OPENGIS WEB SERVICES INITIATIVE 2

REQUEST FOR TECHNOLOGY SUBMISSION

Submitters:

Council for the Central Laboratory of the Research Councils (CCLRC)

British Atmospheric Data Centre (BADC)

NERC Earth Observation Data Centre (NEODC)

Contact:

Dr Andrew Woolf

Rutherford Appleton Laboratory

Chilton Didcot

Oxfordshire OX11 0QX

UK

(A.Woolf@rl.ac.uk)

OVERVIEW

This RFT submission promotes two objectives:

1. The application of OGC technology to the broader earth sciences; and
2. The application of Grid computing to the OGC Technical Baseline.

We believe these objectives carry appeal to a significant scientific user community including earth observation and the climate sciences.

OGC services formalise and standardise functionality that has broad application to the earth sciences. For example, we mention two alternative systems that are becoming widely deployed in the meteorological and oceanographic communities. Live Access Server [i] (LAS) is web server software that enables interactive visualisation of geo-referenced data through a conventional web browser. A considerable amount of oceanographic and meteorological data is available in this manner[1]. LAS may be used either interactively, or in a "batch mode" using HTTP GET or POST with request parameters. There is a clear correspondence between LAS functionality and the OGC Web Map Service. Similarly, a system has been under development since 1995 [ii] for access to raw earth science data using a HTTP GET mechanism. Originally known as the Distributed Ocean Data System (DODS [iii]), it has recently been renamed the Open source Project for a Network Data Access Protocol (OPeNDAP [iv]). There is a clear overlap with OGC Web Coverage Service functionality. A web service alternative for climate data access has recently been proposed [v]. LAS and OPeNDAP will form the technical basis for an upcoming major international ocean forecasting intercomparison exercise, the Global Ocean Data Assimilation Experiment[2]. There would be considerable benefit to seeing standards-based OGC technology expand into the broader earth sciences. This could be facilitated through explicit engagement with the community, perhaps through an Interoperability Program testbed experiment. We are collaborating with the LAS developers to implement a WMS interface.

OGC technology currently is based on a stateless web service model. The Implementation Specifications are closer to the Representational State Transfer (REST) model [vi] than to the Simple Object Access Protocol (SOAP) [vii]. There would be considerable benefits to incorporating statefulness in OGC services. Two examples are handling latency in the case of complex requests, and notification mechanisms in the event of state changes. The addition of state to Web Services is a major benefit of Grid computing [viii, ix]. We believe OGC technology has more in common with the distributed scientific computing problems to which Grid computing has been applied, than with stateless e-business web services.

SUMMARY OF THREADS

Our submission is based around an oil spill scenario outlined below and developed in the next section. This scenario encompasses elements of the following themes:

- Common Architecture
- Image Handling

[1] A representative list of LAS server sites is maintained at http://ferret.wrc.noaa.gov/Ferret/LAS/LAS_servers.html.

[2] See http://www.mersea.eu.org/html/information/overview/MERSEA_project.html#godae_dsp for details.

- Modelling and Simulation

As well, the inbuilt notification mechanisms of Grid Services provide a standardisation of the Web Notification Service functionality considered in the Sensor Web Enablement theme.

SUMMARY OF APPROACH

We consider a scenario germane to the toxic dispersion problem posed by the Modelling and Simulation theme. Figures 1 and 2 outline use cases and a sequence diagram for the scenario. An oil spill incident has occurred in the North Sea. A Disaster Recovery Team wishes to integrate synthetic aperture radar imagery, meteorological and ocean forecast data and spill simulation capabilities to assist in mitigation planning. Each of these elements is accessed in a loosely-coupled, distributed information infrastructure that illustrates the workflow capabilities of Web and Grid Service technology. The steps involved are as follows:

1. One or more Registry services is used to locate sources of remote-sensed SAR imagery, ocean and meteorological forecast data (Grid-enabled Web Coverage Services), and an oil-spill simulation service.

2. An instance of the spill simulation Grid service is created. Through a Control interface, it is provided with the location of the met/ocean forecast data services. Grid security mechanisms are employed for mutual authentication, and delegated proxy credentials are used by the service to obtain forecast data on behalf of the user.

3. The simulation is executed through the Control interface, and the Grid notification mechanism is used by the service to indicate when the simulation is complete.

4. Finally, the user invokes a Grid-enabled Coverage Portrayal Service to retrieve the simulation results (through a Model Output interface) and produce a visualisation. The CPS is also used to render for comparison recent SAR imagery retrieved from an Image Archive Service.

Discovery and processing of remote-sensed imagery in large distributed archives (especially using Grid technology) is considered in the Image Handling theme, while access to meteorological forecast data and simulation services are aims of the Modelling and Simulation theme. The overall approach, based on Grid computing pertains to the Common Architecture theme.

