

Digitaalrepositooriumi DSpace juurutamisest Tartu Ülikooli Raamatukogus

Kalju Kill Kask

Tartu Ülikooli Raamatukogus juurutati 2006. a lõpus maailma üks enamlevinud digitaalse repositooriumi tarkvara DSpace (<http://dspace.utlib.ee>). Vajadus digitaalsel kujul andmete arhiveerimise vahendite järele tekkis varem. Praeguseks on dspace.utlib.ee maailma ülikoolide veebimeetrikas üsna hästi nähtav.

Digitaalsel kujul teabe arhiveerimisest enne DSpace'i juurutamist

- Väitekirjade arhiveerimise tarkvara oli vananenud.
TÜ Raamatukogu alustas Tartu Ülikoolis kaitstavate väitekirjade e-versioonide kogumist 2004. aastal, aluseks rektori poolt kinnitatud väitekirjade publitseerimise kord. Enne kaitsmist avaldati väitekiri raamatukogu koduleheküljel www.utlib.ee arhiveerituna veebiserveris. Pärast kaitsmist väitekiri kataloogiti e-kataloogis ESTER, kataloogikirjes näidati ka väitekirja e-versiooni URL. Tarkvara arendas välja andmebaasi tookordne administraator Ehtel Taevere. Paraku hakkas pärast veebiserveri tarkvara uuendusi tekkima probleeme, uuendamist vajas ka väitekirjade publitseerimise tarkvara.
- Veebiserveris ja failiserverites oli palju faile ning nende turvalisus oli nõrk.
Kõvaketastele oli kogunenud palju digiteeritud materjali pdf-, tiff-, jpg-vormingus. Failde haldamiseks puudus vastav liides, puudusid failide muutmise ja kustutamise kontrollimise vahendid. Kindlaks oli määratud, et tiff-failid on säilituskoopiad eraldi võrgukettal, nendest tehakse jpg-vormingus eelvaated, mis on kättesaadavad raamatukogu koduleheküljel www.utlib.ee ning need kirjeldatakse ESTERis. Aja jooksul hakati säilituskoopiate võrguketast kasutama ka ajutiste failide hoidmiseks, mille tulemusena jpg ja tiff failipuud ei peegeldanud enam teineteist.
- EEVA (Eesti vanema kirjanduse digitaalne tekstikogu).
Paljud tunnustasid ja kiitseid EEVA tarkvara, kuid digifailid pandi lihtsalt kõvaketale. Tarkvara arendajad vahetusid.
- Etteheiteid arhiveerimisele.
Varundamine toimus TÜ infotehnoloogia osakonna lindirobotitele, aga sellist moodust ei peeta korralikuks arhiveerimiseks. Puudus sisu muutumise kontroll, põhimõtteliselt polnud võimalik teada saada, kas üks 50 tuhandest failist on muutunud või kustunud.
- Pealtvalgustuskanneri hange.
TÜ Raamatukogus võeti vastu otsus suurendada digiteerimise jõudlust, käivitati pealtvalgustuskanneri hange. Skaneeritud materjali majutamiseks oli vaja tõsisemat lahendust.

Lahenduse valimine

Kuidas faile arhiveerida, kas tarkvara osta, luua ise või kasutada vabavara? Kus, millist ja millistel tingimustel?

Teabe arhiveerimine?

Tihti kõlab väide, Eesti on nii väike ja mitut süsteemi pole vaja, et suudame küll kokku leppida. Tegelikuses on nii, et iga süsteem on töövahend konkreetse ülesande jaoks, universaalset pole või kui on, siis piiratud võimalustega.

