Access Management at Kielipankki – The Language Bank of Finland

Martin Matthiesen

DELAD Workshop Cork, 16.11.2017
Contents

• CSC?
• A use case example: HS.fi
• Registration of new resources
• Application and Approval Process
• Time for questions
CSC – IT Center for Science

• Non-profit, (mostly) publicly funded company
• Owned by the Finnish ministry of education
• Supporting research & education institutions
  o “We don’t do research, we support research.”
  o Network (FUNET)
  o High-performance super computing
  o Data services
• Home to E-Infrastructures like Elixir, Kielipankki
The Use Case HS.fi

- Helsingin Sanomat newspaper corpus  
- 94 000 articles
- 600,000 sentences
- 8 Million words
- Licensed with the CLARIN ACA +NC license
- Available at https://korp.csc.fi/
Ingredients

We need: The data, a contract and

<table>
<thead>
<tr>
<th>Website</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>lbr.csc.fi (LBR)</td>
<td>“Language Bank Rights” Access Management</td>
</tr>
<tr>
<td>korp.csc.fi (KORP)</td>
<td>The corpus viewer</td>
</tr>
<tr>
<td>metashare.csc.fi (METASHARE)</td>
<td>Metadata management</td>
</tr>
<tr>
<td>pid.csc.fi (PID)</td>
<td>PID management (URN/Handle)</td>
</tr>
<tr>
<td><a href="http://www.kielipankki.fi">www.kielipankki.fi</a> (PORTAL)</td>
<td>License information</td>
</tr>
</tbody>
</table>
Prepare registration of corpus

• Create metadata in METASHARE
• Assign Persistent Identifier, like a URN/Handle (here: URN:HS.fi) in PID
• License text available in PORTAL
• Register corpus in LBR
Dependencies

METASHARE
Metadata server and search

KORP
The corpus server

LBR
Manage access via REMS

URN:HS.fi = metashare entry

LBR sends entitlements for EPPN queried by KORP

URN:HS.fi

Show Metadata

URN:HS.fi

URN:License

Show Terms/License

Applicant

PORTAL
www.kielipankki.fi
License information

Approver
Create metadata in METASHARE

Register a URN in PID ("URN:HS.fi") to point to this page
Make the license text available in PORTAL using templates

<table>
<thead>
<tr>
<th>License type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Choose the appropriate CLARIN license category. Note that RES implies ID, PLAN, BY, NORES, AOA implies NOREAD, to change add the license template.</td>
</tr>
<tr>
<td>○ CLARIN PUE</td>
</tr>
<tr>
<td>□ CLARIN ACA</td>
</tr>
<tr>
<td>○ CLARIN RES</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Resource name (EN) *</th>
</tr>
</thead>
<tbody>
<tr>
<td>The English name of the resource. Must be identical to the name used in Metashare and the Portal’s corpus list.</td>
</tr>
<tr>
<td>The HS.fi News and Comments Corpus</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Kieliarvo nimi (FI) *</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Finnish name of the resource. Again ensure it is the same as in Metashare and /aainolost.</td>
</tr>
<tr>
<td>HS.fi-uutiskomenttelainesto</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>URN of resource *</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fill in the URN (without “<a href="http://urn.fi/%E2%80%9D">http://urn.fi/”</a>, start with “urn:nbn”) to the Metashare article that describes all corpus variants covered by this license.</td>
</tr>
<tr>
<td>urn:nbn:fi:2014052716</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Copyright holder *</th>
</tr>
</thead>
<tbody>
<tr>
<td>The copyright holder according to Metashare</td>
</tr>
<tr>
<td>Helsingin Sanomat, Jutta Saiminen</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tags: ID &amp; Access</th>
</tr>
</thead>
<tbody>
<tr>
<td>Select restrictions for “Identification and Access Conditions”.</td>
</tr>
<tr>
<td>☑ ID: The user needs to be authenticated or identified</td>
</tr>
<tr>
<td>○ AFFIL=EDU: The user needs to be affiliated with a community of researchers through a university or research institution.</td>
</tr>
<tr>
<td>○ AFFIL=META: The user needs to be affiliated with the general community of digital research and technology researchers.</td>
</tr>
</tbody>
</table>
CLARIN ACA end-user license +NC 1.0


Copyright holder: Helsingin Sanomat, Jutta Salminen

The Copyright holder grants the End-User a free, non-exclusive and perpetual (for the duration of the copyright such, as modified, or as part of a compilation or derived work. The permission applies to all known or future modifications of the Resource on other devices and in other formats. Distribution of copies is not allowed.

