



From silos to systems

Jonathan Hodge

Project Leader
Environmental Information Systems

Nicholas Car

Project Manager
Environmental Information Systems

LAND AND WATER
www.csiro.au



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



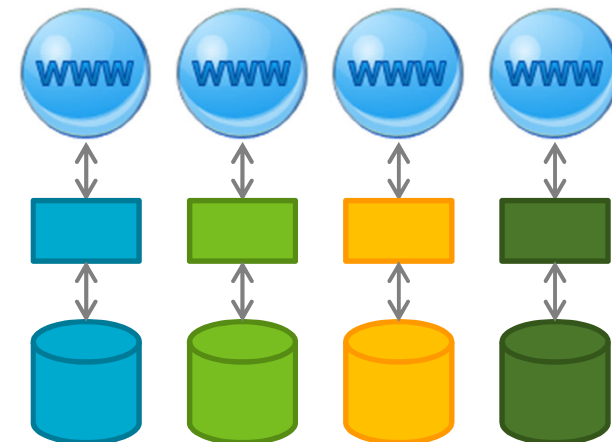
ATLAS OF LIVING AUSTRALIA
sharing biodiversity knowledge



TERN
Terrestrial Ecosystem
Research Network

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?

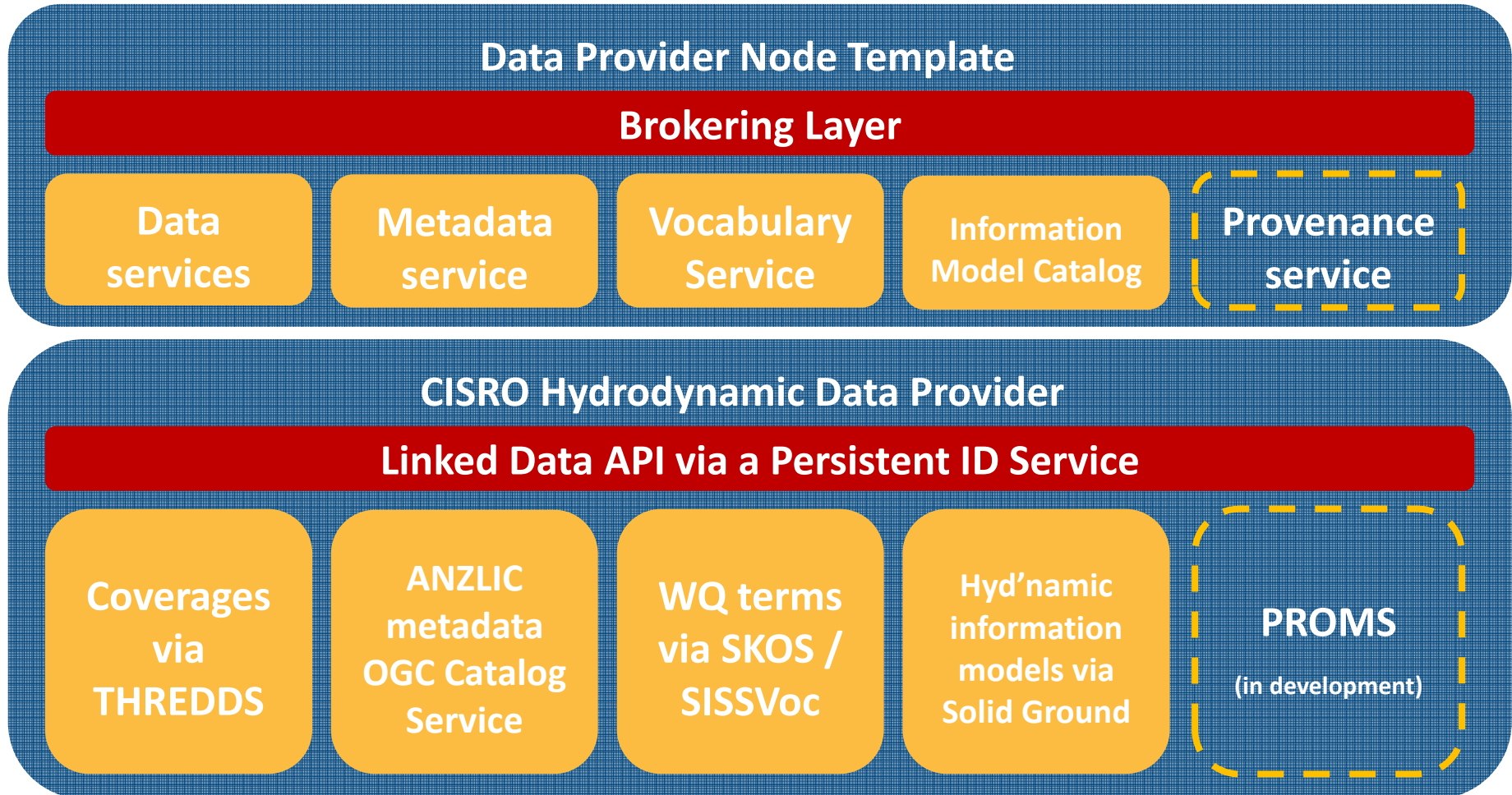
- Data services are rarely indexed
- If they are, it will only be the top level
- If a crawler did go deeper, what should it look at?