Rahvusraamatukogus oli arendamisel digitaalarhiiv DIGAR (digar.nlib.ee). Selle miinuseks oli kasutajaliidese puudumine ning repositooriumi reeglid — teave oleks pidanud vastama selle arhiivi komplekteerimispoliitikale ja edaspidi oleks tulnud arvestada teistsuguste ligipääsureeglitega. Tartu Ülikoolis oli olemas mitmeid süsteeme, mida saaks/saanuks kasutada repositooriumina: veebihaldustarkvara Webmedialt, dokumendihaldustarkvara Livelink (dok.ut.ee), fotode halduse tarkvara, videoservert (video.ut.ee). Ükski nendest lahendustest ei olnud sobiv. Vaadata võis ESTERi võimalike moodulite hankimise ja ETISE arendamise (näit. väitekirjade publitseerimine) ja teiste lahenduste poole (kasvõi Google). Isegi kui need oleksid funktsionaalselt sobinud, jäeti need kõrvale, sest arendusi polnud aega aastaid oodata — pealtvalgustusskänneri riigihange oli välja kuulutatud.

Oma tarkvara arendamine või tarkvara tellimine, valiku põhimõtted?

Koostööst erinevate tarkvaraarendajatega oli selge, et ise me ei suuda luua erinevatele standarditele vastavat tarkvara. Missugune peaks olema vastavus OAI-PHM, OAI-ORE, Z39.50, Dublin Core, MARC, XML, LDAP, XHTML, RSS jm standarditega, kuidas kontrollida failide õigsust? Hea, kui raamatukogu neid standardeid teab ja tunneb, raskem on selgitada arendajatele, mida vaja. Selge oli, et tarkvara elutsükli toetamine on erinevate arendajate ja nende vahetumise puhul nõrk. On möödapääsmatu, et veebiserverite baastarkvara ja riistvara uuendamisel tuleb uuendada ka repositooriumi tarkvara. Parimaks tarkvaratõe indikaatoriks on kasutajate arv, kasutajate foorumid, tarkvara sõltumatus kindlast tootjast ning tema probleemidest ja võimalikest krahhidest. Tarkvara peaks olema võimalik edasi arendada, muuta, sh kasutajaliideseid tõlkida. Samas tuleb omapoolse arendusega olla ettevaatlik — tarkvara standardsete uuenduste kasutamine muutub keerulisemaks. Tähtis on, et erinevate muudatuste korral teavikute veebiaadressid (URLid) säiliks (veebiaadressid võivad minna pikaks ja keeruliseks ning URL nihkub kirje lisamisel andmebaasi).

Konkreetse lahenduse valik

Internetis ringi vaadates selgus, et enamik kolleege maailmas kasutab vaba tarkvara. Pidime vaid välja selgitama, milline nendest on meie jaoks parim. Uurisime selleks võrdlustabeleid ja <http://www.openarchives.org> jm avaldatud ettekandeid valikute kohta. Lähema vaatluse ja

katsetamise alla võtsime Fedora (Fedora Commons), EPrints, DSpace, MyCoRe repositooriumitarkvarad.

Fedora tarkvara valivad suurte süsteemide ehitajad, kellel on arenduseks jõudu. Nimelt puudub Fedoral kasutajaliides, ta on üksnes repositooriumimootor. Tal on mitmeid serveripoolseid võimalusi, näit. pilte suurendada, vähendada, nihutada, mitmekesine objektide esitus, failide seostamine. Tarkvara installimine oli teiste süsteemidega võrreldes üllatavalt lihtne, aga kui see oli tehtud, oli raske aru saada, kus sisend, kus väljund. Tarkvara töötab Javal, valida saab andmebaasimootori. Kui lähtuda sellest, et metainfo on e-kataloogis ESTER ja vajame vaid andmehoidlat, siis oleks Fedora sobinud. Feodrat kasutab Rahvusraamatukogu digitaalarhiiv DIGAR (www.digar.nlib.ee); tarkvara toetab muuhulgas ka VTLS.

EPrints. Teavikud paiknevad teemade hierarhias, kuid üks teavik saab kuuluda vaid ühte hierarhiasse. Servesis kasutatakse PHPd, PERLi.