Additional license terms as defined in the Terms of Service Agreement:

Identification and Access Conditions

• ID: The user needs to be authenticated or identified.

General Use conditions

• BY: Attribution, i.e. acknowledgement of authorship, is required.
• NC: The content is available for non-commercial purposes only.

Distribution conditions

• NORED: The user is not permitted to redistribute the resource.

This license has been made in compliance with copyright agreements by WIPO - the World Intellectual Property Organization. Other intellectual property laws grant rights not mentioned in this license, they are also regarded as part of the rights granted under different legal systems. Additional rights to the Resource may be agreed separately in writing.
Language Bank Rights at Kielipankki

- Manages applications, access rights
- Instance of CSC’s Resource Entitlement Management System REMS
- REMS was originally developed for Bioscience (Elixir)
- Used at Kielipankki since April 2015 as Language Bank Rights (LBR)
- Controls 47 resources (10/2015: 35 resources)
Register the corpus in LBR

• Register the license(s).
• Register HS.fi as a “resource” (using “URN:HS.fi”).
• Create forms to be filled in by applicants.
• Create a workflow that is followed during an application.
• Connect all of the above to a “Catalog item”.
Registration of corpus completed

- Created metadata in METASHARE
- Assigned URN/Handle (here: URN:HS.fi) in PID
- License text now available in PORTAL
- Registered corpus in LBR
Language Bank Rights (LBR) Application and Approval

Applicant’s view

- Name is link to URN (to metadata)
- Applicant can add items to basket (webshop metaphor)
The application

- Fill out form
- Accept licenses
- "Submit" sends application to approvers defined in workflow
The approver view (1/4)

- The application itself
The approver view (2/4)

- Applied items (could be more)

<table>
<thead>
<tr>
<th>ID</th>
<th>Catalog item</th>
</tr>
</thead>
<tbody>
<tr>
<td>62</td>
<td>The HS.fi News and Comments Corpus</td>
</tr>
</tbody>
</table>
The approver view (3/4)

- Applicant
- More info available behind name.
The approver view (4/4)

- History of application
- Who what when
- Comments can be added

(Publications:
Applicant can augment application later with relevant publications, skipped here)

<table>
<thead>
<tr>
<th>Actor</th>
<th>Description</th>
<th>Timestamp</th>
</tr>
</thead>
<tbody>
<tr>
<td>martin.mathiesen</td>
<td>Application created</td>
<td>5.3.2015 14:32:05</td>
</tr>
<tr>
<td>martin.mathiesen</td>
<td>Licence KieliPaniki's General Terms of Use has been approved</td>
<td>5.3.2015 14:32:05</td>
</tr>
<tr>
<td>martin.mathiesen</td>
<td>Licence CLARIN ACA license HS:ii has been approved</td>
<td>5.3.2015 14:32:05</td>
</tr>
<tr>
<td>tero.alto</td>
<td>Application awaiting approval</td>
<td>5.3.2015 14:32:05</td>
</tr>
</tbody>
</table>

The comments are visible to all parties.

Add
The applicant’s view

- Status of application(s)
- Applicant can Close or Update
Recap: Language Bank Rights at Kielipankki

- Based on Resource Entitlement Management System (REMS)
- Connects
  - resource users
  - approvers / licensors
  - licenses
  - resources
- Keeps electronic **paper trail**
- Heavily relies on **SAML2 Single Sign On** (eduGAIN, local federations, CLARIN IdP, Finnish Eduuni)
Thank you!

• The Language Bank: https://www.kielipankki.fi/language-bank/

• Language Bank Rights: https://lbr.csc.fi/

• Demo: https://remsdemo.csc.fi/

• Time for comments, questions.