REVIEW OF CURRENT BEST PRACTICES FOR THE LONG TERM ARCHIVING, DATA STANDARDS, AND ACCESS TO DIGITAL ARCHAEOLOGICAL LANDSCAPE DATA

INTRODUCTION

The aim of this work package was to develop a data management strategy for the *Share-IT* project. To accomplish this objective, a review of current Irish and International best practice was undertaken, addressing issues relating to archiving, data standards and access of digital archaeological landscape data.

ARCHIVING REVIEW

a) Some preliminary observations came from our review of best practice which emphasises the nature and scale of the problem:-

- there has been a rapid growth in the creation of digital data
- it highlighted the speed and ease of short-term data dissemination with little regard for the long-term preservation of digital data.
- digital data is fragile in ways that differ from traditional technologies, more easily corrupted
- digital storage media have shorter life spans
- technological advances are rapid, therefore the time frame in which we must consider archiving becomes much shorter
- b)There are a number of issues relating to the technology and specifically changes in technology associated with digital data:-
- **Obsolescence** - file formats, media, software and hardware, all vulnerable
- **Deterioration** - Standard media used for the storage of digital data –have a finite lifespan and can become corrupted
- **Loss of expertise** -digital data without adequate documentation, and sensible naming strategy relies on existing staff to understand it

c) Properly archived data will open up interdisciplinary re-use opportunities and maximise the return from investing in data.

d) A number of common themes and issues came out when reviewing best practice:-

- The fundamental importance of metadata and the adoption of international standards
- The need to define an access constraint policy
- The financial implications of archiving digital data, and the need to consider detailed cost models
- The OAIS (open archival information system) model provides the framework and terminology for defining an archiving strategy.

DATA FORMATS

The data being considered by share-IT project is limited to three data types, LiDAR, orthoimagery, and geophysical survey.

A key data preservation issue is which file format is selected as the archival version, and is critical to the longevity and future access to the data. Accepting the value of the OAIS reference model it is simplest to consider the appropriate file formats in terms of the three information packages (Submission, Archival and Dissemination). The archival information package is the version which will be held in perpetuity, and as such need to be in a standard non-proprietary format such as ASCII. The choice of this format is critical as the submission format must be able to migrate into it, and the dissemination format be generated from it. A fundamental component of the archival information package is the xml file containing the metadata.

ARCHIVE PROCESSES

A core component of an OAIS compliant archive is Archival Storage. This represents the part of the archival system that manages the storage and maintenance of digital objects entrusted to the archive. It ensures the appropriate structure of the file system, the necessary amount of storage available and other issues related to the physical management of data storage.

RELATED ISSUES RAISED INCLUDE:-

- Encouraging Data Submission - how will we encourage submission of data, and what are the appropriate leverage mechanisms? Contractual obligation? Legislation? Voluntary?
- Copyright – the need to prepare 'Copyright and Liablilty Statement'
- Access – define the rules of access, levels of access based on status? User logins?
- Promotion of the digital archive – little point in preserving data unless it is re-used.
- Cost – need to define a cost model based on the digital archive lifecycle.

METADATA & ISO STANDARDS

Metadata is often described as 'data about data' [a]nd is an integral part of the OAIS model. International standards exist for defining metadata schemas, which are typically stored as xml documents. Important standards related to geospatial data include:-

- Dublin Core - a standard for cross-domain information resource description. A vocabulary of fifteen properties for use in resource description. (ISO 15836:2003)
- ISO 19115 defines the schema for describing geographical information and associated services, including contents, spatial-temporal purchases, data quality, access and rights to use. The standard defines more than 400 metadata elements, 20 core elements. The ISO standards are revised and modified on a regular basis, ISO 19115:2003 is the current version

INSPIRE (INFRASTRUCTURE FOR SPATIAL INFORMATION IN EUROPE)

The INSPIRE Directive sets out to improve the efficiency and effectiveness of public services – those associated with European environmental policy in the first instance – through the provision of a European spatial data infrastructure. It has major relevance to this project as it:-

- defines data standards for spatial data INSPIRE metadata schema is compliant with ISO 19115
- the Department of Environment, Heritage and Local Government is one of the legally mandated organizations
- focused initially on environmental datasets but this can be extended and adapted to encompass cultural heritage data in the future
- it will have become the de facto standard to which everyone should aspire

CULTURAL HERITAGE INCLUSION IN METADATA

The adoption of the INSPIRE directive metadata standard, compliant with ISO 19115, will ensure the geographical description of our datasets is completed to an international standard. Further thesauri, or controlled vocabularies can be added to the Keyword component of the metadata schema. Controlling how the cultural component is described using these resources enhances the ability of users to search and retrieve our data in intelligent ways. More than one thesauri can be defined within a schema and our research identified a number which could be adopted. Examples of cultural heritage thesauri / controlled vocabularies include:-