DSpace. Teavik asub kollektsioonis (collection), kollektsioon kuulub kommuuni (community), alamkommuuni (sub-community). Ühte teavikut saab linkida mitmesse kollektsiooni. Dspace võimaldab luua asutuse erinevate üksuste (raamatukogu, muuseum, teaduskond), teadlaste kollektiivide, konverentside jne jaoks oma dokumendihoidla–repositooriumi. Tarkvara kasutab Javat, SQL-mootoreid. On palju kasutajaid, suuri servereid, arenduste eestvedajad on Massachusettsi Tehnoloogiainstituut (MIT) ja Hewlett Packard.

Fedora, EPrints, DSpace on enamlevinud ja -tunnustatud repositooriumitarkvarad. Pea kõikide OAI repositooriumitarkvarade puuduseks on video vaatamise kehv tugi — video-faili alla laadimise peab üldjuhul ära ootama, alles siis hakkab pleier seda näitama. Video-striimimist tundus mõnede kirjelduste järgi lubavat **MyCoRe** tarkvara, lähemal proovimisel selgus, et ta kasutab selleks väliseid servereid.

Kõikidel süsteemidel on olemas kasutajaliides, v.a Fedora. Enamik süsteemidest sobib väga hästi teadusinfo avaldamiseks, st eelkõige ühe tekstifailipublitseerimiseks ja kirjeldamiseks, arendusi on mitmesuguseid: õpiobjektide tarkvara, videode jm ladustamiseks, kirjeldamiseks ning kättesaadavaks tegemiseks, kollektsioonide kujundamiseks. EPrints'i ja DSpace'i võrdluse kohta lugedes selgus, et esimene võib suurte arhiivide puhul jääda aeglaseks. Johtuvalt tarkvara erinevatest võimalustest, kasutavad mõned organisatsioonid mitut repositooriumi, näit. EPrints ja DSpace.

Juurutamine

2006. aasta lõpul otsustati, et kasutusele võetakse DSpace'i tarkvara. Nagu eelnevalt juttu oli, on DSpace'i tarkvara tasuta. Tarkvara paigaldasime ise, operatsioonisüsteemiks valisime tasuta CentOS Linuxi. Varunduseks kasutame TÜ IT lindiroboteid. Personali vähesuse ja lahenduse võimalikult kiire juurutamise vajaduse tõttu oli tööga seotud peaaegu kogu raamatukogu IT talitus. Kaasatud polnud kataloogijaid, erialareferente repositooriumi tutvustamiseks teaduskondades jt, keda kirjanduses on soovitatud.

Süsteemi nimeks sai dsapce.utlib.ee, et oleks viide DSpace'i tarkvarale. Tarkvara järgi nimetatakse repositooriume tihti (näit. <http://eprints.reclis.org>, <http://dspace.mit.edu>), selle järgi tuntakse ära tarkvara liidestus ja süsteemi tunnustatakse kui standarditele vastavat.

Esmalt anti käiku uute väitekirjade publitseerimise keskkond, millega vahetati välja probleeme tekitanud senine väitekirjade elektroonilise avaldamise süsteem (väitekirjad säilitati ka endises serveris). Vastavalt skaneerimistöde edenemisele lisandusid repositooriumi uued teavikud. Topeltkirjeldamise vältimiseks lepiti kokku, et meta-info on ESTERis, DSpace on andmeladu. dspace.utlib.ee URLide viited registreerisime süsteemis handle.net. Vastav moodul on DSpace'i tarkvarasse sisse ehitatud ja tegelikult olekski raske olnud teda välja võtta. Handle-süsteem garanteerib URLide säilimise Internetis, näit. sellisel kujul: <http://dsapce.utlib.ee/dspace/handle/10062/3> või <http://hdl.handle.net/10062/5676>.

Juurutamise käigus said selgemaks Dspace'i jt repositooriumite olemus, funktsionaalsus ja ka nende kitsaskohad.