- 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?

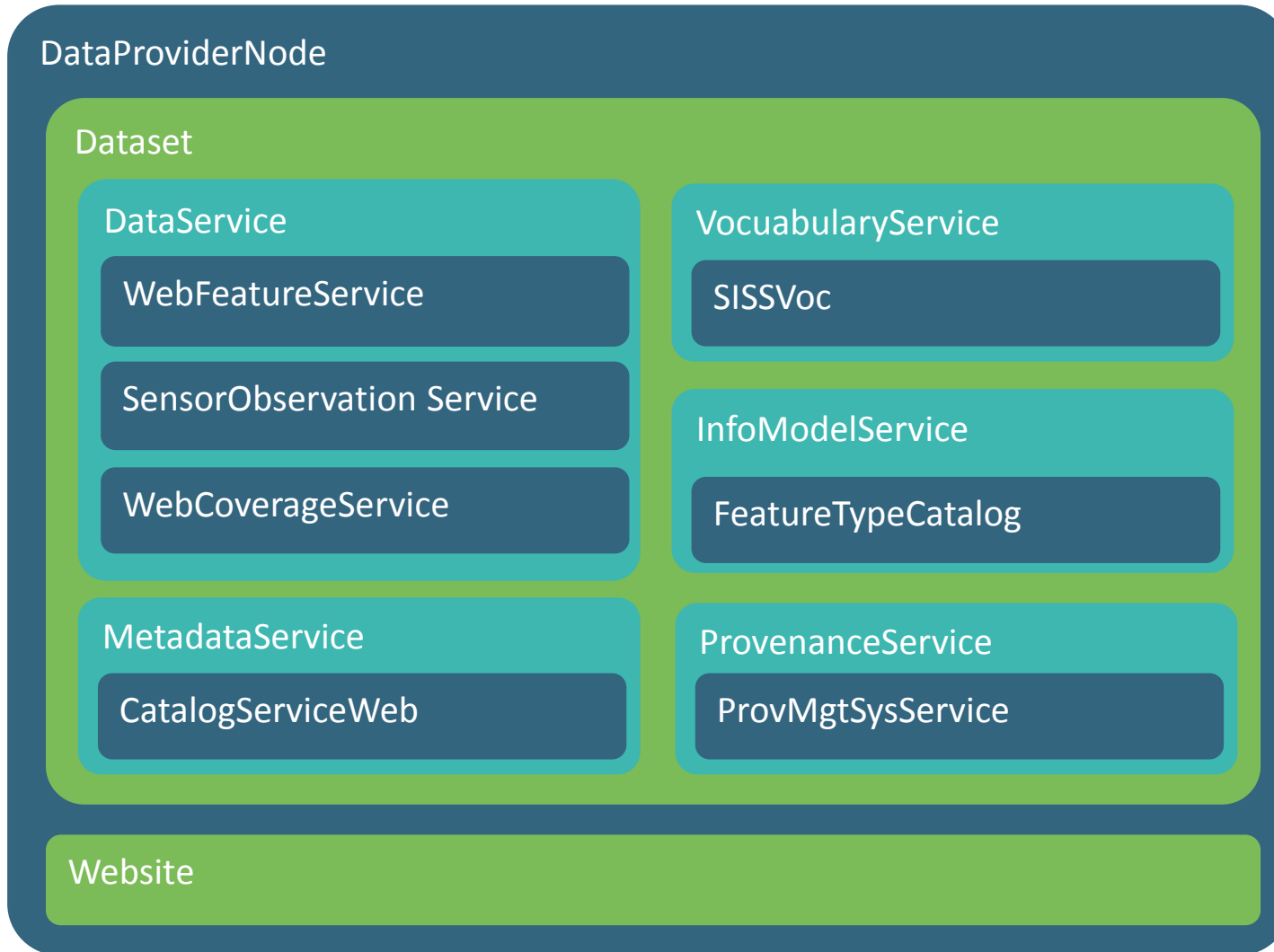
Need some rules

- 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 Ontology - Classes



Data Provider Node Ontology - Classes

- owl:Thing
 - DataProviderNode
 - Dataset
 - DataService
 - FtpService
 - SensorObservationService
 - WebCoverageService
 - WebFeatureService
 - WebMappingService
 - WebProcessingService
 - InfoModelService
 - FeatureTypeCatalog
 - MetadataService
 - CatalogServiceWeb
 - ProvenanceService
 - PromsService
 - VocabularyService
 - SissVocabularyService
 - Website

Dataset Class

:Dataset

a	owl:Class ;
rdfs:subClassOf	:DataProviderNode ;
rdfs:label	"{name}"^^xsd:string ;
skos:prefLabel	"{name}"^^xsd:string ;
:serviceEndpoint	"{uri}"^^xsd:anyURI ;
:isPublic	"{true false}"^^xsd:boolean ;
dcterms:license	"{license_uri}"^^xsd:anyURI ;

Example Data/Metadata class instances

:WebCoverageServiceOcBurdekin

a :WebCoverageService ;
:isPublic "true"^^xsd:boolean ;
:isServiceFor :DatasetOcBurdekin ;
:serviceEndpoint "<http://dpn-oc-vc.nexus.csiro.au/datasets/BDKN-B09>" ^^xsd:anyURI
:dcterms:license : "<http://creativecommons.org/licenses/by/3.0/>" ^^xsd:anyURI

:CatalogServiceWebOcBurdekin

a :CatalogServiceWeb
:isPublic "true"^^xsd:boolean ;
:isServiceFor :DatasetOcBurdekin ;
:serviceEndpoint "<http://dpn-oc-meta-vc.nexus.csiro.au/geonetwork/srv/eng/csw>"
^^xsd:anyURI
:dcterms:license : "<http://creativecommons.org/licenses/by/3.0/>" ^^xsd:anyURI

OWL Class to Persistent Identifier Service

OWL Classes

PID Service Patterns

Dataset → `_view=alternates`

DataService → `_view=data`

MetadataService → `_view=meta`

