# JISC DEVELOPMENT PROGRAMMES

## Project Document Cover Sheet

**FINAL REPORT**

### Project

<table>
<thead>
<tr>
<th>Project Acronym</th>
<th>GRADE</th>
<th>Project ID</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Project Title</strong></td>
<td>Scoping a Geospatial Repository for Academic Deposit and Extraction</td>
<td></td>
</tr>
<tr>
<td><strong>Start Date</strong></td>
<td>1 June 2005</td>
<td><strong>End Date</strong></td>
</tr>
<tr>
<td><strong>Lead Institution</strong></td>
<td>EDINA</td>
<td><strong>Project Director</strong></td>
</tr>
<tr>
<td><strong>Project Manager &amp; contact details</strong></td>
<td>Anne Robertson EDINA The University of Edinburgh Causewayside House 160 Causewayside Edinburgh EH9 1PR tel: 0131 651 3874 email: <a href="mailto:a.m.robertson@ed.ac.uk">a.m.robertson@ed.ac.uk</a></td>
<td></td>
</tr>
<tr>
<td><strong>Partner Institutions</strong></td>
<td>University of Edinburgh, University of Southampton</td>
<td></td>
</tr>
<tr>
<td><strong>Project Web URL</strong></td>
<td><a href="http://www.edina.ac.uk/projects/grade">www.edina.ac.uk/projects/grade</a></td>
<td></td>
</tr>
<tr>
<td><strong>Programme Name (and number)</strong></td>
<td>Digital Repositories Programme 2005-07</td>
<td></td>
</tr>
<tr>
<td><strong>Programme Manager</strong></td>
<td>Neil Jacobs</td>
<td></td>
</tr>
</tbody>
</table>

### Document

<table>
<thead>
<tr>
<th>Document Title</th>
<th>Geospatial Data Repositories – Interoperability Issues</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reporting Period</strong></td>
<td>End of project</td>
</tr>
<tr>
<td><strong>Author(s) &amp; project role</strong></td>
<td>Dr Pragya Agarwal, Lecturer in GIScience, University College London Anne Robertson, Project Manager James Reid, Project Advisor</td>
</tr>
<tr>
<td><strong>Date</strong></td>
<td>January 2008</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://edina.ac.uk/projects/grade/GeoDataRepositoryInteroperability.pdf">http://edina.ac.uk/projects/grade/GeoDataRepositoryInteroperability.pdf</a></td>
</tr>
<tr>
<td><strong>Access</strong></td>
<td>☑ Project and JISC internal ☑ General dissemination</td>
</tr>
</tbody>
</table>
## Document History

<table>
<thead>
<tr>
<th>Version</th>
<th>Date</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>30/03/07</td>
<td>Initial draft for Programme Manager</td>
</tr>
<tr>
<td>1.0</td>
<td>31/10/07</td>
<td>Final draft for review</td>
</tr>
<tr>
<td>1.1</td>
<td>08/02/08</td>
<td>Final version for submission</td>
</tr>
</tbody>
</table>
Introduction
The GRADE Project is one of a cluster of projects in the Digital Repositories Programme funded by the Joint Information Services Committee of HEFCE investigating the interactions between data and institutional (publications) repositories, support for scientific lifecycle, storage and access requirements.

The JISC is bringing together a programme of work relating to digital repositories. Its aim is to bring together people and practices from across various domains (research, learning, information services, institutional policy, management and administration, records management, and so on) to ensure the maximum degree of coordination in the development of digital repositories, in terms of their technical and social (including business) aspects.

Within this context, GRADE is investigating the technical and cultural issues around the reuse of geospatial data within the JISC IE in the context of media–centric, informal and institutional repositories.

GRADE Work Package 5 has the following objectives:

- To investigate and assess interoperability between geospatial data repositories and other types of services/repositories e.g. metadata harvesting into geo-spatial portals, linking e-prints with the datasets referred to in the articles.
- To investigate the linking of geospatial repositories with e-Science infrastructures.
- To investigate and assess the potential of evolving industry driven geospatial interoperability standards, specifically Open Geospatial Consortium/ISO 19100 series standards, as a means of interoperating with repositories within and outside of academia.

Background
The primary rationale for this Work Package is to investigate how media-centric repositories might interoperate with external repositories and how interaction with the external environment (essentially the JISC IE) may best be achieved. This is an essential aspect as not only must a repository be capable of managing data and metadata internally but it must also be capable of providing some means by which these assets can be accessed, communicated, reported and packaged to other repository and non-repository settings e.g. for ‘deep-citation’ (Altman 2001)\(^1\).

