

## ***Spatial Information Services Stack***

**Overview** Summarise the context and objectives of this project.

The SISS project will:

- develop some of the component services and functional capabilities needed to realise a spatial information data commons within Australia supporting linkages with academia, research, public and private sector data and service providers and users
- leverage the Open Geospatial Consortium (OGC) and ISO standards that have been adopted by a number of NCRIS capability areas and institutions in Australia and worldwide for which various reference implementations already exist for information services and portal interfaces
- combine and build on existing expertise in this foundation area in order to assist multiple NCRIS capability areas and institutions that have interest in geospatial web service middleware and clients.
- establish capabilities to enable standards based networks developed by other communities such as the Australian Bureau of Statistics National Data Network to interact with and benefit from the spatial capabilities developed through this proposal

In doing so, the spatial data infrastructures being developed by individual NCRIS capability areas (and other organisations) will be able to interoperate and thus share resources and establish a spatial information data commons that will also be accessible by other communities of interest.

Three principle outcomes will be achieved.

1. SISS will develop and initiate some of the core services needed for a spatial data commons to be operated beyond the lifetime of the project by identified project participants. This will include complete applications or services comprising client interface (e.g. web portal or web service interface), middleware and data repository components. The identified participants will have an enduring custodial role to maintain the service and associated repository for the community.

Core services identified include:

- OGC Catalog Service – providing a service interface to register meta-data. At a minimum this will contain a register of registries that are available. The registers also play a role in governance of standards associated with the spatial data commons (e.g. publication of interoperability profiles, controlled vocabularies, data standards)
  - Discovery portal – a discovery portal providing access to federated registry/catalog services (including but not limited to the one in the previous dot point)
2. SISS will develop common software components that can be deployed with spatial data holdings to make the holdings accessible within the spatial data commons. The expertise needed to maintain and deploy those components will also be developed in a sustainable form within identified project participants. A number of NCRIS capability areas (e.g. Auscope, IMOS) and Australian jurisdictional and domain based SDI projects (e.g. Western Australian Shared Land Information Platform (SLIP), CSIRO Water Resources Observation Network, Marine Portal, Queensland Government Enterprise architecture projects, GeosciML Testbed, CSIRO SEE Grid) share a common service and software component base, usually based around OGC service standards and open source reference implementations of middleware to support these.

Production quality deployment of such services is needed, and SISS can provide a quality assurance and support service by performing development that makes the reference implementations robust, packaged, versioned and testable against interoperability service profiles. Service implementations will be openly available to all organisations.

3. SISS will build a skill base able to work with holders of spatial data to deploy and operate relevant data servers and OGC compliant services and a consequent increase in the number and variety of spatial data sets made available through common access mechanisms.

Whilst the project has a significant emphasis on a baseline of OGC standards it is anticipated that a number of other non-OGC components are likely to be required to fulfil the needs of the NCRIS community in particular niches (e.g. OPeNDAP). It is anticipated these will be more properly identified during the detailed planning phase.

**Users** *Identify the research communities and resource providers that this project serves; and the potential number of users. This should include some NCRIS capabilities or other data federating or collaborating research groups, and any institutions that will participate through setting requirements for or steering this project.*

There is a substantial and diverse user community for these services and the successful deployment in Australia will facilitate both national and international collaborations. The specific users groups to be services during the project period are described below characterised as either:

- Spatial information provider
- Spatial information consumer
- Service builder and/or operator

Notably, many participants are all three, however the project focuses on delivering against these three different needs.

