Consuming SISS
a client’s perspective

Joshua Vote
Software Developer
SSSC 2011
Contents – Timeline of this Workshop

• Example of the AuScope Portal – SISS Consumer

• Discussion and examples surrounding:
  • Querying an Information Model;
  • Querying datasets with a standardised vocabulary;
  • How to find data via a registry

• Further Examples
  • Full SISS Consumption Workshop
    https://www.seegrid.csiro.au/wiki/Siss/SissConsumptionWorkshop
Goals for this workshop

• By the end of this workshop you should be able to...
  • Understand how existing workflows/usecases have been aided by SISS
  • Recognise basic queries on OGC web services and other SISS components
  • Integrate the functionality of multiple SISS components to solve real world problems
• This is NOT a programming exercise
  • However the concepts demonstrated are easily automated.
Demonstration

• Best way to start the learning process is to see a real world example.
• The best example we have of a SISS consumer is the AuScope Portal and ERML workflow.
• You don’t need to be a geologist to follow along
  • As long as you can appreciate that ‘Gold’ is valuable and is extracted from the ground
Earth Resource Markup Language (ERML)

• What is ERML?
  • It’s an information model

• What is an Information Model?
  • Put simply for our purposes - allows multiple data providers to all speak the same language.
  • Aids in interoperability

• It contains a number of complex object relationships which allow you to model various (in this case geological) phenomena
  • Such as mines and mineral commodities.
  • Contains spatial attributes.

• In this workshop ‘ERML’ will be referring to the AuScope profile of ERML.
AuScope Portal

• What is the AuScope Portal?
  • Flagship consumer of SISS
  • [http://portal.auscope.org](http://portal.auscope.org)
  • It is **not** a GIS application; it’s built to serve a few distinct usecases

• Implements a workflow based on ERML that utilises SISS.
  • It also supports many other workflows revolving around ‘Data Discovery’.
  • It’s also the basis for various portals built to enable scientific workflows.

• Stepping through the ERML workflow will highlight the various components of SISS
AuScope Portal and ERML

- **Demonstration** - Get me the names of all gold mines near the SA / WA border.
- Portions of the data are held by various state and territory geologic surveys.
- Each survey has their own vocabulary, spatial projection and database schemas.
- How would you do this now?
- What if the question was get me all the gold mines in Australia?
AuScope Portal and ERML

• **Step 1 – Service Discovery**
  • We need data providers that implement ERML
  • SISS includes a registry – Geonetwork
  • Geonetwork can only offer a guess at what services belong to which information model
    • AuScope Portal handles this by manually adjusting any errors.
  • Our query will return all ERML services that the community is aware of.
AuScope Portal and ERML

• **Step 2 – Asking for the right concept**
  • We are looking for ‘**Gold**’. How do we ask for it?
  • Commodities are defined using survey specific definitions
    • The various providers use ‘Gold’, ‘Au’, ‘15’ all for the same commodity.
    • We are lucky that in this example everyone speaks English.
  • Geoscience Australia has a defined commodity vocabulary
    • ERML information model utilises this standardised vocabulary.
  • Before making a request for gold, we need to know how the vocabulary defines ‘Gold’
    • SISSVoc allows us to make this query.
    • The answer is ‘urn:cgi:classifier:GA:commodity:Au’
Step 3 – Querying the information model

ERML has lots of features that just aren’t required for our example.

- Often means there is a LOT of data we just aren’t interested in.

We are only interested in mines and the commodity they produce.

- Specifically the names of mine and a single commodity.

We need to identify the parts of the information model that are relevant to our task.
• **Step 4 – Making the request**
  
  • We know what we want from the information model however…
    • Each survey will store the data differently (database schemas)
    • Each survey may use different spatial reference systems
  
  • We need a way of mapping or translating our query to the storage model.
    • Geoserver & Application Schemas plugin defines this mapping.
    • Translates in real time between information and storage models.
    • Data providers don’t need to change their data to support a new community agreed information model.

• The actual request and query are made using an OGC WFS request with an embedded OGC Filter
  • The filter contains our information model query
• **Step 5 – Handling the response**
  • Our response is an XML document representation of the information model
    • XML: Easily consumed and transformed.
    • XML preserves the complex relationships within an information model
      • Contrast this to CSV which simply flattens any relationships into a single table.
    • XML can be used for visualising, scientific simulations etc.
  • The content of the response can be interrogated for more information by SISSVoc.
    • Our response may contain ‘urn:cgi:classifier:GA:commodity:Pgvl’ what does it mean?
      • According to our Vocab – ‘Pisolitic gravel’
AuScope and ERML

• We just covered a Geologic workflow
  • The concepts are general enough to abstract to all spatial data.
• A registry can help you find data services for a particular information model.
• A defined vocabulary for the information model removes ambiguity.
• You query against the information model whilst being completely ignorant of the storage model.
  • 1 query can be re-used for each and every dataservice
• You get data automatically transformed to the spatial reference system you requested.
Into the examples!