Components of the JISC IE portals and shared services architecture, specifically Go-Geo! (www.gogeo.ac.uk) and GeoCrossWalk (http://www.geoxwalk.ac.uk/) are relevant in this context. The role of

---

GeoCrossWalk is to support and enhance resource discovery/exploitation by ‘geo-enabling’ new and legacy resources within the JISC IE. Go-Geo!, a portal for the discovery of geospatial data and related resources, would be the obvious point for metadata from geospatial repositories to surface for searching by end users and cross searching by other JISC portals e.g. Intute\(^2\).

The growth of web services and interest in the e-Science Grid computing phenomena, allied with burgeoning numbers of interoperability standards\(^3\) within the wider (common) information environment means that there is something of an embarrassment of riches with regard to choices and the key question to emerge is which standard(s) is (are) optimal within proscribed contexts? It is unlikely (and unreasonable to presume) that a single standard can accommodate all requirements and thus it is necessary to evaluate and assess potential candidates for fitness of purpose. Within the world of geospatial content providers and consumers, interoperability has long been an issue and a raft of interoperability standards have been developed to cover the interoperability of both metadata and data. The same is true of the Grid community. This work package investigates how these standards and protocols might be adopted and attempts to make recommendations as to how domain specific, independently evolved standards relate to those derived from separate domain specialisms.

**Methodology**

This report aims to report on the interoperability aspects of geospatial repositories with specific reference to the GRADE project by providing a SWOT assessment of interoperability aspects of geospatial data repositories then a detailed discussion of the linkages between geospatial repositories, components of the JISC IE and e-Science infrastructures. The primary consideration in this case is that geospatial data has its own specific formatting and managing constraints and limitations, which raises specific issues for interoperability of data repositories that are specifically aimed at storing and handling geospatial data.

**I. Geospatial data repositories**

The distinction between a repository and an archive is vague, partly because the digital library, digital preservation, archiving and other communities often interpret these terms differently, although both things are normally constructed with a goal of ensuring long-term (often in the order of decades) access to the materials within them. JISC defines a repository simply as “deposited content collections.” According to the February 2005 JISC Digital

---

2 http://www.intute.ac.uk/sciences/
3 http://standards.jisc.ac.uk
 Repositories Review\(^4\) ‘a digital repository is differentiated from other digital collections by the following characteristics:

- content is deposited in a repository, by content creator, owner or 3rd party,
- repository architecture manages content as well as metadata,
- repository offers minimum basic services - put, get, search,
- repository must be sustainable and trusted in terms of being well-supported and well-managed.\(^4\)

Geospatial data brings its unique problems to creating a data repository. The first issue is that GI data has proprietary formats and there are not always published guidelines on how to handle these. Shared data requires interoperability ("working together") among the software systems that provide geospatial data, maps, services, and user applications.

An efficient, usable and effective geospatial repository therefore should:

- Improve the quality of research data
- Provide access to reliable working data
- Allow conclusions to be validated externally
- Apply good record-keeping standards to data capture including in lab and field electronic notebooks, which enables scientists to draw conclusions from reliable and trustworthy working data
- Enable large amounts of data to be analysed and developed across different locations by maintaining consistency in working practices and interpretations
- Manage relationships between different versions of dynamic or evolving datasets, and facilitates linkage with other related research and between primary, secondary and tertiary data
- Allow data sets to be combined in new and innovative ways, e.g. historic biodiversity data and GIS data can be combined to investigate trends in ecosystem development

There are a few existing projects, that describe themselves as geospatial repositories or archives. Large government institutions such as NASA\(^5\), NOAA\(^6\), NERC\(^7\) and ESA\(^8\) host decades-old data that serve many of the functions associated with a repository, such as providing services for data preservation, discovery and retrieval.

\(^5\) www.nasa.gov
\(^6\) www.noaa.gov
\(^7\) www.nerc.ac.uk
\(^8\) www.esa.int
Interoperability Aspects of Geospatial Data Repositories

Geospatial interoperability is based on shared agreements (that is, voluntary consensus standards) governing essential geospatial concepts and their embodiment in communication protocols, software interfaces, and data formats. Many repositories are centralised where all data is stored on a central server and retrieved from there. However, many of the data sharing resources over the web make use of distributed services, providing a common platform and interface to search and access the diverse and distributed data from. The centralised data services are easier to maintain and interoperability is less of a concern with better control over the nature of data deposited, metadata entries and storage, and the technical issues related to interface and architecture of the service. In the case of a distributed repository, the issues with technical incompatibilities are much more pronounced, as different services may follow different standards, rules, guidelines and formats. There are more semantic variations as well, since the different services are designed for different purposes. Mostly the data available on the data repositories have been created for a specific purpose, which means that the data creators have abstracted the geospatial reality in a specific way.