### **Spatial information providers**

AuScope will also use SISS outputs to provide an OGC Catalogue Service for:

- Registry of available services, standards, interoperability profiles, reference data sets

AuScope will use SISS outputs to provide OGC WFS, WMS and WCS information services for:

- National Virtual Core Library – observational (Hyperspectral and visible spectrum imagery) and derived data types for drill core
- Global Positioning System (GPS) station data
- Interpreted and observed geology (GeoSciML)
- Models and simulation results
- Geophysics observations (imagery and point observations)
- National topographic and related mapping data (National Mapping and Information Group of Geoscience Australia)

The capacity to integrate relevant data from universities and government agencies underpins many projects, SISS outputs will be adopted as follows:

- Geoscience Australia (National Mapping and Information Group) is coordinating the ANZLIC driven National Elevation Data Framework and data within this framework will be held by the custodian but shared through web services.
- The Geological Survey of Victoria is a member of the international GeoSciML working group and will publish their geological maps, rock property and other geoscience holdings via web services
- The CSIRO Minerals Down Under flagship has data holdings in airborne Hyperspectral imagery and thermodynamic data to be published via web services

## Spatial information consumers

AuScope will adopt SISS outputs to support data integration across several of its activities:

- AuScope Simulation (the CSIRO Minerals Down Under Discovery Theme, the School of Geosciences at Monash University and the Earth Systems Science Computational Centre at the University of Queensland) will develop services that depend on the integration of geochemical, rock property and thermodynamic data
- Modelling and simulation of hazards in real time requires access to data in distributed sources (Geoscience Australia and ANU's Tsunami modelling needs to integrate topographic, DEM, tidal information, earthquake. Realtime simulation of bushfire behaviour needs to integrate data from Bureau of Meteorology and local government data)
- Geophysical inversion (CSIRO Exploration and Mining) requires access to rock property, gravity data etc in government agencies
- Assessment of the Geothermal potential of Australia by a variety of agencies requires access to rock property and geochemical data in the Government geological surveys
- The Australian Resources Industry (and the associated research agencies) have requested the improvement of "the accessibility and consistency of pre-competitive geosciences data through nation-wide protocols, standards and systems - harnessing into a single system all exploration related data including that generated by the private sector and all governments" – Minerals Council of Australia<sup>1</sup>

## Service builder and/or operators

- Building, managing and delivering 3D geological models in the geological surveys requires coordinated, and at times secure, access via online catalogues and registries to geological map data, rock property and other geophysical data from distributed data sources in both the private and public sector. (Geological Survey of Victoria)
- The Australian Bureau of Statistics wants to integrate its mesh blocks with topographic data in government agencies and also add spatial capabilities to their National Data Network
- Researchers in exploration geoscience (e.g. CSIRO Minerals Down Under Discovery Theme, Geological Surveys, Industry and Academic Geoscience researchers, Predictive Mineral Discover CRC) using the Auscope Virtual Core Library will integrate its outputs with distributed data sources (e.g. geological maps, tenement boundaries) from other government and industry repositories
- The Magnetotelluric component of AuScope Simulation (Adelaide University, ANSIR, Research School of Earth Science ANU, Geoscience Australia) aims to have a distributed network of data sources with a series of authoritative reduction programs being available to researchers online

**Needs** *Describe the needs of the research communities or resource providers that this project seeks to address.*

The fundamental motivation arises because the complexity of data integration is rapidly increasing (more data sources and more combinations of interest) and traditional data integration methods have become untenable as the effort required exceeds the available human resources. Additional motivation comes from the greater complexity of national and international issues being addressed and the need to interact with multiple communities of interest through access to data, information and services based on open standards for interoperability.

The solution is to invest in an information exchange layer above data sources and to work with the providers and consumers of data to adopt that layer's access and exchange mechanisms.

