STANDARDS-BASED DATA INTEROPERABILITY IN THE CLIMATE SCIENCES

Corresponding author: **Dr Andrew Woolf**
Address: **e-Science Centre, CCLRC Rutherford Appleton Laboratory**
                          Chilton, Didcot
                          Oxon OX11 0QX, UK
Phone: **+44 (0)1235 778027**
Fax: **+44 (0)1235 445945**
Email: **A.Woolf@rl.ac.uk**

Author: Ray Cramer
Address: British Oceanographic Data Centre
                          Bidston Hill
                          Merseyside CH43 7RA, UK
Email: rnc@bodc.ac.uk

Author: Marta Gutierrez
Address: British Atmospheric Data Centre
                          SSTD, CCLRC Rutherford Appleton Laboratory
                          Oxon OX11 0QX, UK
Email: M.Gutierrez@rl.ac.uk

Author: Kerstin Kleese van Dam
Address: e-Science Centre, CCLRC Daresbury Laboratory
                          Daresbury, Warrington
                          Cheshire WA4 4AD, UK
Email: k.kleese@dl.ac.uk

Author: Siva Kondapalli
Address: British Oceanographic Data Centre
                          Bidston Hill
                          Merseyside CH43 7RA, UK
Email: vpk@bodc.ac.uk

Author: Susan Latham
Address: British Atmospheric Data Centre
                          SSTD, CCLRC Rutherford Appleton Laboratory
                          Oxon OX11 0QX, UK
Email: S.E.Latham@rl.ac.uk

Author: Bryan Lawrence
Address: British Atmospheric Data Centre
                          SSTD, CCLRC Rutherford Appleton Laboratory
                          Oxon OX11 0QX, UK
Email: B.N.Lawrence@rl.ac.uk

Author: Roy Lowry
Address: British Oceanographic Data Centre
                   Bidston Hill
                   Merseyside CH43 7RA, UK
Email: rkl@bodc.ac.uk

Author: Kevin O'Neill
Address: e-Science Centre, CCLRC Rutherford Appleton Laboratory
                   Chilton, Didcot
                   Oxon OX11 0QX, UK
Email: K.D.O'Neill@rl.ac.uk

STANDARDS-BASED DATA INTEROPERABILITY IN THE CLIMATE SCIENCES

Andrew Woolf[1], Ray Cramer[2], Marta Gutierrez[3], Kerstin Kleese van Dam[1], Siva Kondapalli[2], Susan Latham[3], Bryan Lawrence[3], Roy Lowry[2] and Kevin O'Neill[1]

[1]CCLRC e-Science Centre
[2]British Oceanographic Data Centre
[3]British Atmospheric Data Centre

ABSTRACT

Emerging developments in geographic information systems and distributed computing offer a roadmap towards an unprecedented spatial data infrastructure in the climate sciences. Key are the standards developments for digital geographic information being led by the International Standards Organisation (ISO) technical committee on geographic information/geomatics (TC211) and the OpenGIS Consortium (OGC). These, coupled with the evolution of standardised web services for applications on the Internet by the World Wide Web Consortium (W3C), mean that opportunities for both new applications and increased interoperability exist. These are exemplified by the ability to construct ISO-compliant data models that expose legacy data sources through OGC web services. This paper concentrates on the applicability of these standards to climate data by introducing some examples and outlining the challenges ahead. An abstract data model is developed, based on ISO standards, and applied to a range of climate data – both observational and modelled. An OGC Web Map Server interface is constructed for numerical weather prediction (NWP) data stored in legacy data files. A W3C web service for remotely accessing gridded climate data is illustrated. Challenges identified include the following: firstly, both the ISO and OGC specifications require extensions to support climate data. Secondly, OGC services need to fully comply with W3C web services, including support for complex access control. Finally, to achieve real interoperability, broadly accepted community based semantic data models are required across the range of climate data. These challenges are being actively pursued, and broad data interoperability for the climate sciences appears within reach.

1. INTRODUCTION

A spatial data infrastructure (SDI) has been defined [i] as "the technology, policies, standards, and human resources necessary to acquire, process, store, distribute, and improve utilization of geospatial data". It is the broad concept underlying a significant and directed effort to facilitate access to Earth-related data.

Political, economic, and civil endeavours depend increasingly on a capable SDI. Atmospheric and oceanographic (collectively 'climate-science') data, for instance, underpin and inform the global political response to rising anthropogenic carbon emissions [ii]. A number of authors [iii,iv,v] have reported significant agricultural and other economic impacts from improved weather forecast data. And the forecasts themselves depend on assimilating a wide variety of heterogeneous and distributed observational data within state-of-the-art numerical models. Disaster recovery, decision support, and risk management are amongst other activities critically dependent on climate-science data. Underpinning all these activities are infrastructures that – with varying degrees of success – enable data to be located, distributed, and interpreted for a range of purposes.

Several national governments have embraced the strategic importance of establishing extensive SDIs. The USA's National Spatial Data Infrastructure (NSDI) initiative[1] was one of the first to expound a broad vision and strategy for a national SDI. Many others are now

---

[1] http://www.fgdc.gov/

investing significant effort, including Canada[2], Australia[3], South Africa[4] and the European Union[5].

The motivation behind these initiatives may be defined succinctly as an agenda of "interoperability" for spatial data. In broad, interoperability requires agreements on metadata schemas and formats, data models and encodings, and service interfaces for accessing both data and discovery metadata [vi]. A gradual move is underway, away from proprietary, closed, or 'community' infrastructures towards systems based on open, inclusive, internationally-agreed standards. These include, for instance, the FGDC Content Standard for Digital Metadata[6], the OpenGIS Consortium's Web Map, Feature, and Coverage Services (OGC WMS, WFS, WCS), and a series of ISO standards for the "description and management of geographic information and ... services" [vii]. As an example, the 'Geospatial One-Stop Portal' (http://www.geodata.gov) has been established as part of the US NSDI. It provides searching across a large volume of FGDC-compliant metadata, and links transparently to an interactive OGC WMS viewer.

Comparatively speaking, climate-science data is migrating less rapidly to a standards-based approach. The most common network data services are the community Live Access Server[7] (LAS) for visualisation and DODS/OPeNDAP[8] for data delivery. Neither of these are based on the above international standards, though both are evolving to use XML.

Concurrently with the spatial data developments, rapid advances are being made in the business sphere towards open standards for discovering and integrating loosely-coupled networked services ("web services") in value-added processing chains. Service interfaces are described in a standard manner, registered in standardised discovery repositories, and invoked using standardised protocols. This "publish-find-bind" pattern defines so-called "service oriented architectures" that are finding widespread appeal in industry, and form the basis of Grid computing. Web service standards are being developed through the World Wide Web Consortium (W3C) and OASIS.

In surveying the abovementioned standards, a canonical "interoperability framework" for geospatial data begins to emerge. Domain-specific data models may be constructed, based on the conceptual modelling framework and component standards of ISO TC211. Legacy data sources may be encapsulated through wrappers compliant with these data models, and exposed through standardised web service "publish-find-bind" architectures. Obvious candidate services are those (WMS, WFS, WCS) being specified by the OpenGIS Consortium.

