

Teaduse avaandmete säilitamine ja kättesaadavaks tegemine

DataCite Eesti Konsortsium

Teaduse avaandmete avalikustamise põhimõtteid hakati maailmas kooskõlastama juba läinud aastatuhande lõpul ja esimesed edusammud on tehtud. Teadusandmete vabamast kättesaamise ja taaskasutamise võimalusest on huvitatud teadust rahastavad avaliku sektori organisatsioonid, kvaliteetsete teadusajakirjade toimetused, oma teadustööd mainekamaks ja nähtavamaks teha soovivad teadlased.

Teaduse avaandmeteks nimetatakse teadustöö käigus kogutud andmed, mis tehakse Internetis avatud juurdepääsuga kättesaadavaks. Identifitseerimistunnustega teadusandmed on kergesti leitavad ja viidatavad. Olles lingitavad teiste andmetega, moodustavad andmed suuri andmevõrgustikke.

Üldine põhimõte käsitleda teadusandmeid iseseisvalt avaldatud digitaalsete objektidena tekkis eelmise aastatuhande lõpus. 2000. a lõpus hakkas Saksa Teadusfond (Deutsche Forschungsgemeinschaft, DFG) rahastama rahvuslikku töögruppi töötamaks välja teadusandmete avaldamise ja viitamise põhimõtteid. Keskselks punktideks oli andmete muutmatul kujul püsimine peale avaldamist, pikaajaline säilitamine, andmete detailsus, kvaliteedi kindlustamine, autori vastutus andmete õigsuse eest, korrektne viitamine autorile, teadusandmetele viitamise integreerimine teadusartiklite viitamisinfo. 2002. a valmis teadlastest ja raamatukogutöötajatest koosnevalt töögrupilt avaandmetele viitamise aruanne „Conception of citing scientific primary data“.

Analoogia teadusandmete ja teadusartiklite avaldamise vahel viis otsuseni valida sobiva identifikaatorina välja DOI (digitaalse objekti identifikaator, ingl k *digital object identifier*), sest DOI-de kasutamine artiklite publitseerimisel oli teadlaskonnale juba varasemalt tuttav. Teise alternatiivina oli kaalumisel URN (Uniform Resource Name).

DOI süsteem on objekt identifitseerimissüsteem digitaalses keskkonnas. DOI võib anda igale veebis kasutatavale üksusele. DOI kaudu saab infot objekti ja selle leidumuse kohta Internetis. Koos DOI-nimega hoitakse objekti kohta käivaid metaandmeid, sealhulgas ka veebiaadressi. DOI koosneb prefiksist ja sufiksist, mida eraldab kaldkriips. Info digitaalobjekti kohta võib aja jooksul muutuda, kuid selle DOI jääb alati samaks.

Oli tarvis määratleda teadusandmetele sobiv DOI metaandmete profiil, mis oleks kohaldatav erinevatel teadusaladel. Alustati geoteaduste valdkonnast, kus projekti olid hõlmatud kolme Saksamaa teaduskeskuse andmearhiivid ning keskne tehnikaraamatukogu (Technische Informationsbibliothek, TIB) kui ametlik DOI registreerimisagentuur ja rahvusvahelise DOI fondi (International DOI Foundation, IDF) täisliige. Projekti käigus välja töötatud STD-DOI (Scientific and Technical Data DOI) metaandmete nõutav kogum baseerus ISO 690-2 and the Dublin Core standarditele.

Kontseptsiooni väljatöötamisele järgnesid esmase rakendamise ja teenuse väljakujundamise faasid. Esimene DOI – 10.1594/WDCC/EH4_OPYC_SRES_A2 omistati TIB-i poolt 2004. a kliimamudeli katsetusega seotud andmekogule. Aasta jooksul jõudis TIB raamatukogu kataloogi 30 DOI-ga varustatud andmekogule.