<sup>1</sup> [http://www.minerals.org.au/\\_data/assets/pdf\\_file/0017/12266/mr006\\_25\\_energy.pdf](http://www.minerals.org.au/_data/assets/pdf_file/0017/12266/mr006_25_energy.pdf)

Previous developments in Australia and internationally have provided a set of technology selections and subsequent development to meet this need. In many cases most of the technical requirements are already met and only perfective developments, hardening of components, packaging and support are required for deployment to occur. The following technologies have already been identified:

- Geoserver – WMS and WFS
- GeoNetworks – registry/catalog service
- THREDDS, Hyrax - WCS, OPeNDAP
- Portals and Desktop clients – various, we should support a couple

These technologies need to be configured, extended, deployed and supported in ways that allow researchers (especially those not associated with a spatial data custodial organizations) to:

- Publish the spatial datasets they hold and control access to them using the Australian Access Federation infrastructure where desired.
- Publish metadata about the information services they offer that may be of interest to Australian researchers (and other communities) on the Australian Grid.
- Discover and gain access to spatial data online using OGC standard service interfaces
- Find information about the data standards used by the registered information service
- Participate in international research programs by sharing data across international borders due to compliance with international standards
- Locate spatial data “close” to computational capability

**Services** *Describe the result of the project in terms of the service(s) that will be implemented and demonstrated by the project and which could be operated in an ongoing fashion; and the proposed operator of each service.*

The services generated by the SISS project will consist of a large number of data publishing services, a smaller number of exemplar information services consuming and providing spatial information and a limited set of support services. Wider deployment of compatible services would be expected following the project’s completion when the SISS services transition to providing a backbone of an Australian spatial data commons.

### **Data services**

AuScope will provide SISS based access to GPS network station and observation logs, Seismic and Hyperspectral transects, Geological models, maps, Virtual core library. The members of AuScope holding the relevant data will sustain these services into the future, drawing on the Support Services.

ARCS will host (at media cost) SISS compatible spatial data contributed by researchers not associated with custodial organizations – this would include researchers within NCRIS communities including AuScope, IMOS, Atlas of Living Australia, and others.

### **Information Services**

AuScope will provide and maintain a spatial information discovery portal, a registry with multiple registers to support both discovery and governance, and repositories.

### **Support Services**

The support services are limited in number as they intentionally aggregate generic activities required at the heart of any spatial data commons. The intention is that these services will be developed by resources across the project members but the operational aspects will become embedded within the service component of ARCS.

They will include:

Spatial Information Services Stack  
National eResearch Architecture TaskForce

- Packaging of the services stack to ease deployment and ensure consistency of service deployments
- Maintenance and support for the spatial services implementations in line with standards development and user requirements. The spatial services supported will provide the following service interfaces: OGC Web Feature Service, OGC Web Map Service, OGC Catalog Service, OGC Coverage Service and OPeNDAP standards (the detailed requirements specification can be found in the OGC and OPeNDAP documentation). The initial implementations are based on GeoServer, GeoNetworks and THREDDS/Hyrax.
- Reference implementations of client interfaces – these clients should be robust enough for general use and also serve as exemplars for developers wanting to create custom interfaces

While spatial data custodians will deploy SISS outputs themselves, they will require support to ensure the SISS outputs are robust in production environments and evolves with international standards.

Some data sets of national interest may require hosting at ARCS locations to allow data sets to be shared across the AREN and/or placed in close proximity to computing resources.

## **NeAT Characteristics**

**eResearch effect** *What changes in behaviour and activity are expected from the project that will demonstrate the broader adoption of eResearch practice?*

The specific behavioural changes relate to:

- the frequency with which research projects make access to other people's data as an efficient and effective activity within the project,
- the extent to which remote access to that data is perceived to be satisfactory,
- the volume and variety of data that is accessible through standardised mechanisms,
- the extent to which publication of datasets and services is seen as a normal output of research projects

Many research activities require access to data held by government agencies and this stack will facilitate improved access to spatial data of interest.

Thus the SISS should reduce the per unit cost of data publishing and access, particularly for research requiring data from a diversity of domains, but lead to a significant increase in the 'data processing' budget within spatial related research fields.

Research addressing larger problems (systems, cross-disciplinary research) with more meaningful questions and answers should be possible.

