

RaRa virtuaalne labor

Virtuaalse labori
teenuse arendamine



Sisukord

Sissejuhatus

1. Mis on või võiks olla GLAM Lab ehk virtuaalne labor?

2. Disainiprotsess

2.1 Rahvusvaheliste analoogide ja eeskujude kaardistus

2.2 Kasutajagruppide kaardistus

2.3 Kasutajauuring

3. Nägemus RaRa virtuaalsest laborist

Käesolev kokkuvõte ja selle aluseks oleva aruanne on koostatud rahvusvahelise projekti Open Digital Libraries for creative users raames.

Projekti eesmärk on luua teadlastele, uurijatele, tudengitele, õppejõududele ja teistele huvilistele lihtsad, kaasaegsed ja mitmekülgsed võimalused Rahvusraamatukogu (RaRa) digitaalsete kasutamiseks ja sellega töötamiseks, eelkõige rakendamaks sellele teksti- ja andmekaeve metoodikaid. Nende võimaluste loomisele läbi virtuaalse labori teenuse arenduse aitasid kaasa projektipartnerid Hollandi Kuninglik Raamatukogu (Koninklijke Bibliotheek) ja Austria Rahvusraamatukogu (Österreichischen Nationalbibliothek).

Projekti on rahastanud Euroopa Liidu programm "Loov Euroopa". Käesolevas raportis esitatud sisu peegeldab vaid autorite seisukohti. Euroopa Hariduse ja Kultuuri Rakendusamet ning Euroopa Komisjon ei vastuta selles sisalduva teabe kasutamise eest.

Sissejuhatus

GLAM Lab või Virtual Lab (ee virtuaalne labor) on viimasel ajal kultuuripärandiasutustes populaarsust koguv viis oma andmetele rohkem kasutust leida. Enamasti tähendab see nende asutuste kogude põhjal koostatud andmestikke, aga ka tööriistu, mis abistavad ja lihtsustavad andmestikele ligipääsu. Lisaks korraldatakse valdkondlikke üritusi ja/või arendatakse koostöövõrgustikke, mille kaudu kasutajatega kontakti leida ning parimaid praktilisi kogemusi jagada ja meetodeid katsetada.

RaRa-s on mõned aastad toiminud avaandmete portaal (data.digar.ee), mille kaudu on võimalik piiratud vormingutes alla laadida nii Eesti rahvusbibliograafia (ERB) kui RaRa digitaalarhiivi DIGAR metaandmeid ning DIGARis olevate vabakasutuses väljaannete täistekste.

RaRa virtuaalse labori ülesanne on aga teha RaRa andmestikud kasutajatele võimalikult lihtsalt kättesaadavaks, koondada tööriistu ja juhendeid ning abistada andmete kasutamisel, samuti toetada kogukonda üleskutsete, ressursside ja võrgustikuga. Selleks on vaja mitte üksnes veebikeskkonda, vaid ka toetavat terviklikku teenust koos võimeka meeskonnaga.

Mäluasutuste virtuaalsete laborite loomine on algusjärgus. Seetõttu läheneme RaRa labori loomisele ja selle hilisemale ülalpidamisele kui eksperimentaalsele projektile ning arvestame, et vigadest saab õppida.

1

Mis on või võiks olla GLAM Lab ehk virtuaalne labor?

Lühend GLAM tähendab galeriisid, raamatukogusid, arhiive ja muuseume.

Enamasti tähendab GLAM Lab keskkonda, kus asuvad kogude põhjal koostatud andmestikud, aga ka tööriistad, mis abistavad ja lihtsustavad andmetele ligipääsu ja nende abil uute väärtuste loomist. Lisaks korraldatakse valdkondlikke üritusi (konverentsid, seminarid, häkatonid) ja/või arendatakse koostöövõrgustikke (residentuurid, koostöö nt ülikoolidega), mille kaudu kasutajatega kontakti leida, parimaid praktilisi kogemusi jagada ja meetodeid katsetada.

RaRa on otsustanud GLAM Lab-liikumisega kaasa minna ja RaRa avaandmete portaali edasi arendada. Portaali (data.digar.ee) kaudu on võimalik vabalt, kuigi piiratud vormingus alla laadida nii Eesti rahvusbibliograafia (ERB) kui ka RaRa digitaalarhiivi DIGAR metaandmeid ning DIGARis olevate vabakasutuses väljaannete täistekste.