An example of a project developing such an infrastructure for climate-science data is the UK's NERC DataGrid [viii]. This project is aiming to deploy technology for uniform discovery and access to a wide variety of environmental data. The initial focus is on data curated by the British Atmospheric and Oceanographic Data Centres – two of the "designated data centres" of the government's Natural Environment Research Council agency. The architecture is based

---

[2] http://cgdi.gc.ca
[3] http://www.ga.gov.au/nmd/asdi/
[4] http://www.nsif.org.za/
[5] http://www.ec-gis.org/inspire/
[6] The widely-deployed FGDC metadata standard is being harmonised with ISO 19115.
[7] http://ferret.pmel.noaa.gov/Ferret/LAS/ferret_LAS.html
[8] http://www.unidata.ucar.edu/packages/dods/

on integrating standards wherever possible, and the roadmap outlined above is being pursued.

This paper explores the roadmap as it applies to climate-science data. Section 2 outlines the ISO TC211 series of standards for geographic information and services, and describes a compliant data model for a range of atmospheric and oceanographic data. Section 3 examines the service specifications being developed by the OpenGIS Consortium and demonstrates a Web Map Server for data from the ECMWF's 40-year climate reanalysis. Section 4 details the service-oriented architectures being enabled through W3C web service standards and illustrates a web service for remotely accessing gridded climate model data. Challenges remain before the three standards components may fully be integrated, and these are detailed in Section 5. A summary and conclusions are presented in Section 6.

2. ISO TC211 STANDARDS

The Technical Committee 211[9] (TC 211) of the International Organisation for Standardisation (ISO) is developing a range of standards for digital geographic information, and its discovery, distribution, management and use. Table 1 lists the current projects and their status (as of 28 June, 2004).

Table 1: Current ISO TC211 projects (Working Draft, Committee Draft, Draft International Standard, Final Draft International Standard, International Standard)

19101	Reference model
19101-2	Reference model - Part 2: Imagery
19103	Conceptual schema language
19104	Terminology
19105	Conformance and testing
19106	Profiles
19107	Spatial schema
19108	Temporal schema
19109	Rules for application schema
19110	Feature cataloguing methodology
19111	Spatial referencing by coordinates
19112	Spatial referencing by geographic identifiers
19113	Quality principles
19114	Quality evaluation procedures
19115	Metadata
19115-2	Metadata - Part 2: Extensions for imagery and gridded data
19116	Positioning services
19117	Portrayal
19118	Encoding
19119	Services
19120	Functional standards
19121	Imagery and gridded data
19122	Geographic information/Geomatics - Qualifications and Certification of Personnel
19123	Schema for coverage geometry and functions
19124	Imagery and gridded data components
19125-1	Simple feature access - Part 1: Common architecture
19125-2	Simple feature access - Part 2: SQL option
19126	Profile - FACC Data Dictionary
19127	Geodetic codes and parameters
19128	Web Map server interface
19129	Imagery, gridded and coverage data framework
19130	Sensor and data models for imagery and gridded data
19131	Data product specifications
19132	Location based services framework
19133	Location based services tracking and navigation
19134	Multimodal location based services for routing and navigation
19135	Procedures for registration of geographical information items

---

[9] http://www.isotc211.org

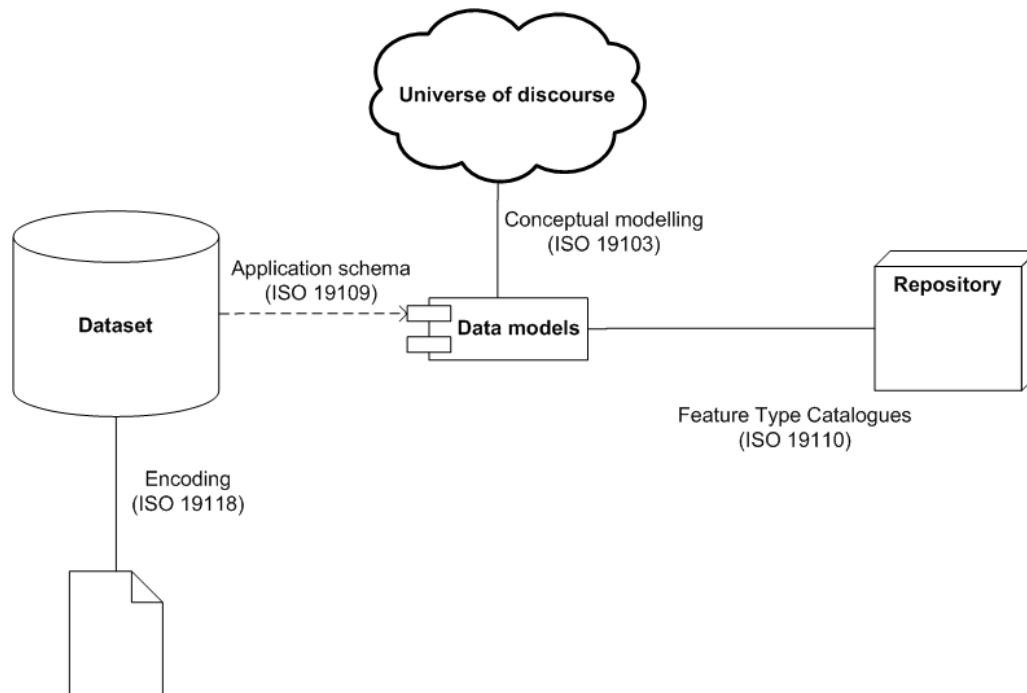| 19136	Geography Markup Language (GML)
19137	Generally used profiles of the spatial schema and of similar important other schemas
19138	Data quality measures
19139	Metadata - Implementation specification
19140	Amendment to the ISO 191** Geographic information series of standards for harmonization and enhancements

The scope of these ISO standards is extensive. ISO 19101 [vii] outlines the overall framework for the series of standards. It prescribes a broad definition of interoperability for geographic information, and proceeds to outline the role of conceptual modelling for both information and services within the standards. A Domain Reference Model provides a high-level description of geographic information structure and content, while an Architectural Reference Model provides a taxonomy of computational services associated with geographic information. Finally, a mechanism for integrating a set, or subsets, of base standards for a particular application (known as 'profiling') is outlined.

2.1 ISO data modelling for geographic information

We don't attempt here a complete review of the ISO standards, concentrating instead on the information modelling aspects of interoperability. In respect of this, the reference model (ISO 19101) states:

> Application interoperability refers to the ability for different GIS applications to use and represent data in the same manner. To do this, semantic interoperability is required. Semantic interoperability refers to applications interpreting data consistently in the same manner in order to provide the intended representation of the data. Semantic interoperability may be achieved using translators to convert data from a database to an application. The schemas and implementations described in the ISO 19100 series of standards support this level of interoperability.

A clear program is outlined for achieving semantic interoperability. First, a conceptual model ('application schema') is formed for the universe of discourse – in this case the logical structure and semantic content of a dataset. It is expressed formally through a conceptual

schema language, general rules for which are prescribed in ISO 19103 [ix]. A profile of the

Figure 1: ISO program for semantic interoperability of geographic information

Unified Modelling Language (UML) is specified as the canonical conceptual schema language for the 19xxx standards. Additional rules specific to application schemas are provided by ISO 19109 [x]. This outlines the General Feature Model, which is a meta-model for geographic 'features' – fundamental components of datasets. Defined broadly as an "abstraction of a real world phenomena", a feature type may represent any important aspect of the universe of discourse. As noted by ISO 19109, "(t)he classification of real world phenomena as features depends on their significance to a particular universe of discourse". Atkinson has suggested [xi] that the granularity with which features should be defined depends on the governance structures available to support those definitions. The General Feature Model defines a meta-model for object feature definitions including 'attributes', 'inheritance relations', 'constraints', 'operations' and 'associations'. In defining geographic feature types, conceptual schemas defined in various ISO standards may be integrated. These include, for instance, temporal [xii], spatial [xiii], quality [xiv], gazetteer [xv], and metadata [xvi] schemas. Having defined geographic feature types, their definitions may be registered for re-use in a Feature Type Catalogue, in accordance with ISO 19110 [xvii]. A mechanism for establishing ISO-approved registers (including Feature Type Catalogues) is described in ISO 19135 [xviii]. Geographic datasets are constructed as instances of an application schema containing a collection of feature instances. A canonical means of encoding (or 'serialising') dataset instances is provided by ISO 19118 [xix]. This defines rules for transforming UML classes (objects) to XML schema (documents). An extensive reference encoding is provided in the Geography Markup Language [xx] for a range of conceptual schemas across various standards (temporal, spatial, coordinates, etc). This overall program for semantic interoperability of geographic information is summarised in Figure 1.