- **The Getty Institute Art & Architecture Thesaurus (AAT)** - controlled vocabulary used for describing items of art, architecture, and material culture. This thesaurus is compliant with two further ISO standards:-ISO 2788 & ISO 5964
- **CIDOC Conceptual Reference Model CRM** - definitions and a formal structure for describing concepts and relationships used in cultural heritage documentation. Accepted as ISO standard 21127
- **Monument Inventory Data Standard (MIDAS)** - is the UK data standard for information about the historic environment. It states what information should be recorded to support effective sharing of the knowledge of the historic environment, and the long-term preservation of those records. and its objective is to complement existing standards such as CIDOC CRM
- **Humanities and Social Science Electronic Thesaurus (HASSET)** - a subject thesaurus which has been developed by the UK Data Archive (UKDA) over the past 20 years.

Controlled vocabularies are also applied to geographic placenames:-

- **The Getty Institute Thesaurus of Geographic Names Online (TGN)** – hierarchal structured definition of geographic descriptions for the world
- **Placenames Database of Ireland** – English and Irish language thesauri, including county, barony and townland names

TOOLS FOR METADATA

Given the creation of metadata has been identified as critical to the archiving model it is important to identify the variety of free and commercially available tools to support metadata creation editing and validation. Among those tested:-

- **INSPIRE Geoportal –** online resource which allows user to create, validate and export as xml document.
- **ESRI ArcCatalog -** a flexible metadata creation and viewing application. Data is input into the fields of a tab-based interface with mandatory fields indicated. Once created the metadata

can then be viewed in a range international standard formats by selecting the appropriate stylesheet

- **ISO Metadata Editor (IME) Tools -** a number of IME application's can be freely downloaded from the internet

RECOMMENDATIONS

1. The OAIS model should be adopted for the archival system.
2. A Submission Information Package guideline document should be created to assist data providers achieve the appropriate standard of data compliance.
3. The metadata schema for 'share-IT' should be compliant with both INSPIRE and ISO 19115 and should have its keywords expanded to include a selection of Thesauri, to standardise geographic placenames and cultural components.
4. 'Preferred' data formats should be defined for each of our three data types:-
    a. LIDAR
       Archive – *ASCII xyz, and xml metadata*
       Dissemination – *ESRI ASCII raster file format*
    b. ORTHIMAGERY
       Archive – *GeoTIFF, and xml metadata*
       Dissemination – *ESRI ASCII raster file format*
    c. GEOPHYSICAL SURVEY
       Archive – *ASCII xyz, and xml metadata*
       Dissemination – *ESRI GRID raster file format*
5. A comprehensive copyright and access policy should be developed in consultation with the data providers and archaeological community (IAI).
6. The cost model needs to be examined and a strategy for financing the archiving process considered in consultation with the wider archaeological community.

SECTION 3: REVIEW OF CURRENT BEST PRACTICES FOR THE LONG TERM ARCHIVING, DATA STANDARDS, AND ACCESS TO DIGITAL ARCHAEOLOGICAL LANDSCAPE DATA

INTRODUCTION

The aim of this work package was to develop a data management strategy for the *Share-IT* project. To accomplish this objective, a review of current Irish and International best practice was undertaken, addressing issues relating to archiving, data standards and access of digital archaeological landscape data.

The review involved contacting organisations identified to be managing similar datasets, and where possible meetings or telephone interviews were arranged to discuss the issues and strategies. From these meetings a number of important international methodologies and standards emerged.

The complexity of developing such a data management strategy became apparent as our research progressed, with important concepts and issues being introduced that took our review beyond our original project remit. They have been incorporated into the following sections of this chapter to satisfy the development of a robust data management strategy

ARCHIVING REVIEW

In recent times the rapid growth in the creation of digital data has highlighted the speed and ease of short-term data dissemination with little regard for the long-term preservation of digital data. However, digital data is fragile in ways that differ from traditional technologies, such as paper or microfilm. It is more easily corrupted or altered without recognition. Digital storage media have shorter life spans, and digital information requires access technologies that are changing at an ever-increasing pace. Because of the speed of technological advances, the time frame in which we must consider archiving becomes much shorter. [11]

'The creation of stable, consistent, logical, and accessible archives from fieldwork is a fundamental building block of archaeological activity' – Hedley Swain, Museum of London[12]

Fortunately nearly all those interviewed during our research agreed with sentiments of this statement. It was generally accepted that only by creating such a structured high quality archive can the evidence, which has been the basis for our archaeological interpretation and understanding, be preserved, re-examined and re-used in the future. Given the often destructive nature of the archaeological process, data may often be irreplaceable, and in the case of landscape data it often presents a unique 'snapshot' in time, an exceptionally valuable resource.