**Broader adoption** *Which additional communities, resource providers or organisations would also be expected to benefit from the provision of the same or similar services should the project succeed?*

As a sanity rule: consumers and service builders are only acceptable project members if the data providers they need are also project members. Consequently only a small number of interested parties are 'in' the project. In addition, the broader adoption of the spatial services stack is dependent on the existence of the associated software components as part of a supported service framework including maintenance and support, as described earlier in the Services section (the unsupported availability of the software components alone is not sufficient).

Consequently a large number of entities to which SISS is relevant are external to the project. These entities are expected to adopt SISS outputs over time, and indeed may do so simultaneously with the project's progress.

- The marine biodiversity program will be able to more easily integrate zoological data with marine sediment data
- Environmental Impact studies, associated with development proposals, will make use of spatial data services
- Many aspects of human geography - most of the studies supported by ABS data would be enhanced if it was part of a data infrastructure/commons, rather than ad-hoc downloads. Development, service provision, socio-economic studies.
- Water resources - requires current inputs from multiple agencies and disciplines. CSIRO WRON project. AWRIS.
- Most projects in eWater space - see eWater-CRC for details see <http://www.ewatercrc.com.au/research/>

- The Australian Government's Social Inclusion Agenda, the establishment and delivery of which can be enhanced through the integration of some of the spatial capabilities that will result from this project.

### **Value adding**

*Identify the components of the project that could be based around generic technologies or be implemented through shared services for which the project would provide an exemplar use case or requirement set.*

The chosen service implementations are designed around generic components which provide a configurable mapping from private data schemas to community agreed standards.

Auscope researchers are actively collaborating with international activities in this area and these activities can also contribute to setting requirements and guiding development:

- CGI GeoSciML Testbed #3 ( Simon Cox, Bruce Simons, Tim Duffy )
- One Geology ( Tim Duffy )
- WRON, AWRIS data services (BRS, BoM, CSIRO L&W)

### **Standardisation**

*Describe the global technology development or standardisation work that will be adopted, adapted or extended within the project and any risk reduction available by collaboration with similar activities occurring elsewhere in the world.*

The OGC and ISO standards are the international standards for spatial data services and information models.

The selected technologies are the reference implementations or are fully compliant with those standards.

The participating communities are, in most cases, directly involved in the development of the international standards involved.

Collaborative links with international activities in this area are already in place via the NCRIS capability areas identified.

The Office of Spatial Data Management also plays a coordination role in the Australian context and has offered to assist this NEAT activity (see Governance).

Profiling (subsets, restrictions, combinations) of standards to serve detailed community requirements is an under-developed area of standardisation. The OGC Policy Working Group is working to specify a meta-model for this activity. Auscope members Simon Cox and Rob Atkinson are participating in this work.

## **Project Scoping**

**Key Participants** *Name any Pfc components, any NCRIS capabilities, or any other institutions or groups that will need to be involved in the project planning and execution.*

The success of the SISS project is of interest to:

- Office of Spatial Data Management (Ben Searle)
- ANZLIC (Ian Batley, CEO of the ANZLIC National Office) (tbc)
- Geoscience Australia (tba)
- State geological surveys (Bruce Simons (Vic), Stephen Bandy (WA) )
- Auscope ( Robert Woodcock )
- CSIRO, Water Resources Observation Network ( David Lemon ), Minerals Down Under Flagship ( David Gray )
- ARCS (Tony Williams), iVEC (Andrew Rohl) and TPAC (Nathan Bindoff)
- Australian Bureau of Statistics (tba)

**Governance** *Outline the arrangements proposed to manage the contributions and user interaction. Eg: a project managed by ARCS, ANDS, or under another NCRIS capability governance, or by subcontract to a named lead agency, or a new J/V between the parties.*

The SISS needs to be aligned with federal government initiatives including the ANZLIC Australian Spatial Data Directory (ASDD) project hosted by Geosciences Australia (GA). The ASDD is a national initiative supported by all levels of government under the auspices of ANZLIC, which aims to improve access to Australian spatial data for industry, government, education and the general community. The AuScope and IMOS NCRIS capability areas are well networked with these communities and can ensure this alignment occurs.