BENEFITS

There is a growing penetration of Web Services in the e-business sphere. The technology facilitates interoperability and just-in-time integration of business processes. Initial applications have been to relatively light stateless computing services. While Web Services encapsulate behaviour, Grid Services encapsulate both state and behaviour, thereby enabling a much richer distributed computing model. Grid computing has seen initial application to scientific computing problems characterised by very large and complex datasets and heavy resource demands. It is our view that these are more characteristic of OGC Web Service applications than the business oriented Web Services developed to date.

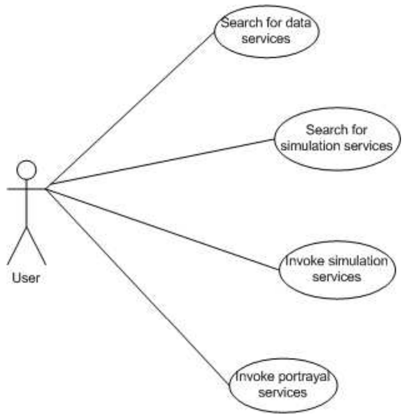

Figure 1: Use cases for oil spill scenario

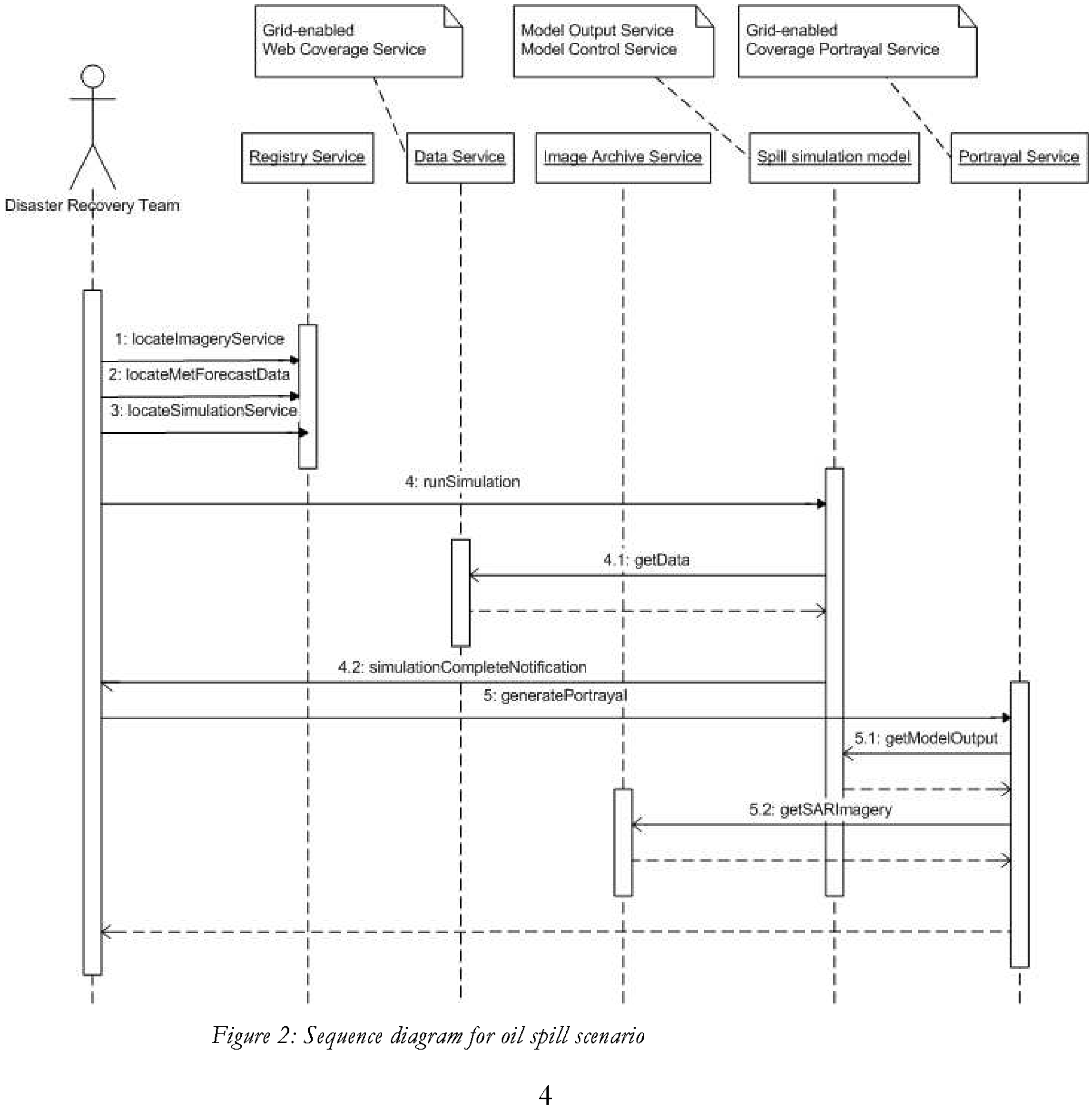

Figure 2: Sequence diagram for oil spill scenario

THE CONSORTIUM

The role of the British Atmospheric Data Centre (BADC, [x]) is to assist UK atmospheric researchers to locate, access and interpret atmospheric data and to ensure the long-term integrity of atmospheric data produced by Natural Environment Research Council (NERC) projects. It is hosting a project (the NERC DataGrid, [xi]), funded under the UK's e-Science program, to develop a Grid infrastructure for discovery and use of a wide range of environmental data sets.