2.2 NERC DataGrid data model

As an example of ISO-compliant data modelling, we briefly describe the data model being developed in the NERC DataGrid project. This data model is intended to apply across the range of data curated by both the British Atmospheric and Oceanographic Data Centres
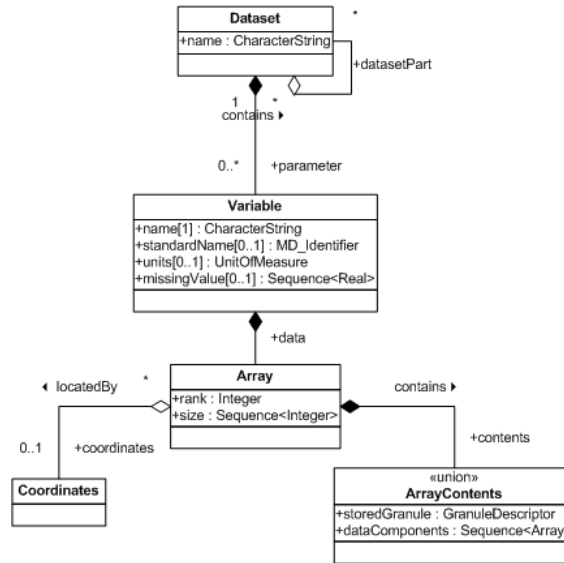

Figure 2: NERC DataGrid abstract application schema

(BADC, BODC). Data types encompass both observational and model data. An initial abstract (weakly-typed) application schema has been described previously [xxi]. We describe below specialisations of the abstract model into feature types currently under development. Key elements of the abstract model are reproduced in UML form in Figure 2.

The initial data model incorporates two crucial elements that apply across almost all environmental data:

1. logical data *structure* (or "shape"), and

2. data *location* in time and space.

At the root level, a named dataset may contain both a number of parameters, and other datasets. A parameter is characterised by its name, physical units, and a flag indicating missing (or bad) data. A standard name from a controlled vocabulary may provide additional parameter type semantics (this includes the namespace authority, for example "BODC data dictionary" or "CF convention").

The parameter's data is structured logically as a multidimensional array, characterised by its rank and size along each dimension. The contents of an array may be either numerical data derived from storage or a further sequence of arrays, one per node of the parent array. This nested hierarchy of multidimensional arrays allows rich and complex data structures to be constructed. An example of its application to a marine science cruise is shown in Figure 3. The research vessel takes a series of salinity measurements down through the water column (a salinity *profile*). This is repeated at a number of different locations. A two-level nesting of one-dimensional arrays may be used to represent this logical structure. A top-level array represents the sequence of individual locations (totalling 50 in this case) at which salinity profiles are measured. Each profile itself may be represented as a one-dimensional array of

measurements. In this example, there are 20 salinity measurements in the first profile, 35 in the second, and 27 in the final profile.
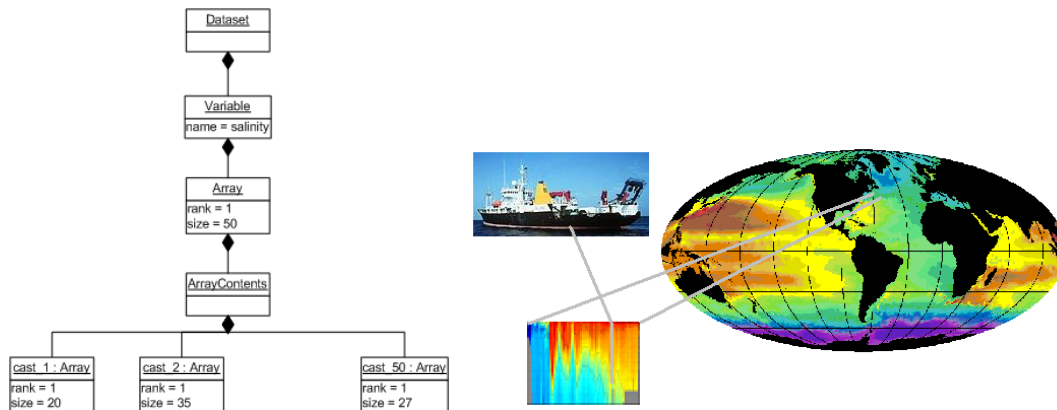

Figure 3: Data model applied to a marine science cruise

Spatiotemporal location of the nodes of an array is accomplished by means of associated coordinates. These are defined with respect to standards-based spatial (ISO 19111) and temporal (ISO 19108) reference systems. Individual ordinates provide values for each axis of the associated reference systems. An ordinate may span one or more dimensions of the corresponding array. Thus a one-dimensional array representing measurements along a radiosonde trajectory will have four associated ordinates providing measurement locations in space and time (Figure 4). Each ordinate spans the same single dimension of the one-dimensional array. A three-dimensional array from an ocean model on a rotated latitude-longitude grid will have three associated ordinates (Figure 4). Each of the latitude and longitude ordinates spans the two "horizontal" dimensions of the model array, while the depth ordinate spans the single third dimension of the model array.
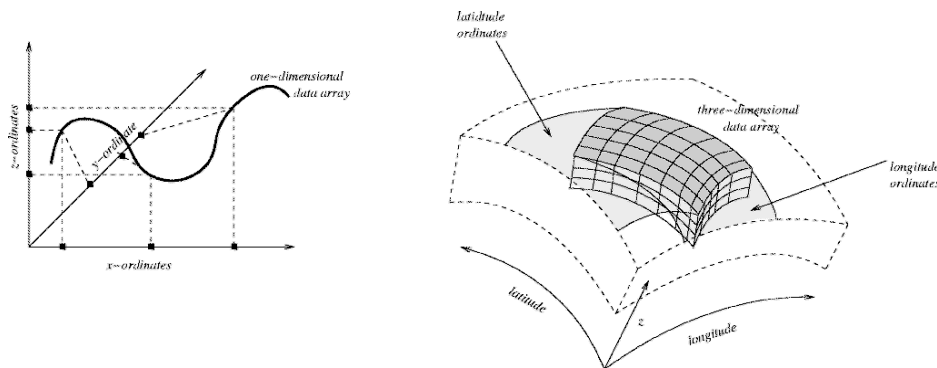