RaRa-s arendamisel olev virtuaalne labor võimaldab pakkuda juurdepääsu rohkematele andmetele, vahendeid nende töötlemiseks ning enamaid ja mitmekesisemaid andmete kasutusvõimalusijuhte.

GLAM Lab ehk virtuaalne labor – ruum, mis võimaldab kas kohapeal või veebi kaudu juurdepääsu digiteeritud ja/või digitekkelistele kollektsioonidele ja andmetele, et nendega eksperimenteerida ja midagi uut luua. (M. Mahey, 2020)

2 Disainiprotsess

Teenusedisain rakendab disainmõtlemise protsesse uute teenuste loomisel või olemasolevate parendamisel. Teenusedisain tähendab teenuse innovaatilist arendamist ja planeerimist, mille keskmes on lähtumine teenuse kasutaja vajadustest ja teenuse pakkuja eesmärkidest. Seega on teenusedisain oma sisult tööriistade/meetodite kogum, mis võimaldab luua teenuseid kasutajakeskselt. Kogu protsess on orienteeritud erinevate lahenduste leidmisele ja nende töökindluse kontrollimisele praktiliselt katsetades ja koosloomes kasutajatega.

Iga arenduse algfaasis on oluline välja selgitada kasutajagrupid ja olulisemad osapooled ning läbi mõelda nende kaasamine arendusprotsessi, nagu ka hinnata, kui mahukat disainiprotsessi iga teenusearenduse puhul läbi viia ja milliseid meetodeid selleks kasutada.

RaRa virtuaalse labori kui teenuse kujundamiseks ja analüüsimiseks kaardistati:

- 1) analoogid ja eeskujud rahvusvahelise kogemuse põhjal;
- 2) intervjuude ja nende analüüsi põhjal kasutajate huvi labori ja selle võimaliku kuju ja sisu vastu.

Tuginedes varasemale kogemusele ja teenusedisaini raames läbi viidud analüüsile koostati visioon, millest peaks labori teenus koosnema (veebileht, andmestikud jne) ning milliseid tegevusi labori käigus hoidmine püsivalt nõuab. Täpsemalt on kõik olulisemad kategooriad kirjeldatud järgnevates peatükkides.

Disainiprotsess



2.1.

Rahvusvaheliste analoogide ja eeskujude kaardistus

Rahvusvaheliselt on virtuaalsed laborid ehk GLAM Lab'id koondunud võrgustikuks International GLAM Labs Community (glamlabs.io), kus on koostatud ka juhend virtuaalse labori loomiseks.

Peamised eeskujud RaRa-le virtuaalse labori teenuse arendamisel on:

- Hollandi Kuninglik Raamatukogu (KB);
- Austria Rahvusraamatukogu (ÖNB);
- Briti Raamatukogu (BL);
- Taani Kuningliku Raamatukogu digihumanitaaria koostööplatvorm;
- Austraalia Rahvusraamatukogu;
- USA Kongressi Raamatukogu;
- Soome Rahvusraamatukogu projekt „Digital Open Memory“;
- Atlas of Digitised Newspapers;
- NewsEye;
- Impresso;
- Europeana;
- ProQuest Text and Data Mining Studio;
- Gale Digital Scholar Lab.

2.2. Kasutajagruppide kaardistus

Kasutajagruppide kaardistus põhines RaRa varasemal ülevaatel kasutajate ringist, kes labori vastu võiksid huvi tunda, ning rahvusvahelisel kogemusel. Teenusedisaini põhimõtetest lähtuvalt jagati need kasutajad eristuvateks rühmadeks. Esialgu püüti leida võimalikult suur hulk potentsiaalseid kasutajaid, kes koondati peamiseks kasutajatüüpideks. Kõikide rühmade juures püüti täpsustada, mille vastu täpselt neil võiks huvi olla ning millised võiksid olla nende oskused. Nii valmis RaRa virtuaalse labori võimalike kasutajate ehk huvigruppide nimekiri.

- Digihumanitaaria ja digisotsiaalia uurijad

Neid huvitavad digimeetodid ja -andmed ning nad tahavad neid uurimistöös kasutada; oskuste tase varieerub, aga vähemalt mõnedel neist on oskus iseseisvalt andmetega ringi käia.

- Üliõpilased ja õppivad andmeteadlased

Võivad kasutada RaRa materjale näidisandmestikeks ja katsetusteks kui midagi põnevat ja ebatavalist. Potentsiaalselt võib nende töö tulemusena valmida kasulikke töövahendeid ka RaRa-le (nt mõne krati näol).