The Council for the Central Laboratory of the Research Councils (CCLRC, [xii]) provides central facilities and technical expertise in support of basic, strategic and applied research programmes across the UK. It is playing a leading role in the UK e-Science program through the CCLRC e-Science Centre [xiii]. As well as hosting the UK national Grid Support Centre, its forty staff work on a range of UK and international Grid projects. The UK and Ireland regional office of the World Wide Web Consortium (W3C) is based at CCLRC.

The NERC Earth Observation Data Centre (NEODC, [xiv]) is, like the BADC, one of NERC's "Designated Data Centres", but with responsibility for airborne- and satellite-derived data (mainly imagery) pertaining to the surface of the Earth. It provides the environmental science community with access to, and support in the exploitation of, data in this archive. It is also planning to develop a suite of GRID-based services to complement these datasets, with the aim of improving their accessibility and ease of use.

ELABORATION

OVERVIEW

The scenario outlined above integrates elements of three OWS2 themes: Common Architecture, Image Handling, and Modelling and Simulation. This scenario is used below as a specific context within which the themes are addressed.

COMMON ARCHITECTURE PAT

PURPOSE AND SCOPE

An important prerequisite for success of the Spatial Web paradigm is Quality of Service (QoS). Availability, fail-over mechanisms, resource provisioning, dynamic discovery are all dimensions of QoS. Each of these may be leveraged from the inclusion of state (including transience and instantiation) in the architectural framework. This inclusion of state is the sole constraint that differs in our approach from those identified for OWS solutions, and is a key functionality added by Grid Services to Web Services.

Grid computing is proving a robust technology for dynamic discovery and access to large distributed data archives. A few examples of Grids being developed for the earth sciences include:

- The NERC DataGrid (http://ndg.badc.rl.ac.uk)
- The Earth System Grid (http://www.earthsystemgrid.org)
- The EU-DataGrid (http://eu-datagrid.web.cern.ch/eu-datagrid/) and WP9: Earth Observation (http://styx.esrin.esa.it/grid/)
- EuroGrid and WP2: Meteo Grid (http://www.eurogrid.org/wp2.html)

Markets for Grid-enabled implementations of OGC technology include a growing number of high-profile international Grid projects. The horizontal technology base would find application across large segments of the earth-sciences community.

Problems and solution summary

We consider the following problems identified by the OWS2 sponsors:

> 1. *Web-based clients and services don't interoperate without specialized code, are not reliable and are not scalable.*
>
> 4. *Not possible to flexibly and dynamically combine (aggregate, compose and/or chain) services to perform new/novel functions.*

Grid computing has evolved "... as a solution to new challenges relating to the construction of reliable, scalable, and secure distributed systems" [viii]. The difficulties of composing complex workflows reliably in a large cross-domain infrastructure are just those problems which Grid computing aims to address. The

Grid Service model incorporates mechanisms for creating and discovering transient service instances. Mechanisms for lifetime management, change management and notification all support enhanced QoS.

OBJECTIVES

Broad objectives identified in the RFT include:

> - *Reference architecture descriptions, proof-of-concept prototypes, middleware components, and 'use scenarios' demonstrating end user benefits;*
> - *Data and Service representation methods, languages and metadata techniques enabling Earth science community-centric services and open tools sets; and*
> - *Technologies managing data integrity, authentication, and heritage, among others.*

The NERC DataGrid project, amongst others, is directly concerned with these objectives. Data and metadata models are being developed that apply across a range of datasets. Initially concentrating on meteorological and oceanographic data, this will extend to earth observation, hydrology, land-use, and others as the project develops. Development of these models draws heavily on the emerging ISO TC211 series of standards, in particular ISO 19111 (Spatial referencing by coordinates), ISO 19115 (Metadata), ISO 19119 (Services) and ISO 19123 (Coverages). A considerable amount of work remains to be done, but profiles of metadata standards for specific communities are being developed.

Authentication is a fundamental requirement for Grid computing. Current implementations rely on public key infrastructure and x.509 certificates. These are being standardised through the Internet Engineering Task Force [xv]. Grid resource owners establish networks of trusted Certification Authorities and only accept appropriately certified credentials. Individual Certification Authorities publish a Certificate Policy and Certification Practices Statement in accordance with RFC2527 upon which trust may be based. Procedures typically include identity verification and procedures for managing credential revocation in the event of compromise. In practice a user's electronic certificate is managed by Grid client software and a pass-phrase is used to generate proxy certificates with limited lifetime. The issue of authorisation is distinct from authentication, and is currently the subject of considerable activity. An extensive set of proposals is being developed for web services security [xvi].