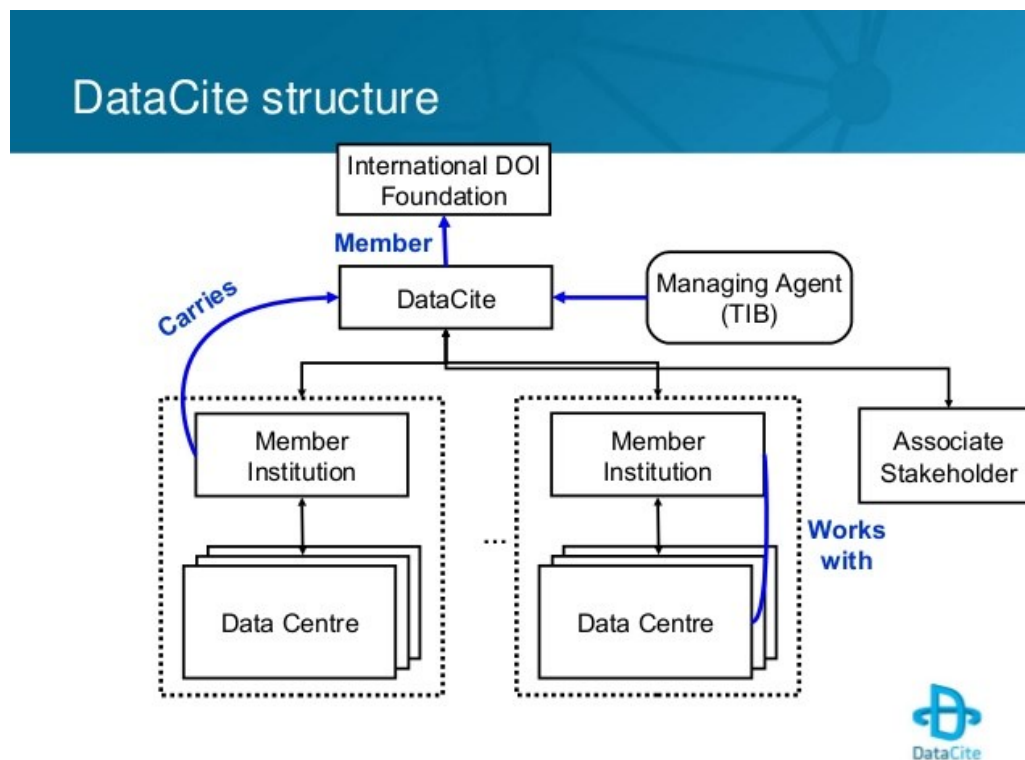
Suurenenud rahvusvaheline huvi kiirendas üldise teadusandmete avaldamise taristu loomist. Saksamaa mudel, kus TIB kui DOI-de registreerimisagentuur millega andmearhiivid olid

lepinguliselt seotud, loodeti üle kanda ka teistesse maadesse. 2007–2008 tutvustati projekti tulemusi nii Saksamaal kui mujal. Mitmed tehnikaraamatukogud ja tulevased andmekeskused väljendasid huvi määrata andmeüksustele DOI-sid ja soovisid pakkuda DOI-de registreerimist ühena oma teenustest.

TIB hakkas TechLib konsortsiumi partneritega arutama võimalust, et tehnikaraamatukogud Euroopas võiksid ühiselt tegutseda teadusandmete DOI-sid registreeriva keskusena. Idee hakkas võtma ilmet 2008. a lõpus, kui ka Briti Raamatukogu ja Prantsuse teadus- ja tehnikainformatsiooni instituut INIST väljendasid huvi ühineda püüdlustega.

2009. a veebruaris toimunud seminaril Pariisis kuus raamatukogu allkirjastasid memorandum (,,Memorandum of Understanding“) vajadusest luua mittetulunduslikel põhimõtetel tegutsev ühendus, et registreerida andmeüksusi ja määrata neile püsidentifikaator. Järgnevatel kuudel loodi koostöösuhted nelja maineka andme- ja teaduskeskusega USAst, Kanadast ja Austraaliast.

1. dets. 2009 asutati Londonis Briti Raamatukogus seitsme asutajaliikme poolt **DataCite - International Data Citation Initiative**. Juriidiliselt on DataCite Saksamaa mittetulundusühing, mille peakontor asub TIB-is Hannoveris.



DataCite on kasvanud ülemaailmseks konsortsiumiks, kuhu kuulub 30 liiget 16 maalt,

<https://www.datacite.org/about-datacite/members>

Nüüdseks on DataCite-i poolt registreeritud DOI-de arv ületanud 4 miljonit. See on toimunud koostöös 350 andmekeskusega üle maailma. Viimastel aastatel välja arendatud tehnilise taristu südamiku moodustab metaandmete varamu DataCite Metadata Store. Statistikaportaali kajastab erinevates andmekogudes registreeritud DOI-de hulka.