Figure 4: Coordinates in the data model. Left: one-dimensional array. Right: three-dimensional array
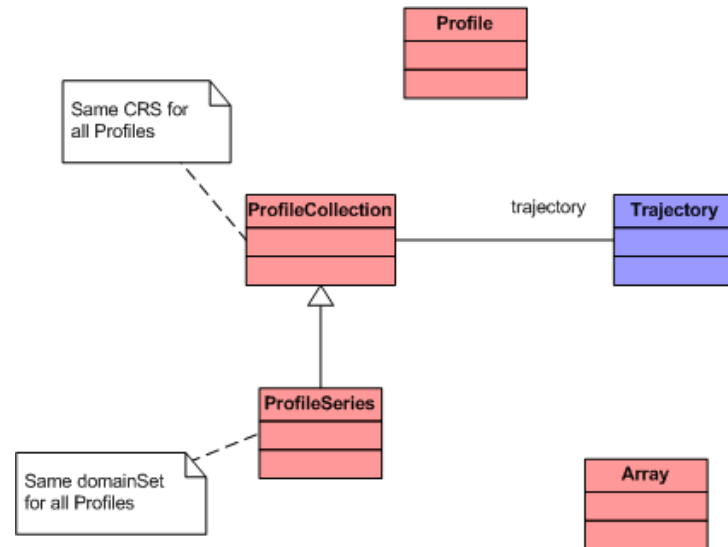
Figure 5: Candidate feature types specialised from NDG abstract data model

This abstract model, based on nested hierarchies of multidimensional arrays is very generic and needs to be specialised into a set of feature types. A range of data types across BADC and BODC have been examined for this purpose. Figure 5 presents an initial set of feature types under consideration for implementation within NERC DataGrid.

They are characterised, briefly, as follows:

Trajectory: a geometric object representing a series of point locations in time and space, e.g. a ship's cruise track or an atmospheric balloon trajectory.

Profile: a series of measurements along a path in time and/or space, e.g. a marine CTD cast, or a radiosonde measurement.

ProfileCollection: a series of Profiles, having the same underlying coordinate reference system, but not necessarily the same lengths or locations, e.g. a cruise section

ProfileSeries: a special case of a ProfileCollection with the same spatial or temporal domain for each Profile, e.g. a thermistor chain (same depths) or scanning radar (same ranges).

Array: a multidimensional gridded array of data, as produced by a numerical model.

A ProfileCollection and ProfileSeries may have an associated Trajectory, representing a notional reference point, either stationary or moving. This would be used, for example, for an airborne lidar.

These feature types fall into two categories – geometry features (Trajectory), or 'coverage' features which define a mapping from some spatiotemporal domain onto a value range (ISO 19123).

3. OGC STANDARDS

The OpenGIS Consortium[10] (OGC) is an international consortium of over 250 industry, government, and academic institutions having the aim of developing open specifications for web-enabled interoperability of spatial data. OGC-conformant services are becoming widely deployed as standard solutions for exchange of geographic information. Specifications are developed through the oversight of various Working Groups and fast-track testbed exercises coordinated by the Interoperability Program. A considerable number of vendors, including, for example, ESRI and Cadcorp, are beginning to offer OGC compliant software.

There is considerable overlap of interests between OGC and ISO TC211. In practice, the two work closely together, principally through a Coordination Group established formally in a 1999 joint agreement. A number of OGC specifications are incorporated in the ISO TC211 program of work. These include the OGC's Geography Markup Language (ISO 19136) and Web Map Server interface (ISO 19128).

We now review some of the flagship OGC specifications.

3.1 Geography Markup Language

The Geography Markup Language (GML) [xxii] is a reference XML schema encoding of a range of ISO conceptual schemas, including parts of ISO 19107, 19108, 19111 and ISO 19123. It is compliant with the encoding rules of ISO 19118. As of February 2004, GML (3.1) has been harmonised with ISO CD 19136 [xx][11].

GML provides a number of integrated XML schemas for feature types, geometry (0, 1, 2 and 3-dimensional primitives and aggregates), coordinate reference systems, topology, temporal information, dictionaries, units of measure, directions, observations and coverages. The document is lengthy (over 550 pages as at version 3.1), but roughly a third consists of verbatim normative XML schemas. It is likely that GML will continue to evolve as a Working Draft in the committee stage before release as an International Standard.

3.2 Web Map Service

The OGC Web Map Service Implementation Specification (WMS) [xxiii] defines a web accessible interface for requesting rendered maps of spatial data. The functionality is similar to that provided by the Live Access Server used in the climate sciences. A WMS request embodies the following parameters:

- required information to be rendered (one or more map 'layers')

- styling of layers

- a bounding box specifying a region of the Earth

- coordinate reference system or map projection to be used

- output image size, format, background colour and transparency

Three separate operations are provided by a WMS server:

---

[10] http://www.opengis.org

[11] Available at http://portal.opengis.org/files/?artifact_id=4700

- **GetCapabilities** (mandatory): returns service level metadata in XML document describing available information content and valid request parameters – reference systems, image formats, etc.

- **GetMap** (mandatory): returns a rendered map

- **GetFeatureInfo** (optional): returns information on particular discrete features that may appear on a generated map

A HTTP GET binding is defined for invoking WMS operations. Figure 6 illustrates a typical WMS GetMap request, in this case for the field 'era40:temperature' over the entire globe using

```
http://localhost:8080/deegreewms/wms?SERVICE=WMS&
 VERSION=1.1.0&REQUEST=GetMap&
 LAYERS=era40:temperature&
 BBOX=-180,-90,180,90&FORMAT=image/gif&
 WIDTH=640&HEIGHT=480&SRS=EPSG:4326
```

Figure 6: Example WMS request

the map projection 'EPSG:4326'[12] (conventional axes of longitude-latitude).

WMS servers can be 'cascaded' to aggregate maps from multiple sources, or to perform processing such as format conversions or coordinate transformations.

A compliant open-source reference implementation of a WMS server is provided by OGC[13].

3.3 Web Feature Service

The Web Feature Service Implementation Specification (WFS) [xxiv] provides a means to obtain GML-encoded 'simple features' from a geo-database. Transactional updates are also supported. Simple features are those that have geometry-valued properties limited to points, lines or polygons, or collections of these. The following five operations are defined:

- **GetCapabilities**: returns service metadata describing supported request types, available feature types and supported transactions (insert, update, or delete), and supported feature filtering operations.

- **DescribeFeatureType**: returns a GML-encoded application schema describing the structure of a named feature type.

- **GetFeature**: returns one or more GML-encoded features matching the specified query. The result set may be limited to a specified subset of feature attributes, or a maximum number of features. In addition, a filter may be specified to constrain the result set (for instance on a spatial domain).

---

[12] Two namespaces are supported for map projections – 'EPSG' corresponding to the extensive set of projection codes defined by the European Petroleum Survey Group, and 'AUTO' for a class of projections including an arbitrary centre of projection.

[13] The 'deegree' reference WMS implementation is available at http://deegree.sourceforge.net/src/demos.html.

- **Transaction**, **LockFeature**: These operations are used to support transactions to create, update or delete features from the WFS server's database.

Both HTTP GET and POST bindings are defined for a WFS server. As with the WMS, OGC provides an open-source reference implementation[14] of a WFS server.

3.4 Web Coverage Server

While a WMS provides rendered maps and WFS provides actual data for simple features, the Web Coverage Service (WCS) [xxv] defines an interface for accessing gridded 'coverage' data, for example remote-sensed imagery. Coverages are returned in one of a number of supported encoding formats. Defined operations are:

- **GetCapabilities**: returns XML-encoded metadata describing the service, supported formats, and available coverage data (summary descriptions including spatiotemporal extent).

- **DescribeCoverage**: describes one or more named coverages in terms of their detailed spatiotemporal domains, range description and value types (e.g. scalar- or vector-valued), supported encoding formats, and coordinate reference systems.

- **GetCoverage**: requests gridded coverage data from a specified spatiotemporal bounding box, with specified resolution and interpolation method (if interpolation is supported), coordinate reference system and format.