Specific objectives pertinent to a Grid computing approach include:

> *1. OWS services describe themselves (their type, their operations, their content, etc) in a uniform way.*
>
> *2. To define an extensible, evolvable framework supporting interoperability standards to create interdisciplinary applications and/or custom data processing systems.*
>
> *3. Define technologies that enable interoperability between data production, storage, archive, and analysis systems*
>
> *6. Enhanced positioning in the IT infrastructure community*

These are discussed in the Architectural Concepts section below.

REQUIREMENTS

The full range of Common Architecture requirements are supported by a Grid computing approach:

> 1) *Enable self-describing data and service representation (e.g., Metadata definition and standards, Markup languages)*
>
> 2) *Enable access and transformation of data and services (e.g., adaptive interfaces, agents)*
>
> 3) *Employ metadata for use as basis of search/discovery of resources*
>
> 4) *Employ shared schemata, interoperability software tools and metadata for data and services*
>
> 5) *Employ techniques to represent semantics, ontologies and thesauri that promote interchangeability*
>
> 6) *Enable automated search of catalogs*
>
> 7) *Define basic interfaces (types and operations) to be reused and extended as appropriate to support aggregation, composition and chaining of existing, new and hybrid services.*
>
> 8) *Use UML models to define abstract interfaces and toolkits to generate appropriate bindings for XML Schema, WSDL, HTTP GET, HTTP POST and SOAP.*
>
> 9) *Define general mechanisms for composing chains of interworking OWS services using enabling emerging or de-facto "standards" such as BPML, BPEL and WSCI.*
>
> 10) *Define an Application Program Interface (API) for a general-use interface to Earth Science data products and applications (e.g., Web portal technologies)*

ARCHITECTURAL CONCEPTS

We describe here specific dimensions of the OWS service architecture that would benefit from a Grid computing approach. First, a brief introduction to Grid Services is given.

The notion of a Grid Service has developed through combining elements of Web Services and Grid computing. It is emerging as the dominant paradigm for distributed Grid architectures. The vision was first outlined in [viii] and is being refined in the Open Grid Services Infrastructure (OGSI) specification. Version 1.0 of this specification [xvii] is, at the date of this submission, in the final public comment stage of standardisation through the Global Grid Forum (GGF)[3]. Grid Services are Web Services that conform to a set of conventions that "... provide for the controlled, fault resilient, and secure management of the distributed and often long-lived state that is commonly required in advanced distributed applications" [xvii]. Standard factory and registration interfaces are also introduced for creating and discovering Grid services. Grid Services currently use the Web Service Description Language (WSDL) to define their public interfaces, but require extensions in two specific areas:

[3] The GGF is the *de-facto* standards body for Grid interoperability. It operates through a document process modelled after the IETF's RFC series. A number of Working Groups consider the full range of issues associated with developing Grid infrastructure standards. See http://www.gridforum.org for further information.

1. For interface (`portType`) inheritance

2. To describe additional information elements on a `portType`.

These deficiencies are being addressed by the W3C Web Services Description Working Group, and the GGF commits to updating the OGSI specification for WSDL compliance once WSDL 1.2 is published in draft.

We refer to [xvii] for details of Grid Service lifetime management (including instantiation), referencing, fault handling, and operation inheritance and concentrate instead on the representation of state within Grid Services. We note, however, that interface inheritance supports requirement (7) above.

Grid Services expose state through public `serviceData` elements associated with given interfaces (`portType`s). These elements may be manipulated by name through base operations defined for all Grid Services in a mandatory `GridService portType`. An example of `serviceData` within a WSDL description of a Grid Service interface is shown below:

```
<wsdl:definitions xmlns:tns="xxx" targetNamespace="xxx">
  <gwsdl:portType name="exampleSDUse">
    <wsdl:operation name=…>
 …
    <sd:serviceData name="sd1" type="xsd:String" mutability="static"/>
    <sd:serviceData name="sd2" type="tns:SomeComplexType"/>
 …
    <sd:staticServiceDataValues>
      <tns:sd1>initValue</tns:sd1>
    </sd:staticServiceDataValues>
  </gwsdl:portType>
 …
</wsdl:definitions>
```

Initial values may be specified for "static" `serviceData` elements. Every Grid Service must provide the mandatory `GridService portType` which provides a number of operations to manipulate `serviceData`. The `findServiceData` operation allows sophisticated queries to be executed against the `serviceData` elements, while the `setServiceData` operation allows values to be set for modifiable `servicedata`.

We reiterate that Grid Service extensions to WSDL are being integrated into the W3C standard to provide for statefulness in conventional Web Services.

Referring now to the OGC Service Information Model, the Capabilities XML document trivially becomes an integral component of an OGC Grid Service's `serviceData`. Introspection provides a standardised mechanism for discovering the service capabilities in the RM-ODP Information Viewpoint. Requirements (1) through (4) above, and objective (1) are satisfied through a standard mechanism.