Interoperability of data has become a necessity for sharing and integrating datasets available over centralised or local and regional data repositories. This causes problems when the users want to join the data from different sources. There can be syntactic, structural, semantic, temporal and geometric heterogeneities between different data sources. From an interoperability perspective, non-exhaustive issues of concern for geospatial data repositories are:

- Documentation of data – data has to be documented to a level that will allow confident reuse (the UK Academic profile\(^9\) of the international geospatial metadata standard ISO19115\(^10\) provides academic users the necessary information to reveal dataset type, its location, quality and completeness, temporal aspects in terms of data collection and content and contact details\(^11\))
- Documentation of data – metadata required for confident reuse can be far more extensive than the Dublin Core\(^12\)-based metadata required for repository deposit
- Which sector to interoperate with – is it with the geographic information community across the board or academia or both – this will influence the appropriate metadata standards
- Automated metadata generation, maintenance and upkeep of metadata for different versions
- Machine-readable metadata generation
- Managing different versions of data - often geospatial data exists in different formats, compatible with different software versions and

---

\(^9\) [http://www.gogeo.ac.uk/files/UKAGMAP%20version1%20with%20intro.pdf](http://www.gogeo.ac.uk/files/UKAGMAP%20version1%20with%20intro.pdf)

\(^10\) [http://www.isotc211.org/Outreach/Overview/Factsheet_19115.pdf](http://www.isotc211.org/Outreach/Overview/Factsheet_19115.pdf)

\(^11\) The UK Academic Profile of ISO19115

\(^12\) [http://dublincore.org/documents/dcmi-terms/](http://dublincore.org/documents/dcmi-terms/)
formats that it was created and managed in and often different versions have inherent incompatibilities

- Reusability of data - data has to be reusable and results replicated across different domains and contexts and multiple applications
- Managing file names for easy access and retrieval and reference - in a shared environment, file names enable easy access and retrieval of most relevant data and also for managing multiple versions of same data. No existing standards exist for file names\(^\text{13}\).
- Managing content in suitable categories for discoverability
- Managing different frames of references for the data - domain specific, co-ordinate and grid system
- Preservation of different forms of spatial data - versioning control and curation of spatial data
- Validation of data - semi or fully automated methods
- Issues with trust - open or closed communities
- Issues with IP rights, licensing and ownership over deposited datasets\(^\text{14}\)
- Issues with privacy of data and institutional/academic vs commercial uses of certain datasets
- How to package data - data exists not in one file but several related files
- Annotation of data for easy access - requires human interpretation or automated
- Identifiers for data and;
- How to provide appropriate information retrieval methods?

From an interoperability viewpoint, a significant amount of research is required on the issues related to general navigational interfaces, which are decoupled from a specific GIS viewpoint. Another issue of particular significance is the access to very large information repositories. Issues with naming datasets, indexing and categorising datasets have either been resolved by using controlled vocabularies and thesauri. However, these often do not encompass complete semantic ranges and therefore retrieval and access of data is not always user-friendly. The support for the use of ontologies as opposed to controlled vocabularies is relevant here - in this case, the question is how these ontologies can be created and managed. Issues of learnability and scalability also need to be addressed to using schemas and community-defined set of words for referencing data. Informal methods such as wikis and peer to peer software can also be used to either exchange information about data or to share data itself. In this case, technical problems related to software compatibility as well as linguistic, semantic, trust and privacy issues are pertinent.

\(^\text{13}\) However this was a user requirement identified within WP1 formal demonstrator, http://edina.ac.uk/projects/grade/FormalGISRepositoryFeedbackFinal.pdf
Most interoperability issues are raised by terminological and semantic differences, organisation of datasets, and technical and version issues for different formats of data available, and are about the reliability and mismatch of data due to different recording dates. The Open Geospatial Consortium\textsuperscript{15} already has proven working standards. In order for a range of web services from a variety of sources to be able to work together, they must conform with a set of technical standards to enable them to ‘talk to each other’. At a simplistic level this is analogous to having a standard type of electric plug across the UK. Typical and generic data types range from observational data, to large-scale experimental data, simulation, modelling, and design. Other issues include active vs passive manipulation of data by the users, and visualisation of data and accessing suitability of data before downloading.