Any output format may be supported, including at least one of GeoTIFF, HDF-EOS, DTED, NITF, or GML. There is, as yet, no official reference implementation of a WCS server from OGC, however the open-source deegree reference WMS server provides also a WCS server.

3.5 Example WMS for climate reanalysis

The abovementioned reference implementation ('deegree') of a WMS server by default supports underlying tiled raster data sources in various image formats (TIFF, GIF, BMP, PNG, JPEG). However, gridded climate simulation data is available typically in the netCDF binary file format. We have thus modified the deegree WMS server to connect to a netCDF file store. Figure 7 shows an example global snapshot of surface air temperature for 27/4/2001 from the ECMWF 40-year climate reanalysis product curated at the British Atmospheric Data Centre. The open-source WMS client 'quickWMS'[15] was used to provide a browser GUI (the equivalent direct HTTP request is that of Figure 6). Figure 8 shows the same map overlaid with rainfall from another WMS source (globe.digitalearth.gov).

---

[14] The Geoserver project is the OGC reference WFS implementation, available at http://geoserver.sourceforge.net.
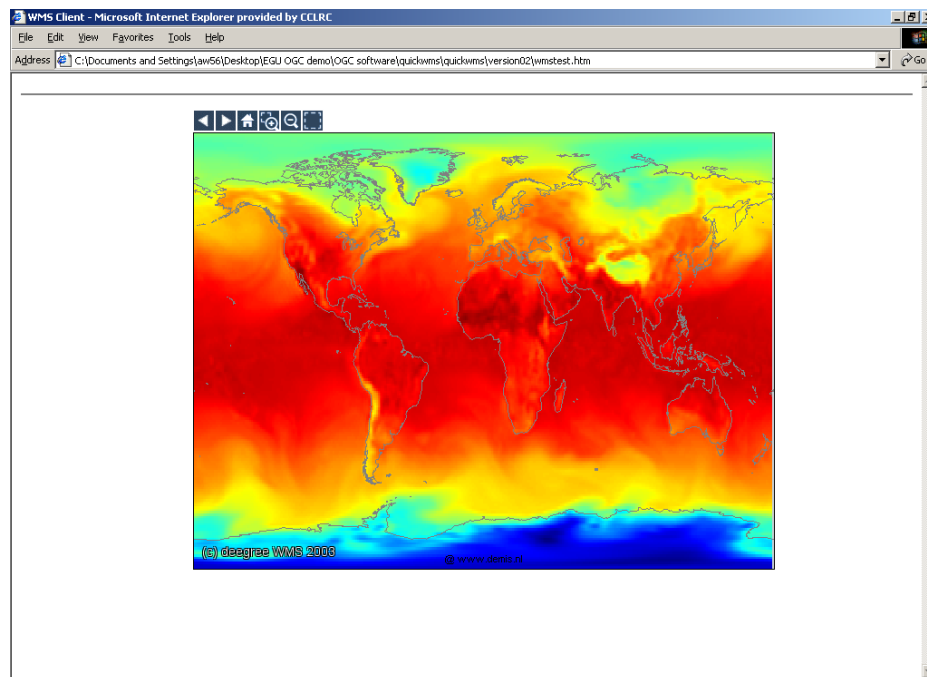[15] Available at http://www.inovagis.org/quickwms/.

Figure 7: WMS server modified for netCDF file store. ERA-40 surface air temperature for 27 April, 2001.
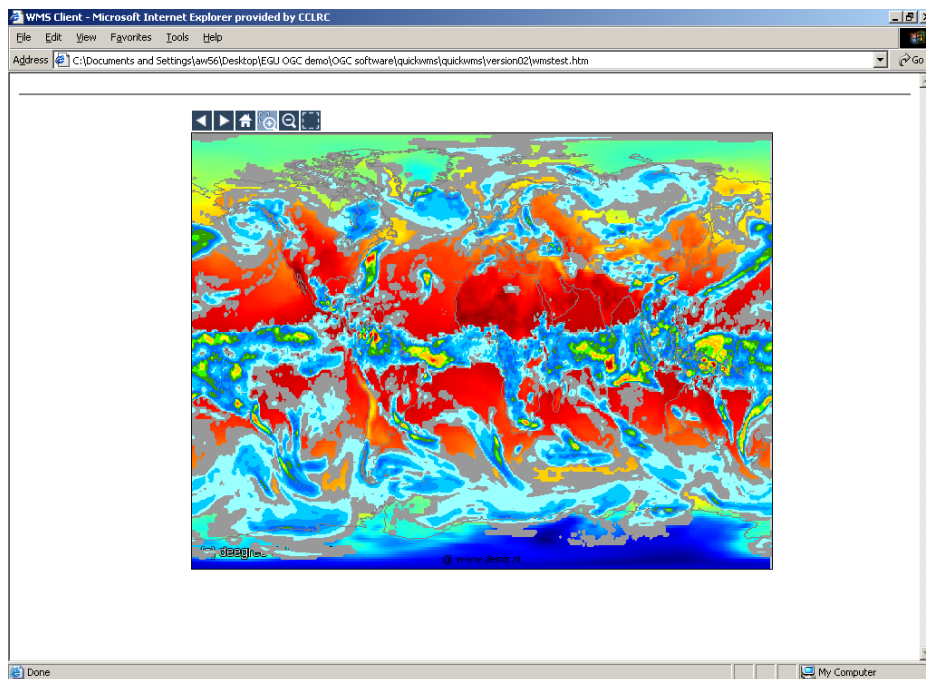

Figure 8: ERA40 surface air temperature overlaid with rainfall from different WMS server

4. W3C WEB SERVICES

The OGC implementation specifications described above are an example of internet-accessible spatial data services. As mentioned earlier, the ISO Architectural Reference Model [vii] envisages a comprehensive taxonomy of geographic information services [xxvi]. For example, human interaction services enable viewing and annotating of geographic information and catalogues, editing of processing chains, etc.; information management services provide access to data, registries, gazetteers etc.; processing services may perform coordinate transformations, geoprocessing etc. The means of invoking OGC services are specified explicitly in their respective Implementation Specifications, and include both HTTP GET and POST bindings. Lengthy details are provided for encoding keyword-value pairs, XML query filters etc., in service requests. A more general and scalable framework is needed if a full spectrum of services is to be realised.

In September 1999, a group of software vendors (including Microsoft Corporation) proposed a general XML message-based protocol for client-server interaction across a network. Known as the Simple Object Access Protocol (SOAP) [xxvii], it has become the cornerstone of what is now described as "service-oriented architectures". These are distributed processing systems where complex applications and processing chains may be assembled dynamically by discovering and orchestrating loosely-coupled networked services. The Web Services Description Language (WSDL) has been developed as a general method for describing service interfaces, locations and bindings. Finally, various mechanisms have been proposed for registration and discovery of services – for example, the UDDI (Universal Description, Discovery and Integration) and ebXML (Electronic Business using eXtensible Markup Language) specifications. The XML-based networked services enabled by these technologies are known as "web services". Figure 9 illustrates the well-known "publish-find-bind" pattern
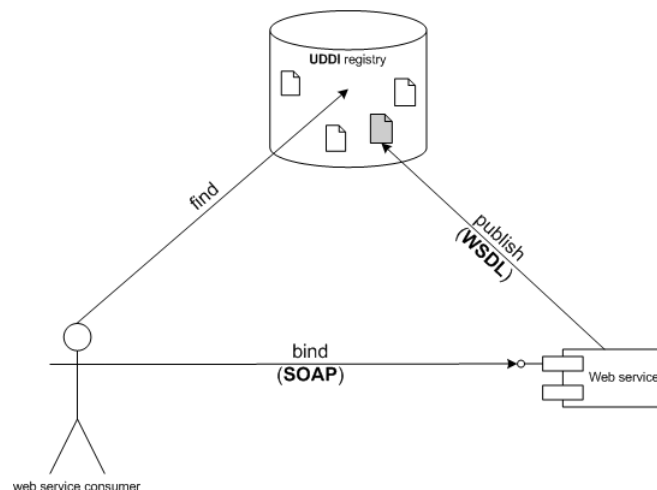

Figure 9: 'Publish-find-bind' pattern of web services

of web service architectures.

