



Eiríkur Rögnvaldsson

Opin málföng — allra hagur

Hugvísindaping

16. mars 2013



Hvað eru *opin og frjáls málföng*?

- Málföng - e. *language resources*
 - sbr. *aðföng, tilföng*
- Hvers kyns „auðlindir“ í máltækni og málvinnslu
 - gögn, gagnasöfn og hugbúnaður
- Opin gögn - e. *open data*
 - *Open data is data that can be freely used, reused and redistributed by anyone - subject only, at most, to the requirement to attribute and share alike.*
 - (<http://opendatahandbook.org>)



Af hverju opin gögn?

Aukinn sýnileiki og áhrif – Vísindaframlög í opnum aðgangi eru aðgengileg breiðari lesendahópi, eru lesin af fleirum og það er oftast vitnað í þau. Þetta leiðir af sér aukin áhrif eða “slagkraft” framlagsins (e. *impact*) fyrir höfunda og rannsóknarstofnanir sem þeir vinna fyrir, og sjóðir sem fjárfesta í rannsóknunum fá meira fyrir peningana.

Frjálst aðgengi að upplýsingum og þekkingu – Óhindrað aðgengi eflir og hraðar á rannsóknar- og uppgötvunarferlinu fyrir *alla þátttakendur í vísinda- og fræðastarfsemi*, ekki bara fyrir þá sem hafa efni á að kaupa áskrift að lokuðum fræðiritum. Til dæmis hafa vísindamenn í þróunarlöndunum aðeins aðgang að litlum hluta alls útgefins fræðiefnis, samanborið við kollega sína í vestrænum ríkjum.

Aðgengi almennings – Meirihluti rannsóknastarfsemi er fjármagnaður gegnum rannsóknarsjóði á vegum hins opinbera (t.d. **Rannís** á Íslandi). Almennir skattborgarar hafa skýlausan rétt á að nálgast niðurstöður rannsókna sem þeir hafa sjálfir borgað fyrir.

Samfélagslegur ávinningur – Opinn aðgangur leiðir einnig til víðtækari ávinnings fyrir samfélagið í heild, óháð því hvort einstaklingar nálgast og nýta sér vísindaefni beint. Til dæmis gagnast hraðari framfarir í læknávisindum og aukið aðgengi að nýjustu þekkingu öllum þeim sem þurfa á læknisþjónustu að halda.



Stefna stjórnvalda, 2007

1. Gæta skal þess að gefa frjálsum hugbúnaði sömu tækifæri og séreignarhugbúnaði þegar tekin er ákvörðun um kaup á nýjum búnaði og ávallt skal leitast við að gera sem hagstæðust innkaup.
2. Leitast verði við að velja hugbúnað sem byggist á opnum stöðlum, hvort sem um staðlaðan búnað er að ræða eða sérsmíðaðan. Opnir staðlar eru yfirleitt ráðandi í frjálsum hugbúnaði.
3. Stefnt verði að því að opinberir aðilar verði ekki of háðir einstökum hugbúnaðarframleiðendum og þjónustuaðilum. Notkun frjáls hugbúnaðar er liður í því.
4. Stefnt skal að því að hugbúnaður sem smíðaður er og fjármagnaður af opinberum aðilum, m.a. í rannsóknar- og þróunarverkefnum, verði endurnýtanlegur. Liður í því er að hugbúnaðurinn sé frjáls. Í upphafi slíkra verkefna skal gera áætlun um endurnýtingu hugbúnaðarins.
5. Stuðlað verði að því að nemendur í menntastofnunum landsins fái tækifæri til að kynna og nota frjálsan hugbúnað til jafns við séreignarhugbúnað.



Drög að stefnu HÍ um opinn aðgang

- Háskóli Íslands leggur áherslu á að sem flestir geti notið afurða þess vísindastarfs sem unnið er innan skólans. Skólinn ætlast því til þess af akademískum starfsmönnum sínum að þeir birti fræðigreinar sínar á vettvangi þar sem aðgangur er opinn og ókeypis, svo sem í tímaritum í opnum aðgangi, safnvistun, forprentagrunnum, eða á annan hátt.
- Akademískir starfsmenn skulu veita vísindasviði án endurgjalds rafrænan aðgang að lokaútgáfu vísindagreina sinna ekki seinna en á útgáfudegi. [...] Skólanum er heimilt að vista greinarnar og gera þær aðgengilegar í opnum gagnagrunnum, s.s. Skemmunni.



Reglur RANNÍS um opinn aðgang

- Niðurstöður rannsóknaverkefna, sem styrkt eru, að hluta til eða að öllu leyti úr sjóðum í umsýslu Rannís, skulu birtar í opnum aðgangi. Tilgangurinn er að sem flestir geti notið afurða vísindastarfa sem styrkt eru af opinberu fé á Íslandi. Krafa um opinn aðgang nær til ritrýndra greina en ekki til bóka, bókakafla eða lokaritgerða nemenda.
- Verkefni sem hlotið hafa styrk úr sjóðum Rannís fyrir janúar 2013 falla ekki undir kröfu um birtingu niðurstaðna í opnum aðgangi. Rannís hvetur þó eindregið til þess að sem flestir vísindamenn birti niðurstöður rannsóknaverkefna í opnum aðgangi.