- Õppejõud

Neil võib olla huvi kasutada andmeid digimeetodite õpetamisel või õppetöö andmetega illustreerimiseks: nt avatakse eesti raamatuajalugu uuest küljest, kasutades selleks RaRa andmeid. Kontakt üliõpilastega võib toimuda ka õppejõudude vahendusel.

- Õpetajad

Neil võib olla huvi kasutada andmeid, et õppetööd illustreerida või tutvustada digitaalset kultuuripärandit ja lõimida see aineprogrammidesse eesmärgiga arendada õpilaste info- ja andmekirjaoskust. Seda saaks saavutada koostöös RaRa haridusvõrgustikuga. Hea näide selles vallas on 2021. aasta

mäluasutuste suveseminaril tutvustatud kooliõpilastele suunatud haridusprogramm „Cultural Remix“ (Mahendra Mahey, Aija Sakova posteresitlus).

- Õpilased

Võivad kasutada RaRa andmeid koolitöös, näiteks uurimistöodes. Oluline on siin koostöö õpetajatega, kes saavad aineprogrammides materjale tutvustada: anda ülesandeks otsida temaatilisi artikleid jne.

- Ajakirjanikud ja muud huvilised

Inimesed, kellel on huvi andmete vastu ja kes soovivad nende põhjal kirjutada tekste, artikleid või mõne lühema ülevaate, nt kui palju ja kuidas kirjutati tennisemängust Eesti Vabariigis 1920–1940.

- Huviseltsid, koduloolased, hobiteadlased

Neil võib olla huvi otsida RaRa kollektsioonist näiteks mõnd konkreetset isiku- või perekonnanime või uurida RaRa postkaardi- või maakaardikollektsioone, et näha, mis on pildil, kust kirjad on saadetud vm.

- Huviseltsid, koduloolased, hobiteadlased

Neil võib olla huvi otsida RaRa kollektsioonist näiteks mõnd konkreetset isiku- või perekonnanime või uurida RaRa postkaardi- või maakaardikollektsioone, et näha, mis on pildil, kust kirjad on saadetud vm.

- Kunstnikud, kuraatorid

Huvi olemasolu näitavad nii KB kui ÖNB (käesoleva projekti partnerite) tegevused. Aga on ka näiteid mujalt: Timo Tootsi loodud säutsumasin RaRa-s 2020. aastal või Briti Raamatukogu kunstiprojekt „Crossroads of Curiosity“.

- Rahvusvahelised teadlased

RaRa kui Eesti ühe andmehaldaja poole on varem pöördunud uurimaks võimalusi, kuidas olemasolevate andmete abil analüüsida Eesti ajaleheartikleid.

- Andmeteadlased/erasektor

Andmed on olulised näiteks eesti keele mudelite arendamisel, mõnele kirjastusele võib olla oluline pidev ülevaade avaldatavast. Andmete põhjal on võimalik luua ka haridusrakendusi, näiteks TLÜ haridusinnovatsiooni keskuse EDUSPACE ja ettevõtte MobiLab koostöös loodud liitreaalsusmäng, mille eesmärk on tutvustada ning suunata lapsi uurima Eestiga seotud sümboleid ja loodust (vt www.tlu.ee/armang).

- Avaandmete entusiastid

Näiteks on Facebooki gruppi OpenESTdata koondunud huvilised, kes võivad tahta andmetega töötada n-ö näpuharjutuseks.

- Statistikaamet jm riigiametid

Kui Eesti rahvusbibliograafia uueneks võimalikult väikese hilinemisega, siis saaksid Statistikaamet ja muud riigiametid sellest otse neile vajalikke andmeid, näiteks andmeid kirjastuste, väljaannete jms kohta. RaRa saaks potentsiaalselt pakkuda selleks kasutajasõbralikke töölaudu (dashboard), vt nt juhtimislauad.stat.ee/.

Pidades silmas ajalisi ja muid ressursipiiranguid, määras töörühm võimalikele huvigruppidele prioriteetsusastme. Sellest sõltus, kas virtuaalse labori praeguses arendusfaasis intervjueeritakse konkreetse huvigrupi esindajat.

2.3. Kasutajauuring

Pidades silmas ajalisi ja muid ressursipiiranguid, määras töörühm võimalikele huvigruppidele prioriteetsusastme. Sellest sõltus, kas virtuaalse labori praeguses arendusfaasis intervjueeritakse konkreetse huvigrupi esindajat.