VocabularyService → `_view=vocab`

InfoModelService → `_view=infomodel`

ProvenanceService → `_view=prov`

Consistent, predictable end points

For example, a dataset:

<http://dpn-oc-vc.nexus.csiro.au/datasets/BDKN-B09>

will have a data service URI of:

http://dpn-oc-vc.nexus.csiro.au/datasets/BDKN-B09?_view=data





(Instead of the original THREDDS end point)

and a metadata service URI of:

http://dpn-oc-vc.nexus.csiro.au/datasets/BDKN-B09?_view=metadata

(Instead of the original GeoNetwork end point on different server)

PID Service mappings

Mapping	Path / Pattern
 Burdekin River Estuary	<code>/datasets/BDKN-B09</code>
 Datasets Index	<code>/datasets/</code>
 Fitzroy River Estuary	<code>/datasets/FITZ-B09</code>
 Pioneer river at Mackay	<code>/datasets/MCKY-B09</code>

```

<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>
        <value>http://dpn-oc-vc.nexus.csiro.au/config/datasets/BDKN-B09/alternates</value>
      </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>
  ...
</conditions>

```

Base URI for Dataset (relative to DPN)

alternates view pattern

action (HTTP 302 redirect)

alternates view *action* location

data view pattern

action (HTTP 302 redirect)

data view *action* location

Ocean Colour Data Provider Node













Datasets

http://dpn-oc-vc.nexus.csiro.au/datasets/BDKN-B09?_view=data

Dataset	Service Endpoints				
Name	Data	Metadata	Vocabulary	Information Models	Provenance
Burdekin River Estuary	data	metadata	vocabulary	info models	provenance
Fitzroy River Estuary	data	metadata	vocabulary	info models	provenance
Pioneer River at Mackay	data	metadata	vocabulary	info models	provenance