Repositooriumina on DSpace mõeldud info ladustamiseks, säilitamiseks, teaviku muutmine pole ette nähtud, avatud faili asendamise võimalus puudub — konkreetne fail (e-raamat, e-väitekiri vm) arhiveeritakse repositooriumis alles siis, kui ta on lõplikult valmis. See tähendab, et kui mingit teksti soovitakse muuta, tuleb olemas olev fail kustutada ning asendada uuega. Selline lähenemisviis teeb failide muutmise ja redigeerimise keeruliseks nii autorile kui raamatukogutöötajale. TÜ Raamatukogu eeskirjad näevad ette, et pdf-faili õigusi ja metainfot pdf-faili sees saab muuta ainult raamatukogutöötaja. Kui vaadata ka muid sisuhalduse lahendusi, siis ilmneb, et veebiajakirjade publitseerimiseks sobib kõige paremini OJS; konverentside veebihalduseks näit. OCS.

DSpace, nagu ka teised repositooriumid, on eriti sobilikud teadustöö tulemuste avaldamiseks, st eelkõige ühe faili, teadustöö teksti, üles laadimist ja kirjeldamist. Kui objekti juurde kuulub mitu faili (näit. 200), siis on see repositooriumis keeruline või lausa võimatu (selleks sobiks ehk Fedora). TÜ Raamatukogus tehakse raamatute skaneerimisel üksikutest lehekülje-failidest üks pdf-fail ja see säilitatakse DSpace'is, vajadusel saab üksikuid pildikaadreid pdf-failist eraldi välja võtta. Piiranguna

selgus, et üle veebi ei saa üles laadida faili, mille maht ületab 500 MB, sellise mahuga on video-failid. Tõsi, paljude failide ja mahtude piiranguid saab vältida toimetades veebiserveris ilma veebiliideseta.

Tuli õppida tundma ja arusaama autoriseerimisloogikast (kollektsoonide, kommuunide, teiste objektide ligipääsu, muutmise, uuendamise jm õigused).

DSpace'i kasutatakse ka raamatukogu dokumentatsiooni arhiveerimiseks ja haldamiseks.

Selgus, et TÜ Raamatukogu polnud sugugi esimene ja ainuke DSpace'i kasutaja Eestis, arvatavasti oli vastav server töös Arhiivikeskusel ja TÜ Viljandi Kultuuriakadeemias.

Arendused ja hetkeseis

ESTERis kirjeldatud väitekirjade e-versioonid koos meta-andmetega importisime veebiserverist DSpace'i, sellega muutusid nähtavamaks nii väitekirjade autorid, tekstid kui ka dspace.utlib.ee server. Oleme jõudnud arusaamisele, et meta-andmed peaksid olema nähtavad mitte ainult ESTERis vaid ka DSpace'is OAI päringute kaudu. TÜ teaduskonnad jt struktuuriüksused on hakanud kasutama DSpace'i ülikooli repositooriumina. Üha enam kostub positiivseid hinnanguid, et DSpace'iga, on raamatukogu teinud hea valiku: see on kooskõlas mitme standardiga, ta on indekseeritud Google Scholaris, TÜ ja raamatukogu reitingud on www.webometrics.info andmetel päris head. Paljud naabridki on otsustanud DSpace'i platvormi kasuks.

DSpace'i uue versiooniga 1.5 peaksid paranema kollektsoonide ja kommuunide veebinäo kujundamise ja üles laadimise võimalused, millega kaoks vajadus mõne kollektsooni veebinäo kujundamiseks serveris www.utlib.ee.

Internetiviited

TÜ Raamatukogu digitaalne repositoorium: <http://dspace.utlib.ee>

Arhiivikeskuse digiarhiiv: <https://digiarhiiv.ee/dspace/>

DSpace, an Open-Source Source for Accessing, Managing and Preserving Scholarly Works:

<http://www.dspace.org>

Open Archive Initiative: <http://www.openarchives.org>

Fedora Commons: <http://www.fedora-commons.org>

Eprints for Digital Repositories: <http://www.eprints.org>

Webometrics Ranking of World Universities: <http://www.webometrics.info>

Nixon, W. Daedalus: Initial experinces with EPrints and DSapce at the Univerity of Glasgow. Online // URL: [http:// www.ariadne.ac.uk/issue37/nixon/](http://www.ariadne.ac.uk/issue37/nixon/) 21. mai 2009

HandleSystem : Unique Presistent Identifiers for Internet Resources: <http://handle.net>