- Esiteks: virtuaalse labori veebikeskkonna arenduseks ja toimimiseks vajalik funktsionaalsus ja kasutajate ootused. Siia kuuluvad sellised teemad nagu suurt huvi pälvinud andmestikud ja juurdepääsuküsimused, andme- ja failivormingud ning tööriistad andmete töötlemiseks.
- Teiseks: laiem valdkondlik kontekst ehk virtuaalne labor kui teenus. Selle arendamiseks on vajalikud eri koostöövormid, mille eesmärk on suurendada teadlikkust virtuaalse labori võimalustest ja julgustada eri kasutajagruppe eksperimenteerima RaRa andmestikega.

Huvipakkuvad andmestikud

Kõige üldisemalt öeldes soovitakse terviktekste, metaandmeid ja/või terviktekste koos metaandmetega.

Terviktekstid, pildid ja muu andmestik

- meediaväljaannete tekstid;
- raamatute tekstid – suured eestikeelsed tekstikogud on väärtuslikud;
- digitaalsed väljaanded üldiselt;
- veebiarhiiv;
- nädalalõppude poliitika- ja uudistesaaated – nende salvestised ja transkriptsioonid, isegi kui meediamonitooringuettevõtted neid juba talletavad;

- pildiline aines koos asukohainfo ja pildiallkirjadega;
- andmestikud, mis on seotud RaRa hoone rekonstrueerimisega;

Intervjuudest tuli välja, et terviktekstidele tahetakse ligi pääseda põhiliselt kahel põhjusel:

- konkreetse teema uurimiseks – nt poliitikauuringuteks;
- keele- ja sõnakasutuse uurimiseks, sel juhul ei olegi oluline tekstide sisu. Küll aga võib mõnele uurijale huvi pakkuda keelekasutus erinevatel ajaperioodidel konkreetsetes valdkonnas, nagu näiteks sõjandus, meditsiini kriisid ja juura.

Terviktekstid koos metaandmetega

Intervjueeritavad rõhutasid korduvalt, et kui uuritakse konkreetset teemat, siis peavad tekstidega kindlasti kaasnema ka metaandmed.

Metaandmed: autor, artiklite pealkirjad, perioodiliste väljaannete nimed, ilmumisajad, märksõnad, teemavaldkonnad, piltide puhul ka asukohainfo.

Tuletatud andmed, s.t andmed, mis on tekkinud analüüsi käigus (nt raamatute ilmumiskohtadega seotud geoinfo, tekstides automaatselt tuvastatud või käsitsi märgendatud isikunimed).

RaRa võiks muuhulgas pakkuda analüüsiks valmis olevaid eeltöödeldud korpuseid, mis oleks temaatiliselt piiritletud, digitaalsel kujul, masinloetavad ja indekseeritud.

Intervjueeritavad tunnevad eriti puudust piisavalt märgendatud materjalidest. See tähendab, et metainfo on kohati puudulik või pole lisatud nt valmiskihte tekstides mainitud organisatsioonide nimedega.

Andmestikega peaks kindlasti olema kaasas ka info andmete kvaliteedi kohta, et uurijad teaksid ette, missugused andmed ja tööd neid ees ootavad.

Kui uurijad juba parandavad andmeid, siis võiks need parandused ka tagasi andmeloojani jõuda.

“Samas muidugi, kui metaandmed ei ole korras, siis [tasuks] need ikkagi üles panna, et on vähemalt midagigi seal.”

“Nimede erikujusid on päris raske puhastada, aga kui on mitu nime ühel väljal, siis seda võiks saada vormindada küll ilusamaks metaandmestikuks.”

Õiguslik raamistik

Kõik intervjuueeritavad, kes teemat puudutasid, pidasid suureks probleemiks just täistekstidele juurdepääsu piiratust.

- Suured meediaväljaanded piiravad tihti juurdepääsu.
- Vabalt kättesaadavad terviktekstid ei moodusta tervikut – need on kättesaadavad katkendlikult või ei ole väljaande kõik aastakäigud tervikuna digiteeritudki. See on eriti suur probleem politoloogidele, kelle jaoks on oluline just ajaline järjepidevus ja tekstikorpuse terviklikkus.
- Terviktekstidele saab ligi ainult üksikutest arvutitest kindlates raamatukogudes ja neid ei saa alla laadida. See ei taga teadlastele vajalikku juurdepääsu.