There has been a significant commitment to web services by all the major software vendors (e.g. Sun's J2EE, IBM's Websphere, and Microsoft's .NET all support web services).

We review below SOAP and WSDL – the two major components of web service technologies.

4.1 Simple Object Access Protocol (SOAP)

SOAP is an XML-based client-server messaging protocol for network services. Two types of messaging models are supported – "RPC-style" and "document-style". The first provides for a remote procedure call invocation, with specific operation parameters and return values. The second allows much greater flexibility, with arbitrary XML documents being sent to a server instead of specific operation parameters; an arbitrary XML document is returned as the result. We review here only RPC-style SOAP messaging.

Consider, for example, a hypothetical web service which provides an operation "getTemperature(lon,lat)" with two arguments – longitude and latitude – and returns the current temperature in degrees centigrade. Figure 10 illustrates the XML SOAP message that would be sent to this web service to find the temperature in London.

```
<s:Envelope xmlns:s="http://www.w3.org/2003/05/soap-envelope">
  <s:Body>
    <n:getTemperature xmlns:n="http://serviceprovider.example.com/TemperatureService">
      <n:lon>-0.1</n:lon>
      <n:lat>51.5</n:lat>
    </n:getTemperature>
  </s:Body>
</s:Envelope>
```

Figure 10: SOAP message to hypothetical geo-referenced temperature service

The XML SOAP response on a summer day might be as shown in Figure 11.

```
<s:Envelope xmlns:s="http://www.w3.org/2003/05/soap-envelope">
  <s:Body>
    <n:temperatureResponse xmlns:n="http://serviceprovider.example.com/TemperatureService">
      <n:temperature>23</n:temperature>
    </n:temperatureResponse>
  </s:Body>
</s:Envelope>
```

Figure 11: SOAP response to temperature request

In this case both the request and response parameters are simple numbers. More complex parameter types may be supplied using the rich data structures possible with XML.

Multiple bindings are possible for actually conveying a SOAP message to a service. Most common is to use HTTP as the delivery protocol. Bindings for both HTTP GET and HTTP POST are defined. However, SOAP messages may also be sent by email (using SMTP as the delivery protocol), for instance.

Finally, the example above shows a SOAP message containing only a 'Body' element. The general structure of a SOAP message (Figure 12) provides also for a 'Header' element to be supplied. This may be used for context-specific information, such as transaction identifiers,
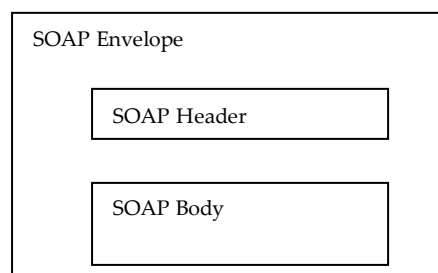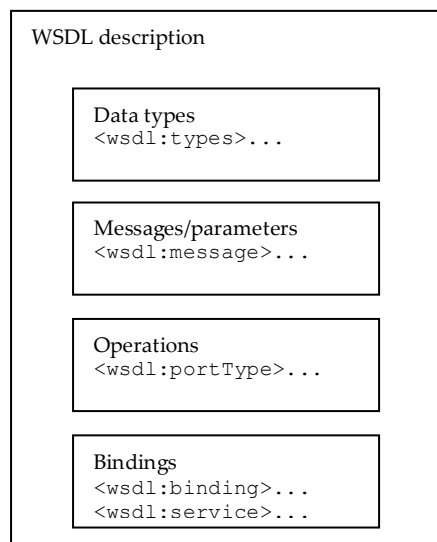

Figure 12: SOAP message structure

security information etc.

4.2 Web Service Description Language (WSDL)

In order to invoke a web service, a client must know what interfaces are provided, what parameters are required, what bindings are supported, and what the 'endpoint address' (URL service location) is. The Web Services Description Language (WSDL) provides a general XML language for describing arbitrary services in sufficient detail to enable client invocations. The

structure of a WSDL web service description is shown in Figure 13.

4.2.1 WSDL data types

The 'data types' section is used to define (with an XML schema) specialised or complex data types that may be used in a service. For instance, to restrict a latitude parameter to a valid range, a 'latitudeType' could be defined as in Figure 14.

```
<wsdl:types>
  <xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema">
    <xsd:simpleType name="latitudeType">
      <xsd:restriction base="xsd:double">
        <xsd:minInclusive value="-90"/>
        <xsd:maxInclusive value="90"/>
      </xsd:restriction>
    </xsd:simpleType>
  </xsd:schema>
</wsdl:types>
```

Figure 14: WSDL definition of 'latitudeType'

Figure 13: WSDL description structure

4.2.2 WSDL message definitions

The set of parameters supplied to a web service operation are defined through a WSDL message description. Figure 15 shows input and output messages for the example temperature service.

```
<wsdl:message name="getTemperatureInput">
  <wsdl:part name="lon" type="n:longitudeType">
  <wsdl:part name="lat" type="n:latitudeType">
</wsdl:message>

<wsdl:message name="getTemperatureOutput">
  <wsdl:part name="temperature" type="xsd:double">
</wsdl:message>
```

Figure 15: WSDL message types for temperature service

4.2.3 *WSDL interface definition*

The description of the operations provided by a web service is done through WSDL 'portTypes'. Each portType defines an interface with one or more operations. There may be multiple portTypes (interfaces) defined for a web service. Figure 16 shows the portType

```
<wsdl:portType name="TemperatureServiceInterface">
  <wsdl:operation name="getTemperature">
    <wsdl:input message="getTemperatureInput"/>
    <wsdl:output message="getTemperatureOutput"/>
  </wsdl:operation>
</wsdl:portType>
```

description for the temperature reporting service.

4.2.4 *WSDL bindings*

The final component of a web service description is the binding information – what transport protocol is required for the SOAP message, and which SOAP messaging model (RPC-style or document-style) is being used. Also needed is an 'endpoint address' (URL) to which SOAP

Figure 16: WSDL interface description for temperature service

messages may be sent. Figure 17 shows the binding information for a hypothetical instance of

```
<wsdl:binding name="TemperatureServiceBinding" type="TemperatureServiceInterface">
  <soap:binding style="rpc" transport="http://schemas.xmlsoap.org/soap/http"/>
  <wsdl:operation name="getTemperature">
    <soap:operation soapAction=""/>
    <wsdl:input>
      <soap:body use="encoded" encodingStyle="http://schemas.xmlsoap.org/soap/encoding"
namespace="http://serviceprovider.example.com"/>
    </wsdl:input>
    <wsdl:output>
      <soap:body use="encoded" encodingStyle="http://schemas.xmlsoap.org/soap/encoding"
namespace="http://serviceprovider.example.com"/>
    </wsdl:output>
  </wsdl:operation>
</wsdl:binding>

<wsdl:service name="TemperatureService">
  <wsdl:port binding="TemperatureServiceBinding" name="TemperatureServicePort">
    <soap:address location="
http://serviceprovider.example.com/TemperatureService:8080"/>
```

Figure 17: WSDL binding description for temperature service

the temperature service.