Table 1 summarises key interoperable issues for geospatial data repositories. Other considerations for scientists and researchers looking to deposit data and share data in a collaboratory environment would be to:

- Use Open Source Software and Open Standards to facilitate exchange and persistence of data through and across different systems
- Ensure good annotation and creation of metadata to enable re-use of data
- Ensure primary, secondary, and tertiary levels of research materials are linked and that links are persistent
- Use unique and persistent identifiers and a consistent citation format

Related work within the GRADE project identified a series of repository functions that a geospatial data repository should offer\textsuperscript{16}. An interoperable repository environment therefore would have to support the following functions:

- The ability to publish data and metadata (via OAI) in a variety of modes (open, closed and limited to defined peer groups).
- Searching across repositories built upon Dublin Core and AGMAP metadata standards – leads to semantic issues
- Data access – with authorisation checks on eligibility and digital rights
- Repository alerts - on new deposits or changes to existing deposits
- Linking – to enable associations between several datasets and associations between datasets and publications
- Data transformations - from one geospatial data format to another and possibly from one coordinate system to another.

\textsuperscript{15} The Open Geospatial Consortium, Inc (OGC) is an international industry consortium of over 350 companies, government agencies and universities participating in a consensus process to develop publicly available interface specifications. OpenGIS® Specifications support interoperable solutions that "geo-enable" the Web, wireless and location-based services, and mainstream IT. The specifications empower technology developers to make complex spatial information and services accessible and useful with all kinds of applications.

\textsuperscript{16} http://edina.ac.uk/projects/grade/InformalGISRepositoryFeedbackFinal.pdf
<table>
<thead>
<tr>
<th>Strengths</th>
<th>Weaknesses</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Existence of ISO series of standards (ISO 19100 series) for geospatial</td>
<td>• Current climate of multiple popular data formats</td>
</tr>
<tr>
<td>data representation</td>
<td>• Lack of agreed standard for geospatial data preservation</td>
</tr>
<tr>
<td>• Existence of ISO Technical Committee (ISO/TC 211 Geographic Information)</td>
<td>• Lack of standard for content packaging of geospatial data</td>
</tr>
<tr>
<td>with responsibility for ISO’s geographic information series of standards.</td>
<td>• No shared definition of a geospatial or geospatial data repository</td>
</tr>
<tr>
<td>• Existence of international standards-making body (OGC) with x-sector</td>
<td>(kind of objects, what formats, standards)</td>
</tr>
<tr>
<td>membership (including academia)</td>
<td></td>
</tr>
<tr>
<td>• Mature metadata standard for documenting geospatial data</td>
<td></td>
</tr>
<tr>
<td>• Number of web service based interoperable standards for data</td>
<td></td>
</tr>
<tr>
<td>dissemination commonly used within GI community(^\text{17})</td>
<td></td>
</tr>
<tr>
<td>• Current climate of multiple popular data formats</td>
<td></td>
</tr>
<tr>
<td>• Lack of agreed standard for geospatial data preservation</td>
<td></td>
</tr>
<tr>
<td>• Lack of standard for content packaging of geospatial data</td>
<td></td>
</tr>
<tr>
<td>• No shared definition of a geospatial or geospatial data repository</td>
<td></td>
</tr>
<tr>
<td>(kind of objects, what formats, standards)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Opportunities</th>
<th>Threats</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Opportunity for GI industry to leverage expertise from eLibrary world</td>
<td>• Dominance by one vendor on format</td>
</tr>
<tr>
<td>on preservation issues (especially standards for content packaging)</td>
<td>• Consensus-making process</td>
</tr>
<tr>
<td>• To implement OGC specifications and standards (especially progressive</td>
<td>• Interoperable specifications too complex to implement</td>
</tr>
<tr>
<td>approach of making data available as web service including feature,</td>
<td>• Education requirements not met (e.g. no consideration made of</td>
</tr>
<tr>
<td>mapping and coverage services).</td>
<td>citation)</td>
</tr>
<tr>
<td>• To advance progress towards an academic spatial data infrastructure</td>
<td>• Continual assessment of alignment with JISC IE</td>
</tr>
<tr>
<td>• Opportunity to implement semantically rich data repository</td>
<td>• Future data preservation standards unable to store all</td>
</tr>
<tr>
<td>• Opportunity to ensure UK institutions meet EU Directive(^\text{18})</td>
<td>attribute/topographic detail and therefore not taken up by community</td>
</tr>
<tr>
<td>requirements</td>
<td>• GI community inherently poor at metadata creation/documenting data</td>
</tr>
</tbody>
</table>