Kolm probleemkohta, mille osas ei ole kohtutel olnud veel võimalust oma tõlgendust anda:

- Kust jookseb kaitse piir tekstikorpuse ja mudeli (nt sagedusloendi) vahel? Kaitse võiks kehtida materjalile, mida kasutatakse mudeli ehitamiseks (tekstikorpus), aga mudel ise ei oleks enam koormatud algsete piirangutega.
- Millises ulatuses on tekst õigustega kaitstud? Kas laused on ka siis autoriõigusega kaitstud, kui materjal on lausete kaupa korpuses segi paisatud? Aga kui lausetes on isikuandmed, nagu näiteks enamikes riigidokumentides?
- Mis on teadus- ja arendustegevus, mis annab õiguse teksti- ja andmekaeveks? Tehti ettepanek, et kui teksti- ja andmekaeve tulemus on avalik ja kõigile kättesaadav, siis peaks seda pidama teadus- ja arendustegevuseks, isegi kui selle tegija on eraettevõtte.

Tuleb leida lahendus, mis annab teadlasele juurdepääsu vajaduseta RaRa-sse kohale tulla, kuid samal ajal ei võimalda kaitse all olevaid tekste RaRa-st välja viia.

Välja pakutud võimalikud lahendused:

- pilvepõhine juurdepääs;

- tekstide „puhastamine“ – kõrvaldatakse isikuandmed;
- keeleandmestiku puhul minimaalseima ühikuna ainult (segamini aetud) lausete pakkumine;
- leping kasutajaga;

Tehniline juurdepääs

Juurdepääsu andmetele võib anda kas tervikandmebaasi koopiana (nt allalaetavad zip-failid), APIde kaudu (nt praegu on RaRa-s suuremate andmehulkade jaoks kasutusel OAI-PMH) või näiteks SPARQL Endpointiga (nt WikiData päringulehe jaoks loodud lahendus, vt query.wikidata.org/).

Juurdepääs piltidele võiks võimaldada ka piltide ristkasutust.

Andmefailide vormingud

Virtuaalse labori platvormil peaks olema võimalik laadida andmestikke alla eri vormingutes. See võimaldab huvilisel endal valida, missugune vorming tema vajadustele kõige paremini vastab, kas siis tehnilisest või ka temaatilisest vaatenurgast lähtuvalt.

Intervjueeritavate jaoks oli probleem, et RaRa andmed on küll raamatukogudes laialt levinud vormingutes, kuid raamatukoguvälistele osapooltele on need vormingud võõrad.

Sobivate vormingutena nimetasid intervjueeritavad järgmisi:

- .json (Pythoni kasutajatele)
- .tsv
- .csv
- .xml

Tööriistad andmete töötlemiseks

Virtuaalse labori pakutavate tööriistade valik sõltub suuresti nii andmehuviliste erinevatest vajadustest kui ka oskustest.

Enamikust intervjuudest käis läbi vajadus otsingu- ja/või filtreerimis funktsiooni järele – et saaks teha konkreetset päringut, mitte üksnes üksikute aastate kaupa, vaid ka teemade, perioodide, sõnade jne kaupa.

“Kihvt oleks näiteks [leida] kõik, mida Andrus Ansip on öelnud Kärkla kohta. Või kõik [Vene] tsaaririigi tekstid 1890ndatest.”

“Oluline on just, et saaks jupikaupa [---] kirjeid alla laadida, terve andmestik võib minna väga suureks. Nt kuidagi otsinguga ära kirjeldada.”

Algajad kasutajad soovivad peamiselt lugeda. Äärmisel juhul võiks neid huvitada visualiseerimisvahendid andmetega eksperimenteerimiseks ja mängimiseks. Keskmise oskusega kasutajad oskavad teha nii kvantitatiivseid kui ka kvalitatiivseid uuringuid, kuid pigem kasutavad teist meetodit. Nemad oleksidki pakutavate tööriistade peamised kasutajad.

Põhjalike oskustega kasutajad on kolmas huvigrupp, nemad kasutavad pigem kvantitatiivseid uurimismeetodeid ning neil on suurem tehnoloogiliste oskuste pagas. Sellised kasutajad töötlevad andmeid enda kirjutatud koodiga ning ei vaja seetõttu virtuaalse labori pakutavaid tööriistu. Siiski vajavad nad võimalust oma koodiga andmeid töödelda.