4.3 A web service for accessing gridded climate data

Woolf *et. al.* [xxviii] described a web service (GADS) for accessing climate data in Grid applications. The service has two operations:

- dataQuery(): used for metadata queries on available datasets, variables, and gridded spatiotemporal domains of variables

- dataRequest(): used for requesting a subset of data in a specified file format (netCDF, HDF, GRIB, raw binary)

The web service provides a number of benefits. Data is delivered in a user-requested format irrespective of the underlying file store; time-series' of files are aggregated by the service to appear as a single logical dataset; file 'metadata' is standardised (delivered data is CF-compliant [xxix] regardless of legacy variable names, axes orders etc.). Furthermore, since the service is deployed as a SOAP web service, invocation is platform- and language-neutral – SOAP clients may be written in any language.

The web service has recently been incorporated[16] into a search and rescue GIS package. The package runs a model to forecast drift location, and requires wind and surface ocean current data. Recent analysis fields from a remote UK Metoffice archive are imported dynamically into the package over some region of interest using the GADS web service. Figure 18 shows a screenshot of the package with overlaid vectors for the imported wind and current fields.
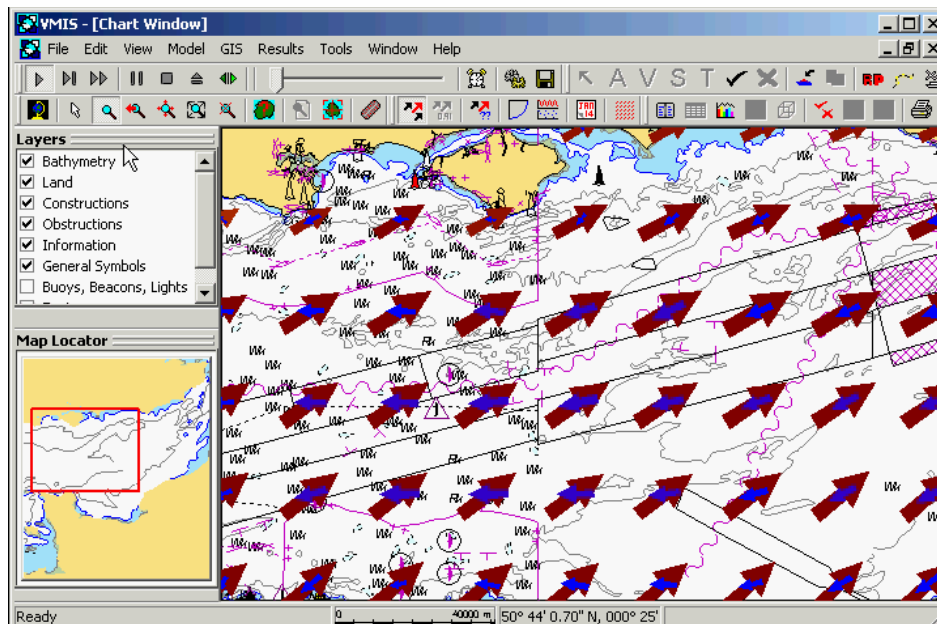

Figure 18: Search and rescue GIS using web service to access remote wind and surface current data (screenshot and SARIS package courtesy British Maritime Technology)

5. CLIMATE DATA CHALLENGES

The roadmap for climate data interoperability discussed in this paper has three components: ISO-compliant data models are constructed for atmospheric and oceanographic data types; legacy data sources are wrapped and exposed through data access web services; in particular, OGC's Web Map, Feature and Coverage Services are being widely deployed.

---

[16] This integration was done by British Maritime Technology and the Environmental Systems Science Centre, University of Reading, as part of the EDAS cluster (http://www.envdatacluster.net).

There exist a number of challenges with this program, however. We consider some of the issues here.

5.1 The vertical dimension

While the ISO standards and OGC specifications are explicitly intended to apply to geographic information in the broadest sense[17], the fact is that they derive from traditional GIS which operates predominantly in two-dimensions over the surface of the Earth. Atmospheric and oceanographic data, on the other hand, have a fundamental spatial component in the vertical. Indeed, a rich spectrum of vertical coordinate systems (see those in Table 2 for example) are used routinely in the climate sciences.

Table 2: Selection of vertical coordinate systems

Coordinate system	Vertical coordinate axis
Pressure	Static pressure
Isentropic	Entropy or potential temperature
Sigma	Normalised terrain-following
z	Lineal distance

ISO 19111 ("Spatial referencing by coordinates") provides for compound coordinate reference systems, where a conventional horizontal CRS is supplemented with a vertical CRS. The datum may be specified as 'vertical' in this case. GML expands this slightly to allow specification of a vertical datum as one of 'geoidal', 'depth' 'barometric' or 'othersurface'. Mechanisms for representing the full richness of vertical coordinates in these standards needs to be determined. It is certainly the case that these coordinate reference systems are not catalogued in standard geodetic tables (e.g. the EPSG tables that are supported by OGC services).

As well, OGC services must allow selection through the vertical dimension. This is an important requirement with gridded ocean or atmosphere simulation data. The current Web Map Server specification only provides for horizontal maps.

5.2 The time dimension

Time is also an important dimension for metocean data.

Numerical simulations are often based on a year of twelve 30-day months. ISO 19108 ("Temporal schema") provides for calendars to be defined on the basis of mappings to and from Julian dates. This may be a sufficient mechanism, but – as with the vertical dimension – it needs somewhere to be formally defined and referenced.

As with slices in the vertical, the OGC Web Map Server must be able to generate slices in time. Hovmuller diagrams showing climate fields on axes of latitude-time or longitude-time are

---

[17] The definition of "geographic information" in the ISO standards is "information concerning phenomena implicitly or explicitly associated with a location relative to the Earth".

very important diagnostics. Similarly, maps of vertical time series' (e.g. vertical radar backscatter, ocean temperature along a moored thermistor chain) are also commonplace. Enabling support for slices through time also suggests the prospect of time animations of vertical or horizontal slices through 4-D fields, again a common and important diagnostic in the climate-sciences.

5.3 Representation of grids

ISO 19123 ("Schema for coverage geometry and functions") provides for two types of gridded coverages – defined on either a 'rectified grid' or a 'referenceable grid'. The first assumes a regularly-spaced grid defined by an origin and linear combinations of offset vectors in each dimension. The second relies on an explicit transformation to be available to convert between grid coordinates and spatial coordinates. This model is reflected in GML which provides for either regularly spaced geo-referenced gridded coverages, or non-geo-referenced coverages.

Neither of these models is particularly suitable for output from finite-difference atmospheric and ocean general circulation models. It is often the case that the finite-difference grids are not regularly spaced in any standard coordinate reference system. For instance, increased resolution over an arbitrary region of the globe may be required. In ocean models, the north pole is often artificially relocated over land to avoid mathematical singularities. Fields from models formulated in spectral coordinates may be transformed to geographic coordinates, but often have fewer longitudinal gridpoints towards the poles.

Similar difficulties occur with remote-sensed imagery from satellites. This was recognised in the ISO Technical Report 19121 which identified a range of new work required "to address imagery and gridded data". ISO 19129 ("Imagery, Gridded and Coverage Data Framework") was instigated as a new work item, and is currently at a very early stage in its development. As this work progresses, it must handle grids used in the climate sciences.