APPROACH

As well as meeting the requirements for self-describing metadata representation, `serviceData` allows state to be exposed. We consider in broad terms how this might be used in the oil-spill scenario outlined in the Overview section earlier. Both the Web Coverage Service and an oil-spill Simulation Service could exhibit significant latency in the case of a complex request. A client (user or agent) of the service would like to query the progress of the data request or simulation. This could be exposed as service state through the `serviceData` mechanism. As well as providing for synchronous introspection of state, the Grid Service specification defines asynchronous notification mechanisms for tracking changes of state. A Grid Service may provide a `NotificationSource portType` to which clients may subscribe (through a `subscribe` operation), and a `NotificationSink portType` to which clients may send notification messages (through the `deliverNotification` operation). Details may be found in [xvii]. In the proposed scenario, the Web Coverage Service would implement the `NotificationSource` to which the Simulation Service would subscribe. Upon completion of the request for required met/ocean forecast data, a message would be sent by the WCS to the Simulation Service through a corresponding `NotificationSink`. This asynchronous notification framework will enable robust Grid workflows to be composed, addressing requirement (9), and objectives (2) and (3) above.

CONCLUSIONS AND RECOMMENDATIONS

Grid technology, and the Open Grid Services Infrastructure in particular, are developing at a considerable pace. Key industry players are actively involved, and GGF is influencing W3C development of the WSDL specification. An Interoperability Program testbed examining the role of Grid technology would assist in positioning geographic information systems as a core Grid customer, in line with objective 6 noted earlier.

IMAGE HANDLING PAT

PURPOSE AND SCOPE

A broad objective identified in the Image Handling PAT is:

- *To make government imagery archives accessible for broader use*

National data centres manage large and valuable community resources. An explicit aim of the NERC DataGrid project is to make access to such archives in the UK much easier than at present, thereby enhancing the return on investment in these centres. The project is focussing initially on data held by the British Atmospheric Data Centre (BADC), the British Oceanographic Data Centre (BODC) and the NERC Earth Observation Data Centre (NEODC). The latter, in particular, curates a large volume of remote-sensed imagery. The demand for such data will only increase as enhanced access technology is deployed. As well as traditional satellite-based sensors, NERC DataGrid has a particular interest in data from airborne instruments. NEODC manages a considerable amount of data from the NERC's Airborne Remote Sensing Facility[4]. An OGC Web Mapping Service interface to NEODC archives (including ATSR and Landsat data) is being investigated. More generally, we believe it would be useful for OGC technology to be applied across a broad range of environmental data sets.

Problems and solution summary

A Grid-based solution would assist with the three problems outlined in the Image Handling PAT:

1. *Very large data volumes*

[4] See http://www.nerc.ac.uk/arsf/home.htm for details.

> *2. Automated processing is required*
>
> *3. Proprietary/closed systems, networks and formats inhibit broader accessibility*

Specific problems associated with very large data volumes (problem 1) include latency (both in assembling a data request, and in transferring results of large requests) and distribution (very large archives may best be distributed). Grid solutions for both are emerging. Grid Service notification mechanisms were discussed in the Common Architecture theme earlier. These provide a mechanism for dealing robustly with latency. A high-performance extension to the FTP protocol (GridFTP) has been developed by the Grid community for transfer of large data volumes. It has been demonstrated in a demanding scenario transferring over 200 GB of climate simulation data in an hour over a high-speed network between Dallas and Chicago [xviii].

An alternative to transferring very large data volumes between processing nodes on wide area networks is to co-locate processing and data services. A Grid-enabled Image Archive Service could offer data mining, coordinate transformation, format translation and other services at the data source.

The service-oriented architecture of Grids is directed at implementing open, interoperable systems (problem 3). The solution lies in applying usage-level metadata internally to services which offer flexible, open, and standard public interfaces and delivery options. By analogy, the Web Map Service and Web Coverage Service offer the ability to request products in a desired format, regardless of the internal data format.

OBJECTIVES

We consider the following OWS2 Image Handling theme objectives:

> *1. Speed availability of information through standard data formats, open and available protocols, and standard validation and verification information*
>
> *6. Develop image processing services as loosely-coupled Web services, e.g., data mining on imagery, enabled by service chaining.*

The use of a high-performance data transport protocol like GridFTP supports objective (1), while an architecture based on Grid Services is consistent with objective (6). These are discussed further in the Architectural Concepts below.

ARCHITECTURAL CONCEPTS

In parallel with the Open Grid Services Infrastructure, a specification for a higher level set of core services is being developed by the Global Grid Forum. The Open Grid Service Architecture (OGSA) Platform [xix] is identifying requirements for core standardised Grid functionalities including data management. Interfaces being considered include data access services, data replication, metadata catalog and storage services.

GridFTP [xx,xxi] is a protocol extension to FTP tuned for high-performance, security and large data volumes. Middleware implementations have been developed by the Globus Project[5]. Important extensions of GridFTP over the FTP protocol include:

- x.509 based security on both control and data channels
- multiple data channel streams for parallel transfer
- partial file transfer
- third-party (server to server) transfers
- TCP tuning

The self-description, notification, and state-introspection mechanisms of Grid Services were described under the Common Architecture theme. Grid Services are suggested as a robust technology for building the scalable, loosely-coupled image processing services outlined in objective (6). Workflow mechanisms for both Web Services [xxii] and Grid Services [xxiii] are under active development.

