Spatial Information Services Stack (SISS) Registry
Where identifiers come together.

Terry Rankine, Guillaume Duclaux
Earth Science and Resource Engineering
CSIRO
Perth, Australia
Terry.Rankine@csiro.au, Guillaume.Duclaux@csiro.au

Abstract—The Spatial Information Service Stack (SISS) [1] Registry Service is crucial to the SISS proposed methodology and provides the ability to ask the first question posed by user of the technology: “What is available?”: Combined with any community agreed standards for data entry, persistence and ultimately quality, communities are now able to share data collections and data services based on Open Geospatial Consortium (OGC) [2] standards with ease. The Catalogue Service for the Web (CS-W) [3] interface to the registry enables multiple clients to discover this metadata collection and its search services via open standards. The AuScope Discovery Portal [4] makes direct use of the SISS CS-W in the presentation of known features from multiple locations for end user consumption.

Geoscience; Registry; OGC; SISS; CS-W; interoperability; GeoNetwork

I. INTRODUCTION

A registry usually exists to provide a formal mechanism for users to announce or discover the availability and state of a resource such as a data collection, a service or a relationship. Such a registry has different requirements for its three main types of users: (1) the resource user views it as a discovery interface, which summarises the resource and provide the connection point to consume it, (2) the resource provider’s view is one of an advertising service, it can announce to the resource users the existence, availability, and next steps to use the resource, (3) the registry manager view is to protect the community, and see inclusion of data into the registry as conditional. The latter decides on the conditions of use for the community, and the needs for quality assurance (QA) and quality constraints (QC). The registry manager decision making process is facilitated by a governance framework designed to assist him taking actions that will support the community and help them to continue to work together.

A registry is often associated with managed repositories. Nevertheless, the need for a registry exists regardless of the existence of managed repositories within an organisation or community. Indeed, community managed repositories exist for this very reason, and in this case, users within the community adopt the roles as registry managers as well as users. This enables the community to utilise the registry within their scientific workflows and in turn the registry helps facilitate the data reuse, data sharing paradigm back to the community. A good example of a community managed repository is illustrated within the AuScope Project [5] and its geoscientific registry/repository solution in its SISS implementation. AuScope is using SISS to deploy this registry service which is a complete OGC Catalogue Service. It provides a web interface to a registry of geoscience data sets; a spatial information discovery portal; and SISS-based access to a number of spatial data repositories including GPS network station and observation logs, seismic data, geological models and maps, and the Australian National Virtual Core Library (NVCL) [6].

In SISS, the principal objective of a registry is extremely simple, and can be summarised in Fig. 1; it is to connect the resource user with the resource provider(s) which can answer their questions. To achieve this, the SISS registry must provide enough information to ensure the resource user can firstly identify what resource provider(s) are capable of, and secondly, have the details required to make valuable use of the provider’s resource. In the case of a SISS enabled resource: (1) the OGC service standards, (2) the service URL, (3) the web service capabilities statement (GetCapabilities()) response which is harvested by the registry, come together to form enough information for the resource to be used by other clients. This can also be seen in Fig. 2.

![Figure 1: Registry role in SISS](image)

II. REGISTRY IMPLEMENTATION

GeoNetwork [7] provides a web based registry application that provides searching, editing, managing, harvesting and sharing metadata. The metadata is built on the following
standards, Dublin Core [8], ISO19115 [9], ISO19139 [10], FGDC (Federal Geographic Data Committee) [11]. The metadata is accessible in multiple access modes: a rendered mode, HTML for the browser; a raw mode, XML/HTTP for generic clients; and a catalogue mode, Z39.50 [12] protocol access. In the web user interface, stylesheets are used to convert the XML metadata for visualisation, searching, and editing. Rules for data validity are provided by the underlying metadata schemas, and are enforced inside the interface. Harvesting other providers’ resources (other registries or OGC services) is also handled by the GeoNetwork engine, and there is a large unified search profile provided by Apache Lucene [13] on the entire catalogue. SISS makes significant use of GeoNetwork’s harvester, and its CS-W interface to its metadata store as shown in Fig. 2.

Numerous other catalogue applications exist (GI-Cat [14]) and there are also a number of registries (Buddata ebXML Registry/Repository [15]) which could be used within SISS. Reviewing them is out of scope of this paper, but the comparisons can be found on the SeeGRID Wiki [16].

III. REGISTRY CONTENTS

A. Data Records

The data records are built on standard metadata templates (Dublin Core, ISO19115, ISO19139, and FGDC) and allow for static thematic vocabularies to populate the manual data entry forms when those latter are available. While manual record entry is possible, it requires an ongoing watchful eye of the creator, maintainer or registry manager to ensure the record does not change significantly, making the metadata useless. The preferred mode for data records, in the SISS registry is harvesting. The harvested data records are directly encapsulated from remote managed repositories, where the record and its management both lie with the curator of the data. The process of harvesting and the agreed level of access is a on a per party basis with the community.

B. Service Records

Service records can also be entered manually but the community QA/QC model should encourage automatic handling for services. Using the harvesting mechanism in the registry, the services can be regularly queried to provide basic metadata (including spatial data), service features and query protocols which can be automatically updated. It allows for resource providers to have the option of automatic inclusion of new data, updated service information, or even extended functionality. Additional information can be added to harvested records such as mappings to known community vocabularies or ‘like’ services. The richness of service records can be improved further with GeoNetwork’s custom service harvesting of OGC services to include snapshots of graphical outputs, and even identification of known community application schemas.