• The exercises will see us manually performing the steps of an automated workflow.
  • We don’t have time to wait for people to follow along but all these examples are packaged with the ‘SISS in a box’ VM.
  • There are also a number of exercises we don’t have time to demonstrate.
  • All exercises are detailed at https://www.seegrid.csiro.au/wiki/Siss/SissConsumptionWorkshop

• You should become familiar with the general *idea* of the queries, don’t bother memorising the syntax.
  • You can always redo these examples for future reference.

• We will be using some live production datasets as well as the ‘slake’ dataset.
  • The data will be coming from a few scientific disciplines.
Surface Lake (slake) dataset

- The slake dataset is based on the Australian Bureau of Meteorology Water Storage dataset
  
  - For the list of supplying organisations and their attribution, see:
    

- A set of spatially located surface reservoirs with associated observations and measurements.

- A reservoir (etymology from French réservoir a "storehouse) or an artificial lake is used to store water.
  
Querying an Information Model

• Define our objectives for the Slake dataset
• I need to analyse metadata for surface reservoirs in Victoria.
  • In particular I’m interested in surface reservoirs that were completed after 1970
• The abstract question – I need data features that lie within a spatial area and that also pass a basic filter.
Querying an Information Model

• Exercise #1 – WFS features on a map
  • “I need to analyse metadata for surface reservoirs in Victoria.”
  • What is the easiest way to get WFS features?
  • Self describing service – GetCapabilities
    • What features can be queried in this WFS?
  • GetFeature – Get all features that pass an optional filter.
  • Response can be transformed and the spatial components easily visualised on a map.
  • The result is most of the data does not meet our requirements – there are many reservoirs created before 1970.
Querying an Information Model

• **Exercise #1 – WFS features on a map**
  • Hopefully you are now familiar with;
    • The steps that are normally taken when first accessing a web feature service
    • A basic WFS GetCapabilities request and response
    • A basic WFS GetFeature request and response
Querying an Information Model

• **Exercise #2 – Filtered WFS features on a map**
  • “I need to analyse metadata for surface reservoirs in Victoria.”
  • Let’s only request surface reservoirs that lie within our area of interest.
    • We’ll have to find our relevant spatial component in the information model.
    • We’ll also have to define our area of interest
  • Now let’s only request surface reservoirs that were created after 1970
    • We’ll have to find our relevant property in the information model
Querying an Information Model

• Exercise #2 – Filtered WFS features on a map
• Building from our previous example you should now be familiar with how an OGC Filter queries an information model
  • More specifically, how it queries an XML representation of the information model.
• You should have also seen how an OGC filter can be integrated into a WFS query.
Querying an Information Model

• **What was covered**
  - A WFS response IS data.
  - Querying a WFS involves
    - Identifying what you want from the information model
    - Generating a OGC Filter / WFS GetFeature request based on the model
    - Refining queries until the exact data you want is returned.
Datasets with a standardised vocabulary

• Why Vocabularies?
  • standardised relationships != standardised terms.
    • We agreed that mines should have a commodity, not what a valid commodity name is.
    • We agreed that each borehole should record their maximum depth but didn’t specify the units of measure.
      • Or we specified the units as ‘m’ (metres, miles or something else?)
      • Overused Example - Ambiguous units of measure doomed the Mars Climate Orbiter.
  
• An information model should standardise it’s content through the use of a defined vocabulary.
  • That way all services speak the same language AND use the same terminology.

• SISSVoc – The SISS Vocabulary service.
Datasets with a standardised vocabulary

• **Our tasks**
  
  • Now switching back over to ERML – “I wish to locate all mines whose commodity is ‘Gold’”.
    
    • Remember each and every geological survey uses their own in house definitions for ‘Gold’.
    
    • Luckily for us they are supporting ERML which uses the GA defined commodity vocabulary.
  