A fundamental principle of service-oriented Grid architectures is resource virtualisation. For an Image Archive Service, this includes providing an interface which abstracts the essential imagery data away from underlying format details. Both problem 3 and objective 1 are addressed with a Grid Service approach.

APPROACH

We describe now how Grid technology relevant to Image Handling could be incorporated into the oil-spill scenario.

Two requirements to be considered for the Image Archive Service include (from the RFT):

- *Large file handling. Images delivered from an IAS using OGC Web Services shall be capable of handling file on the order of 100s of Mbytes*
- *Long transactions. Some WCS access involving subsetting and reprojection of large images, may require an interaction pattern longer than the typical WCS request and response.*

A GridFTP-enabled Image Archive Service (eg WCS) would provide efficient and secure access to large files. Service chaining with large data volumes is facilitated through the third-party transfer mechanism of GridFTP. In our use-case scenario, an intermediate processing service (say a Coordinate transformation Service) may be needed in the data flow between the Image Archive Service and the Coverage Portrayal Service. Third-party GridFTP transfers are used to move the data between these three services. The user's Grid credentials (through a proxy x.509 certificate) are used to authenticate to the respective services.

We outlined in the Common Architecture theme how the incorporation of service state and notification provide mechanisms for dealing with latency. The Open Grid Services Infrastructure specification [ix] supports soft-state lifetime management for Grid Services. A client may send keepalive

[5] The Globus Project (http://www.globus.org) is developing advanced implementations of emerging Grid middleware.

messages to a Grid Service to extend advertised timeouts. In our scenario, the Image Archive Service might use the soft-state lifetime mechanisms of the Coverage Portrayal Service to ensure the latter remains available until the SAR imagery has been prepared. In general, the lifetime management capabilities of Grid Services are designed to provide for long service transaction times (second requirement above).

OGC Web Services conform closely to a key goal of Grid technology – the virtualisation of resources. Thus both the Web Map Service and Web Coverage Service hide data storage details. Internally, however, markup languages may be needed to manage and organise data files. As identified in the RFT, the Earth Science Markup Language is a promising candidate. Other markup languages addressing the issue of usage-level metadata for earth-science data sets include the Climate Data Markup Language (CDML, [xxiv]) and the netCDF Markup Language [xxv]. An Interoperability Program testbed would provide an opportunity to evaluate different schemas. In our scenario, usage-level metadata (in the form of ESML or other) is used internally by the Image Archive Service, allowing image files to be aggregated and presented by the service as a logical data source.

MODELLING AND SIMULATION PAT

PURPOSE AND SCOPE

The earth sciences community use a large range of simulation models for a variety of applications. Examples include high-end meteorological and oceanographic forecast and data assimilation systems, storm surge and flood prediction models, as well as simulations used for process studies. Operational meteorological centres routinely market real-time products to a range of customers in support of logistics and other operations. There is growing interest in the community to the application of novel IT solutions for access to both models and their output. A number of groups[6] are investigating the application of Grid technology to collaborative visualisation and computational steering. NERC DataGrid is devoting considerable effort to supporting Grid Service access to earth science model output data, and is collaborating with the Earth System Grid project on metadata schemas to support search and discovery of model data.

The potential market for standards-based web service access to earth science model output is considerable. This includes disaster management, resource exploration, and the transport sector amongst others.

Problems and solution summary

The problem statement identified in the RFT is:

> *Simulation capabilities are often difficult to integrate into a broader software application infrastructure. Some of the challenges are:*
>
> - *Geospatial data is often not easily exchangeable between simulations/models and geospatial-enabled applications.*
> - *Simulations usually employ a proprietary data model. Yet SISO, W3C, ISO and others have made progress on defining baseline standards for data representation and exchange.*

[6] See, for example, the UK e-Science projects CovisaG (http://www.visualization.leeds.ac.uk/CovisaG/) and RealityGrid (http://www.realitygrid.org/).

- *OGC's XML-based technologies (GML, SensorML and O&M) do not necessarily address the requirements for representing simulated data and simulation parameters, so they may need to be enhanced for this purpose.*

- *Simulation capabilities can be difficult to describe.*

It should be possible to share data between an "OGC Web Services environment" and simulations/models. It should also be possible to interoperate with simulations/models in much the same manner as OGC has done with the Sensor Web Enablement specifications for all types of sensors. In order to incorporate simulations into a Web Service based infrastructure such as the OWS framework, it will be necessary to define the appropriate metamodel for simulations, and also define the means to monitor, parameterize, control, use and manage simulations/models runtime.

There are significant shortcomings of current OGC specifications in relation to meteorological (and hydrological) simulation data which limit the application of OGC technology in the broader earth sciences. Simulations of this type may produce very large volumes of data (mega- to tera-scale), typically stored in binary files rather than database systems. The most common file formats are small in number (eg netCDF, HDF, GRIB) and self-describing to some extent[7] but are yet to be mapped onto OGC schemas. Perhaps the greatest difficulty is due to the geometries associated with simulations. These are invariably grids of considerably greater complexity than supported by GML and the coverage type of the Abstract Specification. Extension of OGC specifications to support climate simulation data will greatly enhance its applicability to a large body of earth science work.