DataCite Metadata Search: <http://search.datacite.org/ui>

DataCite Statistics: <http://stats.datacite.org>

Datacite Eesti, <http://datacite.ut.ee>

TÜ raamatukogu ja TÜ loodusmuuseumi algatusel ning Eesti Teadusagentuuri finantseerimisel liitus Tartu Ülikool 2014. aastal DataCite-ga, saades Eestis õiguse omistada nii üksikobjektidele kui andmekogudele unikaalseid DOI-numbreid. Üle aasta kestnud projekti käigus loodi DataCite Eesti veebipõhine platvorm, mis võimaldab DOI-de omistamist teaduslikele andmekogudele üle Eesti.

Suuremad Eesti ülikoolid (TÜ, TTÜ, TLÜ, EMÜ) moodustasid 28. jaan. 2015 **Datacite Eesti konsortsiumi**. Liikmetel on õigus kasutada kõiki konsortsiumi pakutavaid liidestamisega ja andmekeskuste administreerimisega seotuid teenuseid, sh DOI registreerimine, teadusandmete metaandmete säilitamine, metaandmete otsingumootori kasutamine jms. Konsortsiumi aastamaks tagab nende ülikoolide teadlaste teaduse toorandmetele piiramatu koguse DOI-sid. DataCite Eesti platvormi arendamist koordineerib Tartu Ülikooli Raamatukogu.

Datacite Eesti Konsortsiumis esindab **Tallinna Tehnikaülikooli** TTÜ Raamatukogu direktor Jüri Järs. Konsortsiumi liikmesülikoolide raamatukogud on moodustanud DataCite Eesti töögrupi, kuhu TTÜ raamatukogust kuulub Kai Kalvik, raamatukogu infoosakonna ainespetsialist. Kai Kalvik on DataCite Eesti kontaktisik TTÜ-s.

Esimesed 100 unikaalset DOI-nime registreeriti 2014. a TÜ keemia instituudi QSAR mudelite varamu (QsarDB) uurimisandmetele. Projekti käigus arendati ka teadusandmete repositooriumi Datadoi ning PlutoF ja SARV andmebaaside automatiseeritud lahendusi, mis kõik võimaldavad DOI-de taotlemist ja registreerimist.

PlutoF platvorm võimaldab loodusteaduslike andmete veebipõhist sisestamist, vaatamist ja täiendamist. Uurijatel on võimalik tõsta enda andmed mistahes formaadis PlutoF pilve ja küsida DOI identifikaatorit. 2015. aasta alguses tehti ettevalmistustöid ligi 500 000 UNITE liigihüpoteesi avaldamiseks, mille tarbeks loodi automatiseeritud lahendused andmehulkade koostamiseks, metaandmete lisamiseks ning DataCite süsteemis registreerimiseks. Need andmed moodustavadki hetkel valdava enamuse ESTDOI tunnusega varustatud DOI-dest. ESTDOI-ga on varustatud Tartu Ülikooli kui DataCite-i liikme vahendusel registreeritud andmekeskused (allocator ESTDOI - Tartu University).

DataCite platvormi kaudu saab leida ka geokogude infosüsteemi *SARV: Eesti geokogude infosüsteem ja andmerekpositoorium* kuuluvaid teadusandmeid (kirjeldatud kui - datacentre ESTDOI.GEO - Geoloogilised andmekogud).

TTÜ Geoloogia Instituudil on andmete talletamisel ja infosüsteemide loomisel juba pikaajalised kogemused. Kogude elektroonilise andmebaasi juurutamist alustati 1996. a, rahvusvahelise GeoCAsE andmevõrgustikuga liituti 2010. a.

Eesti geokogude infosüsteemi DOI identifikaatorite otsing: <http://geokogud.info/doi>

Tartu Ülikooli raamatukogu teaduse toorandmete repositoorium <http://datadoi.ut.ee> on loodud toetamaks TÜ ja teiste teadusasutuste üksikteadlaste ning väiksemate töögruppide andmete säilitamist.