WHAT ARE THE ISSUES FOR LONG-TERM STORAGE OF DIGITAL INFORMATION?

Digital data presents a range of challenges very different to those of archiving traditional paper records, where, if controls of temperature and humidity are maintained it becomes a largely passive process. On the contrary, digital archiving is an active process in that it requires regular management and cannot simply be left in static storage.[13] For a number of reasons a coherent digital archive strategy is needed.

There are a number of issues relating to the technology and specifically changes in technology associated with digital data:-

OBSOLETE FILE FORMATS

Backwards compatibility, although a common aspect of most software companies when releasing new versions, cannot be guaranteed into the future. There is a debate as to the real extent of this problem but a number of examples have been highlighted that indicate it is a real issue, e.g. the latest major update to Office 2007 silently disables the ability to open many files that Office would previously handle. They include formats such as Corel Draw, but worryingly for many people also include older MS Office formats, such as Excel 97 and Word 97.[14]

DETERIORATION OF STORAGE MEDIA

Standard media used for the storage of digital data – tapes, disks, CDs, DVDs have a finite lifespan and can become corrupted. The original manufacturer's claims of 'indestructible' CDs are now replaced by most cautious estimates on lifespan as low as 20 years[15] An example of this issue dates back to the 1975 NASA Viking mission to Mars which was searching for evidence of life. The data was compiled by scientists and stored for future generations on magnetic tape. Yet despite the best efforts of NASA using climate controlled storage by the 1990's the tapes had become brittle and begun to crack. Furthermore the tapes which could still be read were in a format which could no longer be de-coded. The data was only salvaged by painstakingly tracking down old printouts and retyping the data.

OBSOLETE MEDIA

There are already examples of storage media which are effectively obsolete. It has been a number of years since computers had a 5¼" drives as standard, and many computers being manufactured now no longer have 3½" drives.

OBSOLETE SOFTWARE AND HARDWARE

The problem of obsolescence continues with software which cannot be run on modern computers, example Wordstar, and of obsolete hardware needed to run software or read the data. An example of this was the BBC Doomsday project in 1986 which only ran on BBC Micro computers, now obsolete, and was recorded on 30cm laser discs, and obsolete media.

LOSS OF EXPERTISE

A crucial factor, often overlooked is the fact that digital data may be stored in ways which only make sense to the current or previous staff, but lack the adequate documentation, and sensible naming strategy others would need to make sense of it. As staff leave the organisation the knowledge, and hence the accessibility of the data may disappear with them. An example of this was the Newham Museum Archaeological Service, active in archaeological fieldwork across London for over 10 years, accumulating significant amounts of data until abrupt closure in 1998. Staff went on to new posts, the computers on which much data was stored were sold off by the council and only a last minute salvage operation ensured the entire contents of hard disks were copied onto floppy disks. When these disks were eventually examined issues of data corruption and obsolete software occurred but the biggest problem was the lack of documentation. The ADS (more later in section) was able to save much of the data through time consuming hard work, a cost which could've been avoided if good practice shad been adopted at the point of creation.[16]

Such an eventuality is a real possibility in Irish archaeology given the current economic climate. If funding is withdrawn from an organisation holding archaeological data whether, or how well data is archived and stored is dependent on the diligence of the staff involved. An example was the Irish Archaeological Wetland Unit (IAWU) in UCD which was closed in 2005 but has thankfully archived its data within the School of Archaeology due to the attentiveness of the unit's director.

Evidence such as the DPC survey results shown in the table below shows that these issues are already having an effect on many organisations

Beyond these technical issues there are a number of more altruistic reasons why we should be archiving digital data

	YES	NO	Don't Know	Don't want to answer
Does your organisation have any inaccessible data?	36%	29%	31%	4%
Has your main type of data been lost?	28%	43%	29%	
Any data in danger of becoming obsolete?	48%	21%	27%	4%
Are any file formats that have been used now obsolete	38%	35%	27%	

Table 3-1 Source, 2005 Digital Preservation Coalition survey

ENABLING FUTURE RE-USE OPPORTUNITIES

The ability to re-use data has considerable benefits in academic research and, ultimately to the wider community. For researchers the benefits of fast access to a wider range of data sets, archived to a reliable and documented standard will be significant. It opens the possibility of a more interdisciplinary approach, with researchers having the option to access and share data from remote locations. Our ability to re-analyse data in the future as intellectual theories and understanding evolves will depend on the quality of the digital archive.