Kogenud andmeuurijatele mõeldud andmetöötlusteks kasutatavatest rakendusliidestest mainiti intervjuudes järgmisi:

- Jupyter Notebook – sellega töötamine nõuab eelnevaid teadmisi ja kogemusi, samas saab seda otse veebis kasutada, puudub vajadus paigaldada tarkvara arvutisse. Võimaldab kasutada nii R-i kui Pythoni programmeerimiskeelt.
- RStudio – R-i kasutamiseks loodud andmeanalüüsikeskkond. Saab kasutada ka Pythonit.

Neile lisaks mainiti järgnevaid keeleandmestiku töötlemiseks sobivaid tarkvarasid:

- Sketch Engine – kasutajaarvu põhise litsentsiga.
- NoSketch Engine – tasuta ja väiksemate võimalustega analoog.
- KORP – vabavaraline, Eestis on olemas edasiarendamise kompetents. Leksikograafidele sellest ei piisa.
- Muu omaloodud tarkvara.
- Muu omaloodud tarkvara.

- Mainiti ka töövahendeid, millega saaks teksti parandada, analüüsida morfoloogiat ja süntaksi, lemmatiseerida, tuvastamaks sõnatüve. Selleks võib kasutada näiteks selliseid Pythoni teekide nagu Stanford ja EstNLTK, mis võimaldavad tekstide automaatset märgendamist.

Eelistatuim oleks võimalus andmed alla laadida. Aga see ei ole tingimata vajalik; võimalust andmeid pilves töödelda ja üksnes tulemused alla laadida peeti täiesti vastuvõetavaks.

Koostöö ja kogukond

Intervjueeritavad tõid välja ootuse laiema valdkondliku koostöö suhtes: asutustel, kes kultuuriandmeid koguvad ja kättesaadavaks teevad, peaks olema ühtne andmebaaside register ning andmestikud peaks olema omavahel lingitud. Mitu intervjueeritavat arvas, et eraldiseisvate andmestike koondamine annaks kasutajatele võimaluse neid võrrelda või kokku sobitada ja luua niiviisi täiesti uusi lahendusi. Sellise lahenduse nimel tasub tegutseda.

Kaasamine andmekvaliteedi parandamiseks

Intervjueeritavad nägid ühe koostöövõimalusena andmete kasutajate, tudengite, aga ka laiemalt ühiskonna kaasamist andmekvaliteedi parandamisse. Näiteks on andmete korrastamiseks ja kvaliteedi parandamiseks võimalik kasutada ühisloomet (*crowdsourcing*), nagu seda on korduvalt teinud BL. Andmekvaliteedi kontrollimiseks ja parandamiseks kasutati koole, õpilaste rühmi, haridusprogramme, vabatahtlikke jt.

Teadlikkus ja oskused olid oluline temaatika, mis intervjuudest kõlama jäi. Esmalt on vaja suurendada teadlikkust sellest, et RaRa-s on andmeid, mis võiks teksti- ja andmekaeve objektiks olla. RaRa saaks pakkuda oma avaandmeid ülikoolides asjakohastel erialadel sisendiks õppetöös. Õppejõud saaksid suunata oma tudengeid RaRa andmeid kasutama. Samamoodi saaks RaRa töötajad ise minna ülikoolidesse külalislektorina, korraldama töötubasid või seminare, et tutvustada raamatukogu kaudu juurdepääsetavaid andmeid ja nende töötlemiseks sobivaid tööriistu.

Teiseks tuleb arendada andmete töötlemise oskusi, eelkõige tudengite seas, või vähemalt julgustada inimesi aktiivsemalt kasutama olemasolevaid oskusi. Selles osas oleks palju abi kasutuslugudest, -juhenditest ja -videotest.

Kolmandaks lootsid intervjuueeritavad, et RaRa saaks tudengitele ja teadlastele pakkuda reaalselt praktilist väljundit oma andmete töötlemisel. Virtuaalne labor võiks välja töötada-pakkuda konkreetseid projekte, mille jaoks andmeid vaja läheb. Igal juhul võiks arvestada võimalusega parandatud andmeid tagasi saada ja laboris autorlust tunnustades avalikustada.



Teenuse loomine

Virtuaalsete laborite rahvusvahelises kogemuses võib teenuse arendamise juures välja tuua viis kõige tähtsamat aspekti.