Et varustada teadusandmed DOI-ga, tuleb andmekeskustel tagada andmete säilimine, nende varustamine vähemalt kohustusliku osaga DataCite-i metaandmetest, piiranguteta juurdepääs metaandmetele, ressursi veebiaadressi õigsus jm.

DataCite Eesti eraldab andmekeskustele DOI prefiks. Andmekeskused saavad ise kujundada DOI sufiks, pidades silmas, et sufiksi unikaalsus oleks tagatud. Enne otsustamist on võimalik taotleda ka proovikontot.

Kuidas uurimisandmeid DOI abil registreerida: <http://datacite.ut.ee/doi.php>

DataCite Eesti konsortsiumi kaudu liitunud andmehoidlate ja teadustaristute kohta lähemalt: <http://datacite.ut.ee/services.php>

Eestis registreeritud DOI-de arv andmekeskuste kaupa: http://stats.datacite.org/?fq=allocator_facet%3A%22ESTDOI+-+Tartu+University%22&#tab-datacentres

Kuidas leida sobivat repositooriumit?

Teadusandmete repositoorium on veebipõhine teadusandmete arhiiv. Repositooriumid võivad olla nii valdkondlikud kui ka institutsionaalsed.

Teadusandmete repositooriumide register **re3data.org** (Registry of Research Data Repositories, avatud 2012, DFG finantseeritud) hõlmab üle 1300 andmehoidla maailma eri riikidest, sisaldades nii avatud kui piiratud juurdepääsuga repositooriume, püsidentifikaatoritega (ARK, DOI, HDL, PURL, URN) ja litsentsidega varustatud või mittevarustatud, valdkonnapõhiseid ja multidistsiplinaarseid repositooriume. 2015. a lõpuks on re3data.org liidetud DataCite-i teenustega ning lõpeb ka teises suure repositooriumide kataloogi Databib liitmine re3data.org-ga.

Registry of Research Data Repositories: <http://www.re3data.org/>

re3data.org kasutamiseks antakse suuniseid Euroopa Komisjoni juhendis „Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020”. Horisont 2020-s on teatud valdkonnad valitud pilootprojekti kuuluvateks, kus nõutakse andmehaldusplaani koostamist (Open Research Data Pilot - Data management and data management plans (DMP)).

Kirjastuselt **Nature** hakkas 2014. a ilmuma veebipõhine, *open access* ajakiri „Scientific Data”, mis avaldab artikleid teaduslikult väärtuslike teadusandmete kohta (Data Descriptors), et andmeid paremini leida, tõlgendada ja taaskasutada. „Scientific Data” autorid on avaldamiseks kohustatud maksma artikli töötlemise tasu (article-processing charge, APC). Andmepoliitika ja Nature poolt soovitatud repositooriumide kohta saab lähemalt lugeda: <http://www.nature.com/sdata/data-policies/repositories>, <http://www.nature.com/sdata/data-policies#repo-suggest>


2012. a alustas **Thomson Reuters** teadusandmete viitamisinfor andmebaasi Data Citation Index loomist. Augustis 2014 teatasid Thomson Reuters ja DataCite omavahelisest koostööst tagamaks, et kõik kõrgetasemelised teadusandmed, mis pärinevad DataCite-ga koostöös tegevatest repositooriumidest, oleksid Data Citation Index-is kajastatud.

Data Citation Index-i, repositooriumide hindamise ja valiku põhimõtetest lähemalt: <http://thomsonreuters.com/en/products-services/scholarly-scientific-research/scholarly-search-and-discovery/data-citation-index.html>

DataCite Metadata Search: <http://search.datacite.org/ui> , – valida Advanced Search

search.datacite.org/ui 🔍 ☆

[Options](#) | [Advanced Search](#) | [About Us](#) | [Contact](#) | [Help](#)

 **Metadata Search beta**

DataCite Metadata Advanced Search
✕

Search

Search in all fields

Field Search

DOI

Title

Creator

Publisher

Contributor

Format

Subject

Date

Description

Resource Type

Alternate Identifier

Related Identifier

Publication Year (to)

Filter

Prefix ▼

Allocator ▼

Datacentre ▼

General Resource Type ▼

Language ▼

Has Metadata ▼

Koostanud: Kai Kalvik, oktoober 2015