Center Registry Specification

<table>
<thead>
<tr>
<th>Identifier</th>
<th>CLARIND-AP3-001</th>
</tr>
</thead>
<tbody>
<tr>
<td>AP</td>
<td>3</td>
</tr>
<tr>
<td>Authors</td>
<td>Dieter Van Uytvanck, Menzo Windhouwer, Peter Wittenburg, Thomas Zastrow</td>
</tr>
<tr>
<td>Responsible</td>
<td>Menzo Windhouwer</td>
</tr>
<tr>
<td>Reference(s)</td>
<td><a href="http://trac.clarin.eu/wiki/CenterRegistry">http://trac.clarin.eu/wiki/CenterRegistry</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Date</th>
<th>Changes by</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>Version 1</td>
<td>2011-12-12</td>
<td>Menzo Windhouwer</td>
</tr>
<tr>
<td>Version 2</td>
<td>2011-12-14</td>
<td>Dieter Van Uytvanck</td>
</tr>
<tr>
<td>Version 3</td>
<td>2011-12-15</td>
<td>Menzo Windhouwer</td>
</tr>
<tr>
<td>Version 4</td>
<td>2011-12-21</td>
<td>Dieter Van Uytvanck</td>
</tr>
<tr>
<td>Version 5</td>
<td>2012-01-16</td>
<td>Menzo Windhouwer</td>
</tr>
</tbody>
</table>

Introduction

The centers form the backbone of the distributed CLARIN-D infrastructure. So far the administration of the center details has been done manually, noting down the entry points and contact information on a web site. In order to streamline the information gathering and to ensure machine accessibility of this data a center registry needs to be built. In this document we specify the requirements for such a registry, in accordance with the CLARIN-D AP3 guidelines.

Requirements

The center registry is a store for records with information about CLARIN centres. This document describes the data that it could contain (Data store), the web application for users to view and change the data (Web Interface) and the machine-accessible API (Web Service).

1. Data store

Each record has these fields:

1.1 Basic Information

- name (string)
- country (controlled vocabulary: subset of ISO 3166-1 alpha-2)
- center type (controlled vocabulary: A, B, C or E)
- center type status (0 or 1; string; to be used to describe the A/B/C/E center status)
- core description (string; possible multiple lines containing some general information on the center, e.g., the collection focus or policy)

1.2 Extended Information

- Long term archiving policy (string)
- CLARIN infrastructure services (0 or 1; string; meaning: non-standard offers like OAI harvesting)
• strict versioning (boolean; meaning: 1 PID always points to 1 version)
• repository system (string)
• PID status (string)
• assessment status (string)
• website (URI)
• references (1 or more; description+URI):
  o description
  o URI, e.g., to link to an assessment reports
• metadata (1 or more):
  o OAI access point (URI)
  o Metadata scheme (1 or more; controlled vocabulary: OLAC or CMDI)
  o Web Services Set (0 or 1; string; the OAI-PMH set that contains the
    webservice metadata records) - if a Web Services set is specified, 1 or
    more Web Service Types must be defined, otherwise there is no Web
    Service Type at all
  o Web Service Type (0 or more; controlled vocabulary: SOAP, REST or
    WebLicht)
• AAI:
  o AAI status (string)
  o member of a national identity federation (0 or more; controlled
    vocabulary: subset of ISO 3166-1 alpha-2)
  o member of the CLARIN Service Provider Federation (boolean)
• organisational information:
  o organisation details (see http://www.clarin.eu/organisations/3434
    for an example)
  o contact information (name, email, phone) of a technical contact person
  o contact information (name, email, phone) of an administrative contact
    person
• monitoring:
  o Service Provider test site (URI, link to the shib_test.pl script)
  o (As the nagios probes will be configured at the monitoring computer
    center no further fields required)

Empty values shouldn’t be allowed, e.g., if the Web Services Set field is left empty it isn’t
instantiated and this is the same as an instance count of zero. It is very probable that
new fields will be needed in the future, so the data model should be flexible. Also it
should be easy to add new values to controlled vocabularies, e.g., to add a new allowed
Web Service Type.

2. Web interface
• initial view: a table like http://www.clarin.eu/centers with a link to the center
  view for each center
• center view: display all fields of a single center
• center editing screen: a webform to change the values of a center
• administrative editing screen: manage controlled vocabularies
• for each field and value (from a controlled vocabulary) it should be possible to
  link to an (external) explanatory web page
• authorization for write-access: SAML-based, initially via the
  eduPersonPrincipalName attribute as received from the CLARIN IdP:
  o everyone should have read access
  o a few administrators (dietu@mpi.nl, menwin@mpi.nl,
    thomas.zastrow@uni-tuebingen.de) have write access
3. Web service

A REST-webservice returns information from the registry as:
- JSON
- CMDI (=XML, for the details of this format see below)

(Inspired by the web service for the Component Registry)

As the public REST-webservice is read-only no authorization should be required.

A CMDI profile has been created:

- link to CMDI profile: http://catalog.clarin.eu/ds/ComponentRegistry?item=clarin.eu:cr1:p_1320657629667
- example instance (.cmdi): http://trac.clarin.eu/attachment/wiki/CenterRegistry/http_www_mpi_nl.cmdi

To start, the following functions should be available:

- return a list of all centers (together with a unique center ID): http://foo.eu/rest/centerregistry/centers
  - optional parameters: /xml (default, ad hoc xml encoding) or /json
  - output: center ID, Basic information fields, if there is at least 1 OAI access point, providing metadata about web services
- return all information about 1 center: http://foo.eu/rest/registry/centers/{centerid}
  - optional parameters: /cmdi (default) or /json
  - output: all fields in the center record

Documentation about the webservice:
- as HTML
- as a WADL file (like eg: http://catalog.clarin.eu/ds/ComponentRegistry/rest/application.wadl)