  • Then after making the request – “I’ve got a mine that includes gold but also mentions ‘urn:cgi:classifier:GA:commodity:As’. How do I find out what it means?”
Datasets with a standardised vocabulary

- **Exercise #4 - Using SISSVoc to make an OGC filter**
  - We need to get the vocabulary defined URN for our commodity ‘Gold’ from the ‘commodity_vocab’ repository
    - The ERML information model defines the usage of the Geoscience Australia vocabulary – therefore our URN’s must come that vocab.
  - As in previous exercises– Identify our properties to query in the model
  - Combine the OGC Filter with the vocabulary URN
Datasets with a standardised vocabulary

• **Exercise #4 - Using SISSVoc to make an OGC filter**
  • Hopefully now you are aware of some of the queries that can be used with SISSVoc
  • You should be able to see how to use a vocabulary service in conjunction with a WFS in order to generate *specific* queries.
    • ‘specific’ is underlined because querying for ‘Gold’ is ambiguous, querying for ‘urn:cgi:classifier:GA:commodity:Au’ is not.
Datasets with a standardised vocabulary

- Making a query is often only half the workflow.
- A WFS response will often contain more vocabulary concepts
  - Units of measure
  - Other related terms
- Being able to interrogate a concept to find its meaning is crucial for incorporating WFS into a scientific workflow.
Datasets with a standardised vocabulary

• **Using SISSVoc to interpret an information model**
  • We now have a commodity response that references ‘urn:cgi:classifier:GA:commodity:As’. What does it mean?
  • We need to identify an appropriate vocabulary service and then query it for more information.
    • We know we are working with ERML which defines a particular vocabulary which makes the first part of the problem easy.
  • Being time limited all I can say is that SISSVoc allows you to query this concept for more information.
    • You can try this in your own time with exercise #5
Arsenopyrite, also unofficially called mispickel, (FeAsS) is the most common arsenic-bearing mineral. In the lithosphere the minerals of the formula M(II)AsS, with M(II) being mostly Fe, Ni and Co, are the dominant arsenic minerals. Realgar Orpiment and realgar were formerly used as painting pigments, though they have fallen out of use due to their toxicity and reactivity. Although arsenic is sometimes found native in nature, its main economic source is the mineral arsenopyrite mentioned above; it is also found in arsenides of metals such as silver, cobalt (cobaltite: CoAsS and skutterudite: CoAs3) and nickel, as sulfides, and when oxidised as arsenate minerals such as mimetite, Pb5(AsO4)3Cl and erythrite, Co3(AsO4)2. 8H2O, and more rarely arsenites ('arsenite' = arsenate(III), AsO33− as opposed to arsenate (V), AsO43−). In addition to the inorganic forms mentioned above, arsenic also occurs in various organic forms in the environment.
Service Discovery

• STOP! We’ve been making assumptions this entire time

• Where are all these data service URL’s coming from?
  • The registry of course, but you already know that.

• So how does a client query the registry?

• With the OGC ‘Catalogue Service for the Web’ (CS/W) specification
Catalogue Service for the Web (CS/W)

• Provides an XML query interface similar to WFS
  • We query for records, not features
  • The response is a set of records encoded to an information model
    • ISO 19115/9
  • Has limited support for the same OGC filters we use with WFS.

• But it also has a number of distinctions too:
  • Pagination
  • Support for inserting/updating records

• Due to time constraints we can’t go into the specifics of a CS/W request
  • But feel free to have a look at exercise #9 for the details.
Service Discovery Architecture

Client/User

Registry (Geonetwork)

Storage

CS/W Request

Metadata Records
So what have I missed?

- **Dataset Visualisation**
  - Rendering a dataservice directly is time consuming
  - WMS is the main way we visualise data and has an exceptionally low barrier of entry
    - See exercises #6 and #7

- **Web Coverage Services**
  - Not all spatial data is easily represented by WFS
  - Datasets containing continuous spatial/temporal over a region a much better represented in a WCS
  - Consumption is trickier than a WFS but it is a completely self describing service.
  - Supports subsetting, data format conversions and spatial reprojections.
  - See exercise #8
For all the handsome software developers

- AuScope Portal source code is available at...
  - https://svn.auscope.org/subversion/AuScopePortal/AuScope-Portal/trunk

- All workshop example source code is at...

- For everyone else – the workshop exercises use configurable URL’s so feel free to use the exercises to demonstrate your own services.
Acknowledgements

SISS Project Sponsors:

AuScope, ANDS and the ARCS Organisations are supported by the Australian Government through the National Collaborative Research Infrastructure Strategy Program and/or the Education Investment Fund (EIF) Super Science Initiative

Australian Government
Department of Innovation
Industry, Science and Research

Workshop Sponsors