5.4 Encoding efficiency

ISO 19118 specifies a normative XML encoding for geographic information. GML is a compliant implementation. Any XML encoding, however, is extremely inefficient for large data volumes. The question of encoding efficiency is recognised in GML for coverages, where the coverage rangeSet (gridded data values) may be encoded in a binary file. The mechanism needs to be supplemented with information on file format, for instance, but is a useful first step. There is no reason to assume in GML that coverage rangeSets are the only data for which encoding efficiency considerations apply. For instance, the grid difficulties referred to above require gridpoint locations to be specified explicitly in many cases – with a resulting domainSet for the coverage as large as the rangeSet.

The supported formats for the Web Coverage Service (GeoTIFF, HDF-EOS, DTED, NITF or GML) are not in common use in the climate sciences. To encourage ease of adoption, the widely-used netCDF format should be included with these.

5.5 Security

The OGC service specifications do not incorporate any security model, but, in common with many other domains, datasets in the climate sciences often need to have restricted access. For instance, national meteorological agencies typically sell their products, so delivery through unsecured OGC services is not feasible. A considerable amount of work is underway for an extensive framework of security for web services [xxx]. This framework would undoubtedly benefit OGC services if adopted in due course.

5.6 OGC web service compliance

As mentioned earlier, explicit and bespoke HTTP bindings are specified in the OGC specifications. This is not a scalable solution, nor does it place the OGC services within the rapidly-growing standard web services arena. SOAP bindings for OGC services should be defined, and WSDL service descriptions provided. This is being considered currently by the OGC Web Services Initiative Phase Two feasibility study[18].

6. SUMMARY AND CONCLUSIONS

We have reviewed a number of emerging standards that, together, offer a roadmap for significant data interoperability in the climate sciences. The ISO Technical Committee 211 is developing a range of standards for characterising geographic data and metadata. Conceptual data models may be defined, drawing on standardised schema for spatial and temporal referencing, geometry, topology, etc. Web services provide a platform from which to build data discovery, access, and processing services for geospatial data. The potential of web-enabled geographic information services is indicated already by the early success of the OGC specifications for Web Map, Feature and Coverage Services.

We have presented examples of these in the climate sciences. An ISO-compliant abstract data model was described for atmospheric and oceanographic data, and specialised to a set of feature types. A Web Map Server was implemented for numerical climate reanalysis data in netCDF files. And a SOAP web service for climate data was integrated into a GIS search and rescue model.

A number of challenges remain before the roadmap is realised. Some extensions to the ISO standards and GML are required – notably in relation to the vertical and time dimensions, and gridded datasets. Encoding efficiency mechanisms need to be expanded. And OGC services must evolve to full web service compliance, and incorporate access control mechanisms.

Finally, a broad governance framework for supporting community development of data models must be established. Interoperability will be hindered if there are multiple definitions of a feature type for radiosonde measurements, for instance. In practice, different communities will need to define and catalogue features at different levels of granularity. Ultimately, ISO-endorsed registries and Feature Type Catalogues will need to be established under the auspices of bodies such as the International Oceanographic Commission (IOC) and World Meteorological Organisation (WMO).

---

[i] US Executive Order 12906, "Coordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure" (1994).
[ii] IPCC, "Climate Change 2001: The Scientific Basis", Cambridge University Press (2001).

---

[18] http://www.opengis.org/initiatives/?iid=7

[iii] Adams, R.M., *et. al.*, "Value of Improved Long-Range Weather Information", Contemporary Economic Policy, **13**(3), 10-19 (1995).

[iv] Mjelde, J.W., *et. al.*, "A review of current evidence on climate forecasts and their economic effects in agriculture", American Journal of Agricultural Economics, **80**(5), 1089-1095 (1998).

[v] Katz, R.W., and A.H. Murphy (ed.), "Economic Value of Weather and Climate Forecasts", Cambridge University Press (1997).

[vi] Nebert, D.D. ed., "Developing Spatial Data Infrastructures: The SDI Cookbook", GSDI Association (2004) [Online: http://www.gsdi.org/docs2004/Cookbook/cookbookV2.0.pdf, 6 July, 2004].

[vii] ISO 19101, "Geographic information – Reference model".

[viii] Lawrence, B., *et. al.*, "The NERC DataGrid prototype", UK e-Science All-Hands Meeting, Nottingham (2003) [Online: http://www.nesc.ac.uk/events/ahm2003/AHMCD/pdf/065.pdf, 6 July, 2004].

[ix] ISO 19103, "Geographic information – Conceptual schema language".

[x] ISO 19109, "Geographic information – Rules for application schema".

[xi] Atkinson, R., "Next steps to interoperability – Mechanisms for semantic interoperability", EOGEO Workshop, University College London, 23-25 June 2004 [Online: http://y.eogeo.org/files/eogeo2004/eogeo-2004-full-atkinson.ppt, 6 July, 2004].

[xii] ISO 19108, "Geographic information – Temporal schema".

[xiii] ISO 19107, "Geographic information – Spatial schema".

[xiv] ISO 19113, "Geographic information – Quality principles".

[xv] ISO 19112, "Geographic information – Spatial referencing by geographic identifiers".

[xvi] ISO 19115, "Geographic information – Metadata".

[xvii] ISO 19110, "Geographic information – Methodology for feature cataloguing".

[xviii] ISO 19135, "Geographic information – Procedures for registration of geographical information items".

[xix] ISO 19118, "Geographic information – Encoding".

[xx] ISO 19136, "Geographic information – Geography Markup Language".

[xxi] Woolf, A., *et. al.*, "Data virtualisation in the NERC DataGrid", UK e-Science All Hands Meeting, Nottingham [Online: http://www.nesc.ac.uk/events/ahm2003/AHMCD/pdf/094.pdf, 6 July, 2004].

[xxii] Cox, S., *et. al., ed.*, "OpenGIS Geography Markup Language (GML) Implementation Specification", Version 3.00, OpenGIS Consortium document 02-023r4 (2003) [Online: http://www.opengis.org/docs/02-023r4.pdf, 6 July, 2004].

[xxiii] de La Beaujardière, J., *ed.*, "Web Map Service Implementation Specification", OpenGIS Consortium document 01-068r3 (2002) [Online: http://www.opengis.org/docs/01-068r3.pdf, 6 July, 2004].

[xxiv] Vretanos, P.A., *ed.*, "Web Feature Service Implementation Specification", OpenGIS Consortium document 02-058 (2002) [Online: http://www.opengis.org/docs/02-058.pdf, 6 July, 2004].

[xxv] Evans, J.D., *ed.*, "Web Coverage Service (WCS), Version 1.0.0", OpenGIS Consortium document 03-065r6 (2003) [Online: http://www.opengis.org/docs/03-065r6.pdf, 6 July, 2004].

[xxvi] ISO 19119, "Geographic information – Services".

[xxvii] Snell, J., *et. al.*, "Programming Web Services with SOAP", O'Reilly & Associates, Inc. (2002).

[xxviii] Woolf, A., *et. al.*, "A Web service model for climate data access on the Grid", International Journal of High Performance Computing Applications, **17**(3), 281-295 (2003).

[xxix] Eaton, B.¸ *et. al.*, "NetCDF Climate and Forecast (CF) Metadata Conventions" (2003) [Online: http://www.cgd.ucar.edu/cms/eaton/cf-metadata/CF-1.0.html, 6 July, 2004].

[xxx] IBM and Microsoft, "Security in a Web Services World: A Proposed Architecture and Roadmap" (2002) [Online: http://www-106.ibm.com/developerworks/webservices/library/ws-secmap/, 6 July, 2004].