http://dpn-oc-vc.nexus.csiro.au/datasets/BDKN-B09?_view=metadata

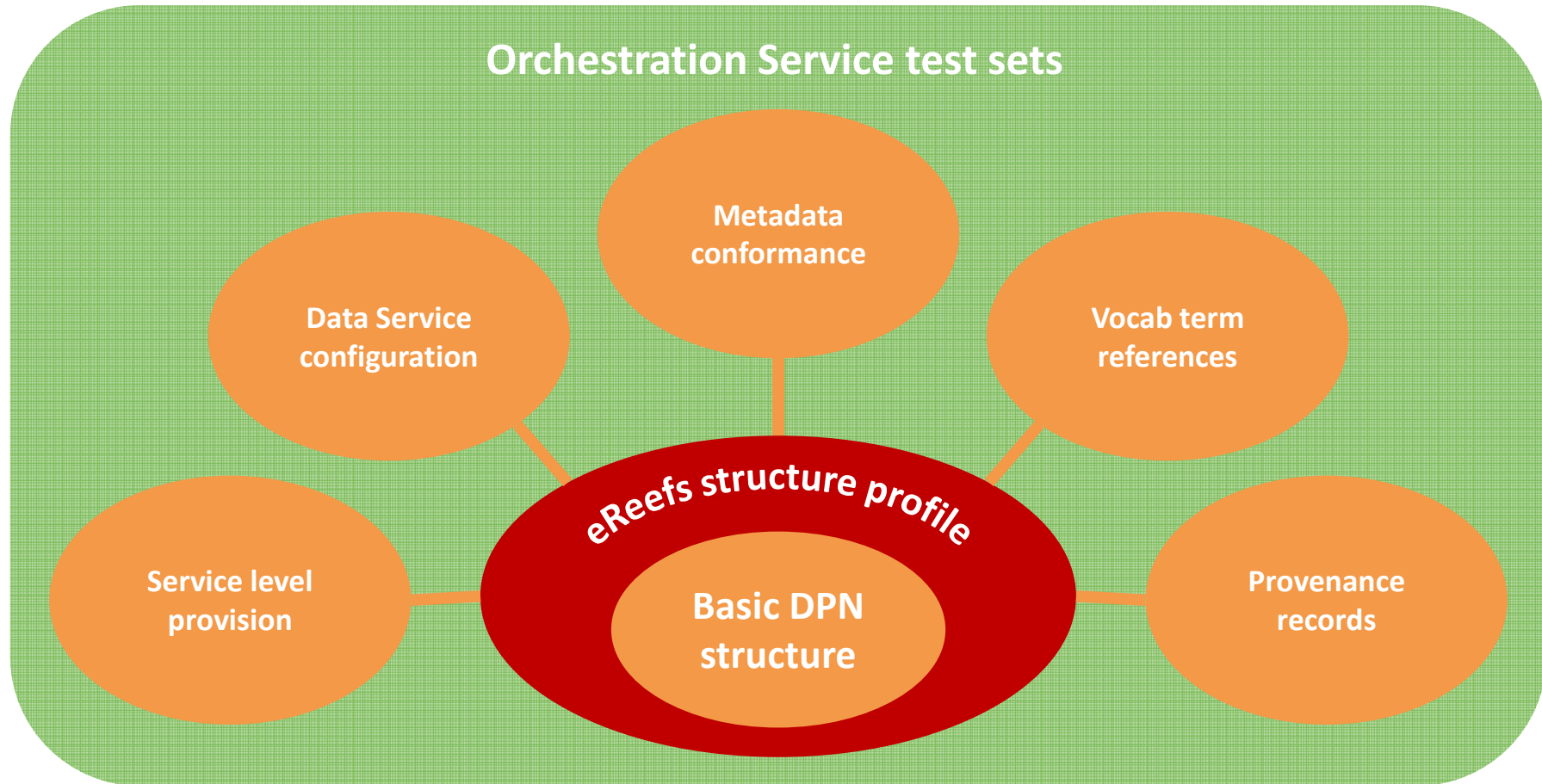
Catalog <http://thredds0.nci.org.au/thredds/catalog/u83/modis/ereefs/mwq/interim/BDKN-B09/catalog.html>

Dataset	Size
 BDKN-B09	
 2012/	
 2011/	
 2010/	
 2009/	
 2008/	
 2007/	
 2006/	
 2005/	
 2004/	
 2003/	
 2002/	

