarkistoa ja keräsin tietoja niiden sisällön määrästä ja eri aineistotyypeistä
- Rajasin aineistot vapaasti saatavilla oleviin kokotekstitiedostoihin (ml. kuvat ja äänitteet)
 - Ei käyttöoikeudeltaan rajattuja aineistoja eikä julkaisuja, joista on saatavilla pelkät viitetiedot tai abstrakti
- Joissakin tapauksissa joissa määrien selvittäminen tai jaottelun tekeminen ei ollut helppoa, jouduin turvautumaan arvioon
 - Kaikkiaan em. arkistoissa n. 126.000 tietuetta, joista 60%:iin liitetty vapaasti käytettävä kokotekstitiedosto (tai tiedostoja)

Julkaisuarkistojen sisältö: aineiston määrä

- Avoimesti saatavilla olevien kokotekstiaineistojen (itemien) määrä julkaisuarkistoissa
 - Doria 19.500
 - Theseus 18.000
 - Jyx 15.500
 - Helda 8.500
 - TaY (3 arkistoa) 5.400
 - Aalto/TKK 1.900 (+ tuhansia väitöskirjojen osajulkaisuja)
 - Oulu 1.500
 - UEF 1.400
 - TTY 550
 - Hanken 300
- Yht. 72.500 kokotekstijulkaisua



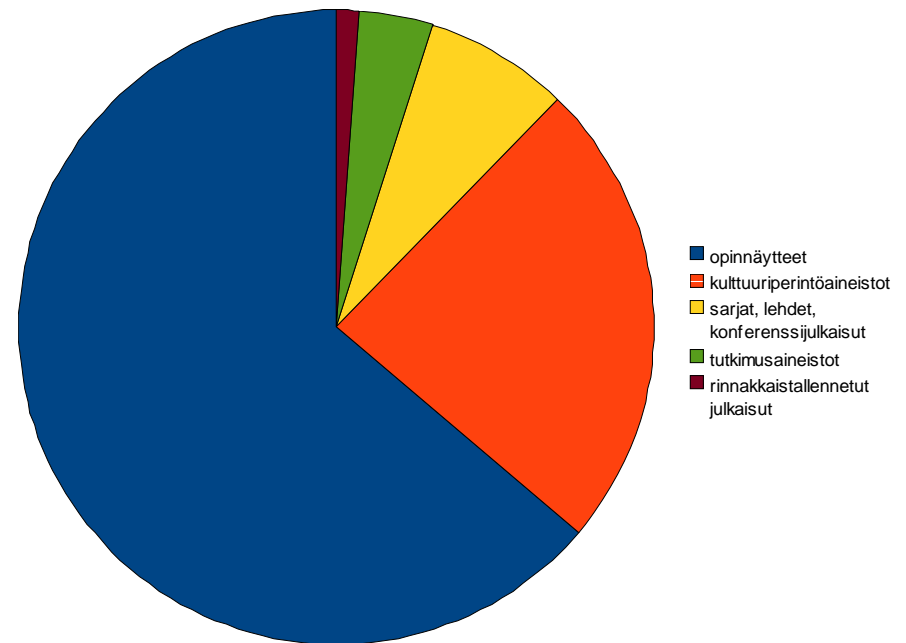
Aineiston määrä: pohdintaa

- Palveluita pyöritetty monin paikoin toistakymmentä vuotta: onko 72.000 OA- kokotekstijulkaisua siihen nähden hyvä tulos?
- Kaksi ryhmää: 4-5 isoa julkaisuarkistoa ja joukko pienempiä
 - Doria, Helda, Jyx ja Theseus isoja arkistoja, yhteenlaskettuina myös TaY:n arkistot kuuluvat samaan ryhmään
 - Loput selkeästi pienempiä (alle 2000 kokotekstijulkaisua)
 - TKK:n väitöskirjoihin tosin liitetty osajulkaisut, jotka voisi laskea myös erikseen (jolloin TKK olisi samassa suuruusluokassa TaY:n kanssa)
- Isot vs. pienet julkaisuarkistot – mitä preferoidaan?
 - Iso arkisto: ylläpidon kustannustehokkuus, koon tuoma näkyvyys, kaikki aineistot löydettävissä samasta paikasta
 - Pieni arkisto: organisaation oman itsenäisen työn tulos, oma brändi näkyy varmasti
- Julkaisujen määrä vs. julkaisujen ja palvelun laatu?

Julkaisuarkistojen sisältötyypit

- Karkea jaottelu kokotekstijulkaisujen sisältötyypeistä:
 - Opinnäytteet 46.000 (väitöskirjat 10.000; gradut, lisurit, kandintyöt ja amk-opinnäytteet yht. 36.000)
 - Kulttuuriperintöaineistot 17.300
 - Sarja- ja konferenssijulkaisut + lehdet 5.300
 - Tutkimusaineistot 2.700
 - Rinnakkaisjulkaisut 850

(TKK:n väitöskirjojen osajulkaisuja ei laskettu mukaan)



Huomioita julkaisuarkistojen sisällöistä

- Opinnäytteiden osuus musertava: yli 63% julkaisuarkistojen sisällöstä
 - Kaikki arkistot sisältävät opinnäytteitä
- Kulttuuriperintöaineistoja 24% sisällöstä
 - Tosin aineistotyyppiä vain kolmessa arkistossa (Doria, Jyx, Helda)
- Sarjajulkaisujen, lehtien ja konferenssijulkaisujen osuus 7%
- Tutkimusaineistoja vain yhdessä arkistossa (Helda), osuus 3%
- Rinnakkaistallennettujen artikkelien osuus edelleen minimaalinen, n. 1%
 - TKK:n väitöskirjojen osajulkaisut nostaisivat määrän n. 5%:iin
 - Joka tapauksessa nurinkurista siihen nähden miten keskeisessä roolissa rinnakkaistallennus on ollut arkistojen markkinoinnissa
 - Ovat olleet tulossa ”ihan kohta” vuodesta toiseen
- Yksittäisten arkistojen sisältöprofiileissa huomattavia eroja

Julkaisuarkistoaineistojen käyttö?

- Kokonaiskuvan saaminen vaatisi tarkempaa selvitystä, eikä yhdenmukaisia vastauksia ehkä saisi sittenkään
- Joitakin havaintoja KK:n ylläpitämien arkistojen (Doria ja Theseus) Simplestats- ja Google Analytics –tilastojen pohjalta:
 - Vuosittain miljoonia latauskertoja: vaikuttavaa, mutta jaettuna yksittäisten julkaisujen kesken puhutaan kymmenistä tai sadoista latauskerroista
 - Tieteelliselle julkaisulle / kulttuuriperintöaineistolle tämäkin ihan OK...
 - Tieteellisten aineistojen yleisö joka tapauksessa rajallinen
 - Käytön määrä vakiintunut, kasvaa lähinnä julkaisujen määrän kasvun myötä (jos sittenkään)
 - Valtamedian huomio saattaa hetkellisesti nostaa yksittäisten julkaisujen kävijämäärää, mutta menee nopeasti ohi