We mention some prior work related to the problem statement:

- The NERC DataGrid project is developing metadata and data models to support access to earth science model output data, as well as observational data.

- A number of markup languages (eg CDML, ESML, NcML) have been developed for describing usage-level information about simulation output. These were mentioned above in the Image Handling theme.

- The CCLRC e-Science program has an ongoing project investigating the automated capture of metadata from climate simulations[8].

- A web service model has recently been proposed for access to climate simulation data [v].

OBJECTIVES

Modelling and simulation objectives in the RFT include:

1. Pick an M&S domain(s) and type of simulation/model to demonstrate integration between GIS-based applications and a legacy simulation system. In particular, provide the means to exchange geospatial data.

[7] Usage-level metadata is contained within the file.

[8] See project documentation at http://www.e-science.clrc.ac.uk/web/projects/climatesimulations.

> *2. Construct interfaces through which an external user can monitor, control/use and manage simulations.*
>
> *3. Define/adopt/enhance a "Simulation Model Language" that allows appropriate model description.*
>
> *4. Where possible, use approaches and technologies that are already developed and accepted in the marketplace.*

Meteorological forecast data has a very broad market and is relevant both to the oil-spill scenario and other toxic dispersion problems. It would be a promising candidate domain. We note that EuroGrid WP2 (http://www.eurogrid.org/wp2.html) is concerned with developing a Grid control interface to a regional forecast model. Global-scale meteorological assimilation and forecast systems, on the other hand, are far too computationally expensive to implement as a web service. Web service control of an oil-spill simulation model, however, is more tractable. The required inputs to such a model include the aforementioned simulated meteorological and oceanographic forecast data.

REQUIREMENTS

RFT requirements to which our approach is addressed include:

> *1. Simulated Data Representation*
>
> *2. Simulated Data Access*
>
> *4. Simulation Management Service*

ARCHITECTURAL CONCEPTS

As in the Common Architecture and Image Handling themes, we propose an architecture based on Grid Services. A central theme with Grid computing – compatible with OGC Web Services – is that of virtualising resources. Essential semantic capability and content is abstracted onto a service interface. Computational services may be invoked without regard to the underlying implementation and data resources may be accessed without knowing storage or location details. Thus, simulated data representation and access is seen as a data abstraction/virtualisation problem. An extension of OGC application schemas in relation to simulation coverage data will support requirements 1 and 2.

In our oil spill scenario, a Data Service provides access to meteorological forecast simulation data. This might be an OGC Web Coverage Service enhanced to support simulation output. Alternatively, a new Model Output Service might be necessary, drawing on an enhanced GML specification for the application schema.

APPROACH

We refer once more to our oil-spill scenario. There are two model interactions in this scenario:

1. Search/discovery and access to meteorological and oceanographic forecast data to drive the oil-spill simulation model

2. Control of the oil spill simulation (involving model setup, simulation parameter specification, and invocation)

We concentrate on the first of these and outline some required developments of OGC specifications.

Conceptually, simulation model output is a geographic feature of coverage subtype. To access simulation results, the Web Coverage Service could be extended, or a new Model Output Service defined. The major requirement for either of these is an extension of the coverage type in the Abstract Specification Topic 6 to handle gridded simulation output. Currently, the coverage type provides only for non-georeferenced grids, or grids with uniform spacing with respect to some coordinate reference system. In practice, a much richer description of grid coverages is needed to support simulation output. We list here a number of relevant factors:

- Most meteorological simulation models are formulated on one of a class of finite-difference grids (the so-called Arakawa grids, types A through E [xxvi]).
- With respect to the earth, these grids are often non-uniformly spaced in any conventional spatial reference system. This is done in order to increase the model's resolution over some region of interest.
- For ocean models, it is common to apply a conformal mapping to the grid [xxvii] and artificially relocate one (or both) of the curvilinear grid poles over land, say North America.
- Some models are formulated in the spectral, rather than the spatial domain, and an intermediate transformation is required to obtain output in the spatial domain. This *grid-point* data may have decreased resolution towards the poles[9] (fewer zonal gridpoints with increasing latitude).

All of these characteristics require extensions to the GridCoverage defined in Topic 6 of the Abstract Specification, and the Grids schema of GML. A community-led activity (the "CF" convention [xxviii]) is underway, with active and widespread support, to develop standard metadata encoding methods for meteorology and oceanography simulation data. This includes mechanisms for representing a range of georeferenced grids used in these models.

There is additional complexity associated with the vertical coordinate in simulation models. As well as a discretisation in height or depth, a pressure or terrain-following coordinate may be used. In this case, model layers are not horizontal with respect to the geoid. The Web Coverage Service could be used in this case by extending the elevation-related parameters (in the capabilities XML and the `getCoverage` operation), though a better solution would be a general extension of three-dimensional spatial reference systems for simulation data.