Opin gögn og íslensk málstefna

- Íslensk málstefna
 - samþykkt á Alþingi 12.3. 2009
- Tillögur um málföng og opin gögn:
 - Að stöðugt verði unnið að uppbyggingu og eflingu mállegra gagnasafna sem eru forsenda fyrir þróun og smíði margs kyns máltækni-búnaðar.
 - Að málleg gagnasöfn og hugbúnaður til að vinna með íslenskt mál verði gerð opin og frjáls eftir því sem kostur er.



Stöðluð nýtingarleyfi

- Opnum og frjálsum gögnum og hugbúnaði er yfirleitt dreift með stöðluðum nýtingarleyfum
- [GNU leyfi](#)
 - (Lesser) General Public License
 - aðallega notuð fyrir hugbúnað
- [Creative Commons leyfi](#)
 - CC BY, CC BY-SA, CC BY-ND
 - CC BY-NC, CC BY-NC-SA, CC BY-NC-ND
 - aðallega notuð fyrir texta



CC BY



- Attribution
 - This license lets others distribute, remix, tweak, and build upon your work, even commercially, as long as they credit you for the original creation. This is the most accommodating of licenses offered. Recommended for maximum dissemination and use of licensed materials.
- LGPL (Lesser General Public License)
 - er með u.þ.b. sömu skilmálum



CC-BY-SA

- Attribution-ShareAlike
 - This license lets others remix, tweak, and build upon your work even for commercial purposes, as long as they credit you and license their new creations under the identical terms. [...] All new works based on yours will carry the same license, so any derivatives will also allow commercial use. This is the license used by Wikipedia, and is recommended for materials that would benefit from incorporating content from Wikipedia and similarly licensed projects.



CC BY-NC-ND

- Attribution-NonCommercial-NoDerivs
 - This license is the most restrictive of our six main licenses, only allowing others to download your works and share them with others as long as they credit you, but they can't change them in any way or use them commercially.
- ÍSLEX er dreift með þessu leyfi



META-NORD og META-NET

- META-NORD er verkefni sem Norðurlönd og Eystrasaltlönd stóðu að. Það er eitt þriggja systurverkefna innan META-NET sem hafa að markmiði að efla og staðla málföng sem nýst geti í margvíslegum máltæknaverkefnum
- Megintilgangur verkefnanna var að skapa tæknilegar forsendur fyrir margmála upplýsingasamfélagi í Evrópu þar sem allir geti notað móðurmál sitt við öflun og úrvinnslu upplýsinga



Verkefni META-NORD á Íslandi

- Vefsíða – vefir.hi.is/metanord
 - fésbókarsíða - [META-NORD Ísland](https://www.facebook.com/META-NORD-Island)
- Efling máltækni á Íslandi
 - samning málskýrslu á íslensku og ensku
 - vitundarvakning, kynning, útbreiðsla, áframhald
- Söfnun og uppbygging íslenskra málfanga
 - athugun og frágangur á réttindamálum og leyfum
 - val og söfnun málfanga og viðræður við rétthafa
 - lýsing, uppfærsla, stöðlun og frágangur málfanga



Stöðluð skráning málfanga

- Málföng eru skráð í META-SHARE
 - dreifða gagnahirslu með mörgum útstöðvum
- Staðlaðar upplýsingar eru skráðar
 - tungumál, tegund, stærð, snið, nýting, höfundur, umráðamaður, leyfi o.fl.
- Einnig hvernig eigi að nálgast málföngin
 - sumum er hægt að hlaða niður af META-SHARE
 - öðrum af eigin vefsíðum
 - stundum þarf að hafa samband við rétthafa



META-SHARE



META-SHARE

Register

Login

Browse Resources

Community

Documentation

Statistics

META-SHARE

2,172 language resources at your disposal

Type in your keywords, please...

Search



META-SHARE

What is it? - About the project

META-SHARE, the open language resource exchange facility, is devoted to the sustainable sharing and dissemination of language resources (LRs) and aims at increasing access to such resources in a global scale.

META-SHARE is an open, integrated, secure and interoperable sharing and exchange facility for LRs (datasets and tools) for the Human Language Technologies domain and other applicative domains where language plays a critical role.

META-SHARE is implemented in the framework of the [META-NET Network of Excellence](#). It is designed as a network of distributed repositories of LRs, including language data and basic language processing tools (e.g., morphological analysers, PoS taggers, speech recognisers, etc.).



Pökk fyrir áheyrnina!

META  NORD

META  NET