- Oluline on tippjuhtkonna mõistmine ja toetus. On tähtis, et juhtkonna tasemel saadaks aru, mida soovitakse saavutada ja kuidas seda tehakse.
- Detailne ülevaade raamatukogu digikogudest – missugused digitaalsed ressursid ja varud kogude ja andmete näol raamatukogul üldse olemas on. Iga kollektsioon nummerdatakse ja varustatakse infoga, kuidas kogu moodustati, mis vormingus on selles andmed jne.
- Raamatukogu sees tuleb panna paika selged töövood, kuidas määratletakse digiainesele autoriõigusest tulenevad litsentsid ja defineeritakse kasutusvõimalused.
- Mõelda läbi, kuidas pakkuda ligipääsu: APId, zip-failid, andmevormingud vm, kas panna andmestikele juurde DOld jms.
- Ühendada inimesed andmetega – s.t raamatukogu poolt andmete loojad ja haldajad, mujalt kasutajad ja tihti ka täiendajad. Need inimesed tuleb viia üksteisega kokku ja näidata, et nende töö on väärtuslik ja seda väärtustatakse.

Juhendid ja näited

Enamik intervjueeritavatest leidis, et ei piisa, kui andmed tehakse kättesaadavaks koos erinevate töötlemistööriistadega. Virtuaalse labori arendamisel tuleb silmas pidada, et kasutajate oskused on eri tasemel.

Andmestikel ja tööriistadel peavad olema sisukirjeldused: millised andmed on saadaval ning milliseid tööriistu nende töötlemiseks pakutakse. Samuti peaks veebikeskkonnas olema kasutuslood: huvitavad näited, kuidas andmeid kasutati ja kuidas saaks veel kasutada.

RaRa võiks pakkuda uurijale uurimistöö läbiviimisel abi, mida keegi teine ei saaks anda ning ka uurimisteemasid. Suunav tugi üleüldiselt julgustab uurijaid rohkem RaRa kogusid kasutama.

Nõustamine võimaldab juhtida uurijaid suundades, kuhu nad ise ei oskaks minna, viia nad kurioosumite ja põnevate materjalide juurde, mida ainult kogudega väga hästi tuttav inimene leiab.

Meediatöö

Virtuaalse labori teenuse edukus sõltub raamatukogu oskusest seda turundada ning sellele avalikkuse ja meedia tähelepanu tõmmata. Mida suurem on võimalus, et teadlase projekt köidab avalikkuse tähelepanu, seda suurem on ka tema huvi pakutava andmestiku vastu.

3 Nägemus RaRa virtuaalsest laborist

Tuginedes ülalkirjeldatud rahvusvahelisele kogemusele ja intervjuudele potentsiaalsete huvilistega, puudutavad RaRa ülesanded sel alal kõige üldisemalt kolme teemat.

- Mida näeb kasutaja (*front-end*) ehk tegevus, mis on suunatud kõigile kasutajatele ja on labori n-ö avalik külg: veebileht ja selle sisu ning sisu haldamiseks vajalikud tegevused.
- Mida teeb arendaja/haldaja (*back-end*) ehk nn töö tagatoas: et oleks ressursse, mida jagada, et need oleks kvaliteetsed ja kasutatavad.
- Töö kogukonnaga on labori edu alus. Kaasamine vajab nii regulaarseid tegevusi kui varasemaid valmislahendusi.

Labori veebikeskkond

Veebilehel on oluline roll labori töövahendite, andmestike ja tegevuste edastamisel ja tutvustamisel, nii et need peavad olema kergesti leitavad. Labori peamised elemendid on andmestikud ja tööriistad. Oluline on, et veebileht suunab kasutajaid ka ise andmekogude täiendamisse panustama; mõnikord toimuvad kaasamisaktsioonid, milles tuleks tavalugejaid osalema suunata.

Andmestikud

Andmestikele ja ka tööriistadele peaks virtuaalse labori veebikeskkonna kaudu olema kõikidel huvilistel juurdepääs autentimata. See tagab võimalikult laia kasutajaskonna, kes eksperimenteerivad ja kasutavad eri võimalusi ning koguvad informatsiooni.

Samas võib teatud andmestikele juurdepääsu võimaldamisel olla siiski vajalik ka autentimine, et kindlaks määrata materjali kasutaja õigused ja kohustused. Sellisel juhul peaks veebiplatvorm võimaldama ka veebi kaudu lepingu sõlmimist.

Labori peamine sisu on andmestikud, mis vajavad RaRa poolt sobivale kujule viimist, tutvustuste kirjutamist, uuendamist ja haldamist. Seda võib teha RaRa sees või mõnede usinamate kasutajate abiga.

Tööriistad virtuaalses laboris

Virtuaalse labori veebiplatvormi arendamisel võiks arvestada vähemalt kahte tüüpi andmehuvilistega: need, kes oskavad ja soovivad andmetele juurdepääsuks kasutada pigem koodi ning need, kes pigem kasutaksid lihtsamaid tööriistu.