Finally, there is a considerable amount of legacy software in use supporting the major file formats extant: netCDF, HDF and GRIB. An application schema for simulation data based on GML would need to support at least these in the `gml:FileValueModelType`.

---

i Live Access Server [Online: http://ferret.wrc.noaa.gov/Ferret/LAS/ferret_LAS.html; Available June 1, 2003]

---

[9] These *Gaussian grids* are designed to ensure alias-free transformations between the spectral and spatial domains.

ii Gallagher, J., and G. Milkowski, "Data Transport Within The Distributed Oceanographic Data System", Proceedings of the Fourth World Wide Web Conference, Boston, Massachussetts, USA, December 1995 (http://www.w3.org/Conferences/WWW4/Papers/67/).

iii Distributed Oceanographic Data System [Online: http://www.unidata.ucar.edu/packages/dods/; Available 1 June, 2003]

iv OPeNDAP [Online: http://opendap.org; Available 1 June, 2003]

v Woolf, A., K. Haines and C. Liu, "A Web Service Model for Climate Data Access on the Grid", Int. J. High Perform. Comp. Appl., *in press.*

vi Fielding, R.T., "Architectural Styles and the Design of Network-based Software Architectures", PhD dissertation, University of California, Irvine, 2000.

vii Simple Object Access Protocol [Online: http://www.w3.org/TR/SOAP/; Available 1 June, 2003]

viii Foster, I., C. Kesselman, J. Nick and S. Tuecke, "The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration", Open Grid Service Infrastructure WG, Global Grid Forum, June 22, 2002 [Online: http://www.globus.org/research/papers/ogsa.pdf; Available 1 June, 2003]

ix Tuecke, S. *et al*, "Open Grid Services Infrastructure (OGSI)", Global Grid Forum [Online: http://www.gridforum.org/Meetings/ggf7/drafts/draft-ggf-ogsi-gridservice-23_2003-02-17.pdf; Available 1 June, 2003]

x BADC [Online: http://badc.nerc.ac.uk/; Available 1 June, 2003]

xi NERC Data Grid [Online: http://ndg.badc.rl.ac.uk; Available 1 June, 2003]

xii CCLRC [Online: http://www.cclrc.ac.uk; Available 1 June, 2003]

xiii CCLRC e-Science Centre [Online: http://www.e-science.clrc.ac.uk; Available 1 June, 2003]

xiv NERC Earth Observation Data Centre [Online: http://www.neodc.rl.ac.uk; Available 1 June, 2003]

xv PKI x.509 Charter [Online: http://www.ietf.org/html.charters/pkix-charter.html; Available 1 June, 2003]

xvi Web Service Security (WS-Security) [Online: http://www-106.ibm.com/developerworks/webservices/library/ws-secure/; Available 1 June, 2003]

xvii Open Grid Services Infrastructure (OGSI) Version 1.0 [Online: http://prdownloads.sourceforge.net/ggf/draft-ggf-ogsi-gridservice-29_2003-04-05.pdf; Available 1 June, 2003]

xviii Allcock, B., *et al*, "High-Performance Remote Access to Climate Simulation Data: A Challenge Problem for Data Grid Technologies", *Supercomputing 2001* [Online: http://www.globus.org/research/papers.html#sc01ewa, Available 1 June, 2003]

xix The Open Grid Services Architecture Platform [Online: http://www.gridforum.org/Meetings/ggf7/drafts/draft-ggf-ogsa-platform-2.pdf; Available 1 June, 2003]

xx GridFTP: Protocol Extensions to FTP for the Grid [Online: http://www-fp.mcs.anl.gov/dsl/GridFTP-Protocol-RFC-Draft.pdf; Available 1 June, 2003]

xxi The Globus Project, "GridFTP: Universal Data Transfer for the Grid" [Online: http://www.globus.org/datagrid/deliverables/C2WPdraft3.pdf; Available 1 June, 2003]

xxii Web Services Choreography Working Group [Online: http://www.w3.org/2002/ws/chor/, Available 1 June, 2003]

xxiii GSFL: A Workflow Framework for Grid Services [Online: http://www-unix.globus.org/cog/projects/workflow/, Available 1 June, 2003]

xxiv Climate Data Markup Language, Chapter 6 of the User Guide for the Climate Data Management System [Online: http://esg.llnl.gov/cdat/cdms.pdf; Available 1 June, 2003]

xxv The netCDF Markup Language (NcML) [Online: http://www.vets.ucar.edu/luca/netcdf/; Available 1 June, 2003]

xxvi Arakawa, A. and V.R. Lamb, “Computational Design of the Basic Dynamical processes of the UCLA General Circulation Model”, Meth. Comp. Phys., **17**, 173-265 (1977).

xxvii Murray, R.J., “Explicit Generation of Orthogonal Grids for Ocean Models”, J. Comp. Phys., **126**, 251-273 (1996).

xxviii NetCDF Climate and Forecast (CF) Metadata Convention [Online: http://www.cgd.ucar.edu/cms/eaton/cf-metadata/, Available: 1 June, 2003]