The project will be managed in three parts to meet the structural requirements of ARCS and Auscope:

Data Services	A support and engagement activity with data holders
Information Services	A co-development effort to build new information services
Support Services	SISS development and hardening and user support systems

The project will operate under the management of a single project manager. With the overall governance of a steering committee, proposed as follows:

- CEO Auscope - Scott McTaggart (agreed)
- Executive Director ARCS – Tony Williams (tbc)
- General Manager, Office of Spatial Data Management – Bean Searle (agreed)
- Geological Survey of Victoria – Alan Willocks (tbc, supportive)
- Geoscience Australia Rep – Chief of an appropriate division in GA (tba)
- CSIRO Land and Water – Research Group Leader - David Lemon (agreed) may be changed to CSIRO L&W Chief – Neil McKenzie

## **Project Scale**

*Identify the overall scale expected in the project, eg. 1 to 3 years, total effort in any year, and nominate any parties that have indicated a willingness to participate through providing resources. (funded or in-kind, people or facilities).*

Because the project could be made arbitrarily large by including more participants, the full 3 years is proposed. However the organisational and relationship complexity demands full commitment from any organization wishing to participate.

Overall scale is estimated as:

Project Member	Year 1	Year 2	Year 3	Total	Funded	Contributed	NeAT
Data Services	2	3	4	9	3	6	33%
Information Services	3	3	4	10	3	7	30%
Support Services	4	4	2	10	6	4	60%

Allowing cash cost per EFT of \$125,000, the estimated NeAT cost is of the order of \$1.5M.

The NeAT support assumes the bulk of the effort in Data Services will come from the participants seeking to make their data accessible through the SISS. That effort is assumed to represent full time people committed over short periods so that the table is an average estimate. The NeAT funding is assumed to provide a full time liaison and assistance service for that purpose.

## **Major Steps**

*Identify the key steps that will be visible to users as the services develop. Note that some observable deliverable is needed every half year and projects may be reviewed based on the achievement of these steps.*

It is anticipated the following services

### **July 2008-Dec 2008**

- Project plan accepted and included in ARCS and AuScope business plans
- ARCS Hosting OGC Catalog Services and Discovery Portal
- Auscope hosting OGC Catalog Service and Discovery Portal
- Auscope GPS data WFS service deployed
- Geological Survey of Victoria GeoSciML testbed collaboration deployed

### **Jan 2009- Dec 2009**

- ARCS SISS support service fully established
- CSIRO Minerals Down Under Flagship deploys laterite geochemical data service (Western Yilgarn)
- Auscope NVCL WFS service deployed
- CSIRO Minerals Down Under Flagship deploys airborne hyperspectral data service
- CSIRO Minerals Down Under Flagship deploys thermodynamic data service
- CSIRO Minerals Down Under Flagship deploys Northern Yilgarn hydro-geochemistry

### **Jan 2010 – Dec 2010**

- Auscope Virtual Rock Laboratory and Tsunami workflows utilise registry and information service infrastructure for service discovery and data management support
- Auscope deploys WCS for geophysics imagery with large data set support

**Jan 2011-June 2011**

- Auscope Earth Model and portal service infrastructure fully established and using the service stack
- Broader adoption well underway

**Dependencies**

*Identify dependencies that exist to activities or developments external to the project.*

The full gains from the SISS will arise when it is integrated with other services simultaneously under development.

Within ARCS, these include:

- Authorisation solutions making use of the Australian Access Federation
- Large data transport and interoperation with Grid facilities (e.g. data staging from the spatial services to an HPC facility)

Within AuScope these include:

- GeosciML Testbed 3 – GeoNetworks and GeoServer related developments
- One Geology – Geoserver will be used for the WFS+GeosciML cookbook
- AuScope Grid
- AuScope NVCL

The SISS project is known to provide essential inputs to IMOS and the Atlas of Living Australia within NCRIS.