Table 1: SWOT analysis of Interoperability Aspects of Geospatial Data Repositories

\(^\text{17}\) For example, Web Mapping Service, Web Feature Service and Web Coverage Service. Making data available via these means was something that was reported by users of the demonstrator repository within work package 1 as desirable and reported within work package 2 as a means of making data available for reuse by others.

II. Relation between JISC IE, geospatial repositories, and e-science programmes

The vision of the UK e-science programme of a common e-infrastructure in which resources of storage and computation are connected by high bandwidth networks, and ‘middleware’ in each of these resources allows the creation of virtual computing environments that cross organisational boundaries, and this in turn enables collaborative research (e-science or more accurately, e-research), engineering, and public services\(^\text{19}\). The sharing and collaboration of resources is at the centre of e-science research and activities. In this way, there is close synergy between what GRADE and e-science activities are trying to achieve. The most commonly expressed goal of the grid infrastructures is to underpin collaborative e-research. As the volume of digital data grows, so new research is increasingly dependent upon effective storage, management and retrieval methods and so there is increasing recognition of the need to develop links between virtual learning and research environments, digital libraries/data repositories, methods of data curation and grids.

The e-science efforts in developing and maintaining a stable GRID environment provide opportunities for the geospatial community to share large datasets, resources, operations and capacities across organisations and nodes on the GRID. This enables more effective sharing of data in a trusted and secure environment. This has implications on the way that geospatial repositories can be structured and managed in a distributed environment and effective implementation of standards ensured. This will also allow for connections on a larger European and/or multinational GRID for uploading, managing and sharing data. Therefore the e-science activities are critical in providing a platform for management of a centralised geospatial repository and ensuring standardised development. Nevertheless, as compared to a centralised geospatial repository, this will require major organisational effort, as well as technological expertise to provide the networking, the management of resources on each site, and the support of users of this e-infrastructure.

The e-infrastructure is being developed to provide a foundation for information creation (e-research), data curation and discovery (digital libraries/data repositories and services that expose and virtualise these), and e-learning. This fits perfectly with the needs of the geospatial community in providing an infrastructure for shared resources.

Thus, there is a need to look at the specific needs of the geospatial community and explore the application of GRID to geospatial repositories. GRADE provides a valuable resource and avenue for making this possible.

\(^{19}\) http://www.rcuk.ac.uk/escience/default.htm
Both e-infrastructures (as encouraged and established by e-science initiatives) and geospatial repositories have a lot in common. Both deal with massive databases, complex modelling, and based in collaboration and shared resources. One example of this is the Solid Earth and Environmental GRID project\(^{20}\), which deals with geospatial data for mineral exploration spread across 8 government departments and the primary need is to provide integrated access to current data and archives, but also to maintain an integrated skills base as these are also dispersed.

Geospatial repositories therefore provide an effective demonstration of e-science principles by putting e-science methods of ‘collaborative’ environments into practice in data and compute intensive settings.

The JISC IE (Information Environment) aims to be an environment that will enable convenient access to a comprehensive collection of scholarly and educational materials held at UK academic institutions.

The JISC IE also outlines a common e-infrastructure that will help facilitate learning, teaching research and administration (see figure 3). This proposed architecture for integrating digital libraries/data repositories, portals, virtual learning environments (VLEs), shared services and other academic services must be considered in parallel with the UK e-science programme in the context of developments in centralised geospatial repositories.

As interoperability in semantics and metadata progresses, so services in both the UK e-science programme and the JISC IE will be more easily accessible to a widening spectrum of academic (and perhaps non-academic) users, both from portals and GIS packages, and more importantly from a broadening

\(^{20}\) https://www.seegrid.csiro.au/twiki/bin/view/Main/WebHome

\(^{21}\) Andy Powell UKOLN, University of Bath, 2003
spectrum of software services invoked by the users. The aim of the Information Environment is to help provide convenient access to resources for research and learning through the use of resource discovery and resource management tools and the development of better services and practice. The Information Environment aims to allow discovery, access and use of resources for research and learning irrespective of their location.