Leida tuleb tasakaal selle vahel, mis võib olla ka uustulnukale õpitav ja kui palju koodikasutust võib osutada juba eemalepeletavalt paljuks.

Tööriistad hõlmavad teiste laborite pakutavat vaadeldes üpris erisuguseid vahendeid – graafikutest, kus ehk saab veidi andmetega mängida kuni andmestiku uurimiseni mõne keerulisema vahendi abil või andmestike jaoks mõeldud parandusalgoritmidega.

Mitmed tööriistad võivad piirduda kirjeldusega ja lingiga väliskeskkonda. Tihti koostatakse andmete uurimiseks keskkondi R-is Shiny lahenduste abil (*shiny.rstudio.com*). Veebileht võiks võimaldada neid paigaldada ja kui vaja, ühendada meie oma serveris hoitud suuremate andmefailidega.

Erilabor tekstitöötluks

Üks eraldi andmetöötlusvahend, mis on RaRa-s juba osaliselt välja arendatud, on virtuaalne keskkond, kus saab jooksutada koodi ja teha lihtsamat andmetöötlust. RaRa on praegu läbi rääkinud ETASiga, kes saab pakkuda Jupyter Notebooki keskkonda, andmete hoidmist ja turvalist sisselogimist.

Praegu on selle lahenduse kaudu ligipääsetavad avatud võrgus kättesaadavad tekstid digiteeritud Eesti artiklite (DEA) kollektsioonist.

Kasutajate loodud andmestikud ja blogipostitused

Labori andmestike ja tööriistade tutvustamisel ja illustreerimisel jäävad olulist rolli kandma kasutusnäited. Need näited on tihti lisanud kasutajad ise piltidega blogipostitustena või Markdown-dokumentidena.

Sama lugu on andmestikega, mis võivad pärineda kasutajatelt. Sellisel juhul peab olema suhteliselt standardne võimalus informatsiooni sisestada ja väga hea oleks, kui see informatsioon oleks säilitatav lihttekst-failina (*plaintext*), mille saab andmestikule lisada (nt viide, sildid (*tag'id*) jm ühes readme.md-failis).

Kui kogu andmestiku metainformatsioon on ühes failis ja standardstruktuuriga (mida veebileht kas oskab lugeda või mis on veebilehele kergesti üle kantav), on seda kerge uuendada ja see ei kao kunagi ära.

Kogukonna loomine ja kaasamine

Virtuaalse labori veebiplatvormiga peab kindlasti kaasnema töö kogukonnaga. See on pidev töö ning RaRa digitaalsete kogude ja andmete kasutamise suurendamisel võib see olla isegi veebiplatvormist olulisem.

Vältima peab olukorda, kus arendatud on detailirohke ja innovaatiline veebiplatvorm, kuid andmehuvilised ei tea seda, ei oska seda kasutada või ei vasta see siiski huviliste vajadustele, kuna nende soove ei ole kas algsel või edasisel arendamisel arvesse võetud.

Kogukonna loomiseks ja sidemete hoidmiseks on oluline vähemalt järgmine:

- Ühenduse hoidmine ülikoolidega, eelkõige õppejõudude ja koolitajatega, et oleks ülevaade, mida parasjagu ülikoolides tehakse.
- *Crowdsourcing* ehk ühisloome või rahvahange ja huviliste kaasamine.
- RaRa enda kollektsioonide loojad ja hoidjad on olulised. Peaks jälgima, et inimesed, kes on RaRa-s vastutanud andmestike loomise eest (nt veebiarhiivi, rahvusbibliograafia jne koostajad), näevad, mis kasu nende kogudest ja andmete avamisest võiks tulla.
- RaRa virtuaalne labor peab pidevalt pildil püsima ja motiveerima huvilisi tegudele.
- Labori meeskond peaks olema valmis huvilisi nõustama ja muul viisil aitama.
- Silmas tasub pidada ka võimalust seostada virtuaalse labori tegemised teiste RaRa ettevõtmistega (nt kratid või ühisloome), nt kas tekstide automaatmärgendamiseks tehtud mooduleid on võimalik ehk mõnes tudengiprojektis rakendada.

Kindlasti jääb RaRa-le vähemalt osaliselt harimise ja propageerimise roll – nii vahendite kasutamine, koodi kirjutamine kui ka üldine teadmine, et andmetega võib palju toredat teha, on olulised teemad.

