From silos to systems

Jonathan Hodge
Project Leader
Environmental Information Systems

Nicholas Car
Project Manager
Environmental Information Systems
Data evolution in Australia

• Large amounts of high quality environmental data sit with government and research organisations

• Many of these organisations tend to be risk averse and slow to adopt technology

• Strong push in recent years towards open data and open licensing

• Major national investments are changing the way that Australia delivers data
Data evolution in Australia

• Many now deliver OGC services
• Map and metadata services common
• More complex services less common
• Proliferation of silos of information
• Proliferation of portals
• Minimal cross-domain or cross-organisation integration
Information evolution in Australia

• Some systems, such as the AODN, have started to aggregate these services

• Need to allow these service silos to deliver more widely
An integration example – websites

• Build a website, put online

• Search engines index and allow discovery

• But, you can have more control

• ‘Webmaster tools’ (sitemap, owner verification, etc)
  – Controlled by provider
  – More correct
  – Automated verification and testing
Webmaster tools for data?

• Data services are rarely indexed
• If they are, it will only be the top level
• If a crawler did go deeper, what should it look at?

• Things we want to know:
  • How else can I access the data (html, json, xml, other)?
  • How do I find the metadata?
  • Where did the data come from?
  • Is there a consistent answer to these questions for:
    – Different datasets from one service?
    – Different services from one provider?
    – Different providers?
Need some rules

• Need to develop a structured approach to discovering information about data services

• Need to answer what, how and where (and more)

• What is available?

• How do I access it?

• Where is it? Where has it come from?
eReefs Data Provider node

Data Provider Node Template

- Brokering Layer
  - Data services
  - Metadata service
  - Vocabulary Service
  - Information Model Catalog
  - Provenance service

CISRO Hydrodynamic Data Provider

- Linked Data API via a Persistent ID Service
  - Coverages via THREDDS
  - ANZLIC metadata OGC Catalog Service
  - WQ terms via SKOS / SISSVoc
  - Hyd’namic information models via Solid Ground
  - PROMS (in development)
Data Provider Node Ontology - Classes

DataProviderNode

Dataset
- DataService
  - WebFeatureService
  - SensorObservation Service
  - WebCoverageService

- MetadataService
  - CatalogServiceWeb

- VocabularyService
  - SISSVoc

- InfoModelService
  - FeatureTypeCatalog

- ProvenanceService
  - ProvMgtSysService

Website
Data Provider Node Ontology - Classes
Dataset Class

:Dataset

a
rdfs:subClassOf :DataProviderNode
rdfs:label "\{name\}\"^^xsd:string
skos:prefLabel "\{name\}\"^^xsd:string
:serviceEndpoint "\{uri\}\"^^xsd:anyURI
:isPublic "\{true|false\}\"^^xsd:boolean
dctterms:license "\{license_uri\}\"^^xsd:anyURI

From silos to systems  |  Jonathan Hodge and Nicholas Car11 |
Example Data/Metadata class instances

:WebCoverageServiceOcBurdekin
  a :WebCoverageService ;
  :isPublic "true"^^xsd:boolean ;
  :isServiceFor :DatasetOcBurdekin ;
  :dcterms:license : "http://creativecommons.org/licenses/by/3.0/" ^^xsd:anyURI

:CatalogServiceWebOcBurdekin
  a :CatalogServiceWeb
  :isPublic "true"^^xsd:boolean ;
  :isServiceFor :DatasetOcBurdekin ;
  :dcterms:license : "http://creativecommons.org/licenses/by/3.0/" ^^xsd:anyURI
## OWL Class to Persistent Identifier Service

<table>
<thead>
<tr>
<th>OWL Classes</th>
<th>PID Service Patterns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataset</td>
<td>_view=alternates</td>
</tr>
<tr>
<td>DataService</td>
<td>_view=data</td>
</tr>
<tr>
<td>MetadataService</td>
<td>_view=meta</td>
</tr>
<tr>
<td>VocabularyService</td>
<td>_view=vocab</td>
</tr>
<tr>
<td>InfoModelService</td>
<td>_view=infomodel</td>
</tr>
<tr>
<td>ProvenanceService</td>
<td>_view=prov</td>
</tr>
</tbody>
</table>
Consistent, predictable end points