In this way, there are close synergies between JISC IE and the geospatial repositories with the JISC IE providing standards for development and management of geospatial repositories and the development of geospatial repositories can provide a suitable and innovative application platform for the JISC IE principles.

Moreover, the JISC IE strategy is to provide a national and global networked environment, and therefore would provide the geospatial community an opportunity to identify stakeholders with whom to actively engage and collaborate. The Information Environment is clearly key to the goals of achieving an interoperable distributed national electronic resource for geographically referenced data.

GRADE could effectively operate under the wider JISC IE umbrella, especially as a metadata provider. For example in work package 122, it was demonstrated that GRADE could be cross searched by the pilot service set up by the Pilot Engineering Repository Xsearch (PerX) project23. A critical requirement for such an engagement would also be the use of open source software and standards for geospatial data repositories which will enable more seamless integration across different resources and data providers. Some kind of reference model will be required to allow for compatibility and integration across these different resources. The primary geospatial topics will have to direct this model’s primary organising structure. In choosing standards, the first decision is whether the intended activities involve Data or Data Access, Metadata or Catalog Access, Maps or Visualization, Spatial Reference Systems, or other Geoprocessing Services. Two different model specifications are generally prescribed:

- **Abstract models** specify what information or requests are valid in principle, irrespective of individual computing environments. They define essential concepts, vocabulary, and structure (type hierarchy) of geospatial services and information transfer. These models set the stage for creating implementable specifications, and for extending existing ones to new environments.

- **Implementation specifications** tell software developers how to express information or requests within a particular distributed computing environment (e.g., World Wide Web, CORBA, .NET). Implementation specifications generally include access protocols, object models, and naming conventions.

---

22 http://edina.ac.uk/projects/grade/FormalGISRepositoryFeedbackFinal.pdf
23 http://www.icbl.hw.ac.uk/perx/index.htm
Such specifications are specific to, and directly usable within, their target computing environment.

For distributed computing as in a JISC IE or the e-science framework, the service and information viewpoints are crucial and intertwined. For instance, information content is not useful without services to transmit and use it. Conversely, invoking a service effectively requires that its underlying information be available and its meaning clear. Hence the underlying semantic of the data is important for effective retrieval of appropriate information within such frameworks.

The primary challenges for interoperability in geospatial repositories and across the JISC IE and e-science GRID frameworks can therefore be summarised as follows:

• Standards- effective regulation and monitoring of standards
• Data and data access- massive databases cause a number of problems in management and access of data as well as effective distribution of resources
• Metadata and catalogue access- metadata standards and harvesting is absolutely critical for the management of geospatial data and for development of an effective and efficient geospatial repository
• Maps and visualisation- visualisation of large datasets requires massive computing resources
• Geographic reference systems- reference frameworks for datasets vary across domains and regions. One of the primary challenges for an integrated geospatial repository will be to align the different reference frameworks
• Geoprocessing services- Although there have been some initiatives, primarily led by the OGC to define standards for geoprocessing services, mainly those over the web, these standards are not enforceable and/or legally binding. It is important to engage the OGC activities within development of geospatial repositories to ensure collaboration between various services forming part of the distributed framework.
Conclusions
The primary challenge for media repositories operating within the JISC IE is how to align their domain standards with those standards underpinning the JISC IE. Media repositories will need to find a way to adhere to their industry standards while implementing the appropriate digital library interfaces so interoperability with other services within the JISC IE continues. In particular, metadata to enable the confident reuse of geospatial data (such as AGMAP) is more complex than Dublin Core-based repository deposit data. If this more detailed metadata is to be packaged with the data item itself, attention needs to be given to the development of content packaging standards for geospatial data. This is particularly pertinent because geospatial data are often not made up of one single file but are a bundle of associated files. Add to this the fact that many researchers will create a series of data sets for one geographic location with associated visualisation/rendering information and wish to deposit that series of data as one repository item, then the need for a packaging standard for geospatial data is clear.

Recommendations

- If Dublin Core is used as the foundation for repository interoperability, mappings between DC onto OGC metadata profiles should be developed.
- Outstanding issues relating to security and IPR are urgently required to be resolved. In the security context, e-Science approaches may offer some solutions.
- Allied to the above, existing licensing frameworks for geospatial resources within an interoperable repository environment need reviewed and engagement with data providers is required.
- OGC and ISO standards should be adopted by repository developers to assist in interoperability between repositories, components of the JISC IE, e-Infrastructure and for purposes of internationalisation.
- Effort should be placed on developing content packaging standards for geospatial data.