IV. ACCESSING THE REGISTRY

The registry requirements in SISS were designed to be generic, which enables other choices in registries to be used. SISS has focused on standards based access to the registry, and has chosen not to use custom interfaces (such as the

![Figure 2: SISS Usage of the GeoNetwork Architecture](image-url)
GeoNetwork Web User Interface, or the GI-Cat User Interface). Machine to machine registry interaction is done making use of the OGC CS-W standards, and GeoNetwork provides a direct CS-W interface to its registry (Table 1).

<table>
<thead>
<tr>
<th>Operations available via CS-W</th>
<th>Operations provided</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Capability class label</strong></td>
<td><strong>OGC_Service.GetCapabilities()</strong></td>
</tr>
<tr>
<td><strong>CSW</strong></td>
<td><strong>CSW Discovery.DescribeRecord()</strong></td>
</tr>
<tr>
<td></td>
<td><strong>CSW Discovery.GetRecords()</strong></td>
</tr>
<tr>
<td></td>
<td><strong>CSW Discovery.GetDomain()</strong> (optional)</td>
</tr>
<tr>
<td></td>
<td><strong>CSW Discovery.GetRecordById()</strong></td>
</tr>
</tbody>
</table>

Building on the CS-W interface (Table 1), we developed code and libraries which enable queries, filter construction, and result processing for the AuScope Discovery Portal, which can also be reused by other clients. The DescribeRecord() operation allows a client to discover elements of the information model supported by the target catalogue service. These elements can then be built into filter queries which form the search capabilities of the GetRecords() method. GetRecords() does a search, based on the previous filter and the profiled (brief, summary, full) results are returned. By default, CS-W provides two reduced profiles for the registry items by defining ElementSetName: “brief” and “summary”, which are appropriate during the search process. The records IDs are then fed to GetRecordById() queries using “ElementSetName=full” to include the full metadata profile. The returned result contains all elements defined by either standard, ISO19115 or ISO19139 allowing the richness of the full record to be used within other clients.

V. WORKING WITH COMMUNITIES

Communities are loosely based on science domains (i.e. AuScope – focused on geosciences), but often exist purely on data consumption (i.e. Bureau of Meteorology – providing precipitation data). This adds complexity on forming a governance model, and but does not prevent it from being developed.

Without a community based governance model operating within the registry, the QA/QC are unable to function, and the trust the community had in the records and no longer sees the value in maintaining such a resource. The governance model is like the registry, both extend beyond organisational bounds, and operate in the same manner. The governance model will vary, and with multiple parties it will not be simple to establish, however the effective management of the registry demands an effective model to ensure value is returned to the community.

This community based approach for the registry does not limit it to a single instance; there is a need for linking communities together. The SISS registry service is harvestable by Australian National Data Service (ANDS) [17], which enables the registry to be realised beyond the current targeted community, resulting in greater data discovery and reuse of existing resource providers as shown in Fig. 3. Researchers can now cross community boundaries in their quest for data, and in turn can expand their science to make use of the newly discovered resources. It should be noted that the current practice of harvesting content from the geospatial representation inside the SISS registry (ISO19139) into the ANDS catalogue (RIF-CS [18]) does not capture all of the information or the inter-relationships present in ISO19139. The ANDS catalogue draws on the common elements in the scientific datasets and provides pointers back to the individual scientific community registries.

VI. PERSPECTIVE

The registry service is still not adequately integrated with other components in SISS. There is no current way to make use of the SISS Vocabulary service inside the registry. Ontologies access provided by the Vocabulary service would provide much richer linkages within the registry, and then between registries.

The registry service can also be improved by storing metadata about the information models supported by the services. This would enable more accurate selection of services and once again allow smarter links between services.

The registries internal metadata store could also be significantly improved if newer XML database technologies (eXist-db [19] or Apache Xindice [20]) where used, as it would avoid re-processing the records into the complete documents for string storage in the existing storage model.

Persistent identifiers such as URNs [21] or ANDs Persistent ID service [22] should be integrated into the registry. This would enable the resources to be shared in a much broader scope, improve the QA/QC when performing registry harvesting, and simplify the linking between resources.

REFERENCES

  http://www.auscope.org.au/content.php/category/id/15
[8] Dublin Core Metadata Initiative. Dublin Core Metadata Element Set,
    Geographic information -- Metadata -- XML schema
    Geospatial Metadata (version 2.0). [Online].
    http://www.fgdc.gov/standards/projects/FGDC-standards-
    projects/metadata/base-metadata/v2_0698.pdf
    retrieval (Z39.50) -- Application service definition and protocol
    http://code.google.com/p/buddata-exml-registry/
[16] SeeGRID. SeeGRID Wiki. [Online].
    https://www.seegrid.csiro.au/wiki/bin/view/Main/WebHome
[18] International Standards Organisation. Registry services for libraries
    and related organizations. [Online].
    http://www.iso.org/iso/catalogue_detail.htm?csnumber=44936
    http://exist.sourceforge.net/
    http://xml.apache.org/xindice/
[21] W3C. URIs, URLs, and URNs: Clarifications and Recommendations
    1.0. [Online]. http://www.w3.org/TR/uri-clarification/
[22] ANDS. ANDS Persistent Identifier Service. [Online].