For example, a dataset:

will have a data service URI of:
(Instead of the original THREDDS end point)

and a metadata service URI of:
(Instead of the original GeoNetwork end point on different server)
# PID Service mappings

<table>
<thead>
<tr>
<th>Mapping</th>
<th>Path / Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burdekin River Estuary</td>
<td>/datasets/BDKN-B09</td>
</tr>
<tr>
<td>Datasets Index</td>
<td>/datasets/</td>
</tr>
<tr>
<td>Fitzroy River Estuary</td>
<td>/datasets/FITZ-B09</td>
</tr>
<tr>
<td>Pioneer river at Mackay</td>
<td>/datasets/MCKY-B09</td>
</tr>
</tbody>
</table>
<path>/datasets/BDKN-B09</path>  
<title>Burdekin River Estuary</title>

<conditions>
  <condition>
    <type>QueryString</type>
    <match>_view=alternates</match>
    <actions>
      <action>
        <type>302</type>
        <name>location</name>
      </action>
    </actions>
  </condition>
  <condition>
    <type>QueryString</type>
    <match>_view=data</match>
    <actions>
      <action>
        <type>302</type>
        <name>location</name>
        <value>http://thredds0.nci.org.au/thredds/catalog/u83/modis/ereefs/mwq/interim/BDKN-B09/catalog.html</value>
      </action>
    </actions>
  </condition>
...
## Ocean Colour Data Provider Node

### Datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Data</th>
<th>Metadata</th>
<th>Vocabulary</th>
<th>Information Models</th>
<th>Provenance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burdekin River Estuary</td>
<td>data</td>
<td>metadata</td>
<td>vocabulary</td>
<td>info models</td>
<td>provenance</td>
</tr>
<tr>
<td>Fitzroy River Estuary</td>
<td>data</td>
<td>metadata</td>
<td>vocabulary</td>
<td>info models</td>
<td>provenance</td>
</tr>
<tr>
<td>Pioneer River at Mackay</td>
<td>data</td>
<td>metadata</td>
<td>vocabulary</td>
<td>info models</td>
<td>provenance</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Dataset</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDKN-B09</td>
<td></td>
</tr>
<tr>
<td>2012/</td>
<td></td>
</tr>
<tr>
<td>2011/</td>
<td></td>
</tr>
<tr>
<td>2010/</td>
<td></td>
</tr>
<tr>
<td>2009/</td>
<td></td>
</tr>
<tr>
<td>2008/</td>
<td></td>
</tr>
<tr>
<td>2007/</td>
<td></td>
</tr>
<tr>
<td>2006/</td>
<td></td>
</tr>
<tr>
<td>2005/</td>
<td></td>
</tr>
<tr>
<td>2004/</td>
<td></td>
</tr>
<tr>
<td>2003/</td>
<td></td>
</tr>
<tr>
<td>2002/</td>
<td></td>
</tr>
</tbody>
</table>
Barry - Orchestration Service

Orchestration Service test sets

- Metadata conformance
- Vocab term references
- Provenance records
- Data Service configuration
- Service level provision

Basic DPN structure

eReefs structure profile
Multiple use and re-use

Consumer Node
eReefs Main Portal

Other Consumer Nodes

Other Community Central Node

eReefs Community Central Node

Data Provider Node
Hydrodynamic

Data Provider Node
Ocean Colour

Other Data Provider Nodes
Summary

• Provides a mechanism to link existing systems

• Relatively low effort to build the brokering layer

• Relatively low ‘barrier of entry’ for data providers

• This gives us a true, distributed architecture in a manageable, configurable, linked structure

• Moves us from stand-alone silos to interoperable systems