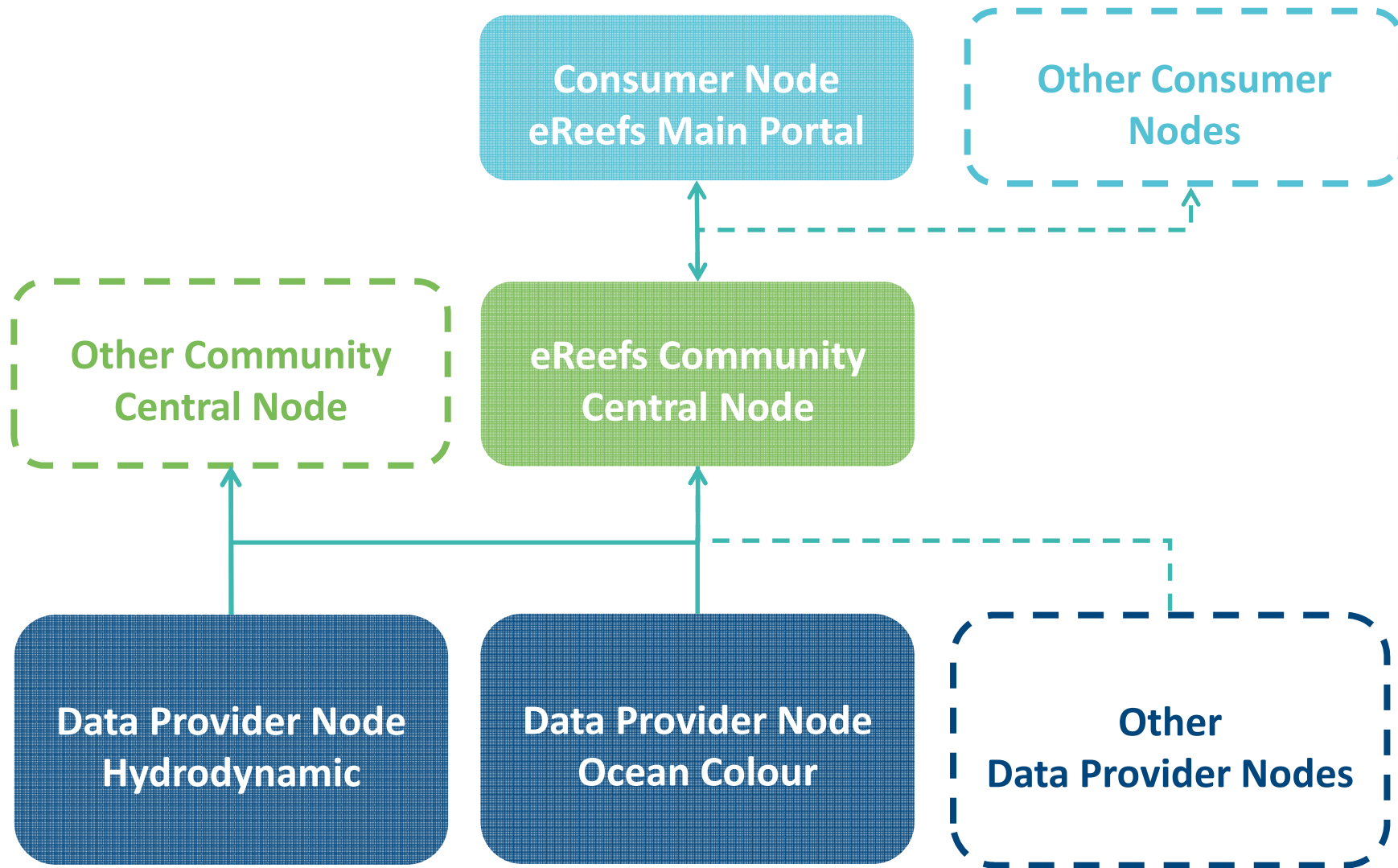
[ANUSF TDS Server at ANUSF](#)
[THREDDS Data Server \[Version 4.2.3 - 20110113.2322\] Documentation](#)



Barry - Orchestration Service



Multiple use and re-use



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

eHeets is a collaboration between:



GREAT BARRIER REEF
foundation



Australian Government
Bureau of Meteorology



Queensland
Government

Thanks to the rest of the team:

David Lemon

Peter Fitch

Simon Cox

Bruce Simons

Jonathan Yu

Shane Seaton

and the project coordinators:

Andy Steven

Andreas Schiller

Jonathan Hodge

Project Leader

t +61 7 3833 5515

m +61 409 577 945

e jonathan.hodge@csiro.au

w www.csiro.au

Nicholas Car

Research Projects Officer

t +61 7 3833 5600

m +61 427 300 292

e nicholas.car@csiro.au

w www.csiro.au

LAND AND WATER

www.csiro.au