The following institutions and organisations were consulted in a review of best practice in digital archiving; their main foci are briefly outlined below. This review contributed to an improved overall understanding of the structures required in digital preservation, and contributed to the design of an appropriate archiving strategy for geospatial archaeological data.

DIGITAL HUMANITIES OBSERVATORY (DHO)

The DHO, a project of the Royal Irish Academy, was established to manage and co-ordinate the increasingly complex e-resources created in the arts and humanities. It aims to enable research and researchers in Ireland to keep abreast of international developments in the creation, use, and preservation of digital resources. It will fulfil these objectives by:

- serving as a knowledge base in Ireland via consultations with project partners;
- setting national standards to ensure the interoperability, preservation, and long-term accessibility of digital resources;
- establishing a central repository which will provide access to a wide variety of interdisciplinary, multilingual, and multimodal digital resources created on the island of Ireland

MIDA – THE MARINE IRISH DIGITAL ATLAS

The Marine Irish Digital Atlas (MIDA) [17]provides a single source where people interested in coastal and marine information can visualise and identify pertinent geospatial datasets and determine where to acquire them. MIDA offers both digital geospatial data and information, incorporating text and multimedia elements, related to coastal and marine resources in Ireland. Integrating the latest advances in web-based mapping techniques, the atlas is built around an interactive map, which allows anyone to identify, visualise, and query those datasets relevant to their interests. The atlas displays data layers from numerous coastal and marine organisations both within Ireland and abroad, thus providing the best single resource for finding and viewing existing Irish coastal and marine data.

The key goal of the MIDA project has been to develop a Marine Irish Digital Atlas as an updateable web GIS based data archive and informational resource.

ENVIRONMENTAL PROTECTION AGENCY (EPA) - ENVIRONMENTAL RESEARCH CENTRE

The Environmental Research Centre (ERC) was established under the National Development Plan 2000-2006 to allow for a more structured approach to environmental research and to provide stronger environmental support to the plan. Its overall purpose is to help ensure that development is environmentally sustainable. The aims of the ERC may be listed as follows:

- to allow for a more structured approach to environmental research,
- through the development of advanced innovative techniques
- and systems, and addressing priority environmental issues,
- thereby supporting environmentally sustainable development.

In addition, the EPA collates all datasets generated during the research projects. The ERC ensure safekeeping and management of these valuable datasets. Detailed information (Metadata) on these datasets can be obtained from the ERC website[18]. All datasets will be made available usually upon a 12-month period after the publication of the final report.

SAFER-Data is a web-based interface to the Environmental Data Archive maintained by the Environmental Research Center (ERC) in the Environmental Protection Agency (EPA)

SAFER-Data provides a user friendly interface for a variety of users:

- Public Users - those interested in finding out about environmental research, exploring data, and possibly downloading data and/or reports to their own computers for further studies
- Researchers - both researchers looking for data and information about other research projects and also researchers uploading their environmental data and information for archival on the Secure Archive For Environmental Research Data System
- EPA Users - interested in exploring information about environmental research currently being carried out and results of research projects which have concluded.

COIMBRA GROUP

Founded in 1985 and formally constituted by Charter in 1987, the Coimbra Group is an association of long-established European multidisciplinary universities of high international standard committed to creating special academic and cultural ties in order to promote, for the benefit of its members, internationalisation, academic collaboration, excellence in learning and research, and service to society. It is also the purpose of the Group to influence European educational policy and to develop best practice through the mutual exchange of experience.

Amongst its aims are to be recognized as an expert body, able to advise its members and EU institutions on various matters relating to higher education, such as Information Technology (IT) as applied to new teaching methods and lifelong learning.

LIFE (LIFE CYCLE INFORMATION FOR E-LITERATURE)

LIFE (Life Cycle Information for E-Literature) is a collaboration between University College London (UCL) and the British Library.

The LIFE Project has developed a methodology to model the digital lifecycle and calculate the costs of preserving digital information for the next 5, 10 or 100 years. For the first time, organisations can apply this process and plan effectively for the preservation of their digital collections.

The LIFE Project has completed its second phase ("LIFE2"), an 18 month project running from March 2007 to August 2008. The LIFE2 Project Final Report and supporting documentations can be viewed from the LIFE2 Documentation Page.[19]

ARCHAEOLOGICAL DATA SERVICE (ADS)

The ADS was founded in 1996 for the purpose of preserving digital data produced by archaeologists based in the UK, and making it available for re-use.

The ADS supports research, learning and teaching with high quality and dependable digital resources. It does this by preserving digital data in the long term, and by promoting and disseminating a broad range of data in archaeology. The ADS promotes good practice in the use of digital data in archaeology, it provides technical advice to the research community, and supports the deployment of digital technologies.

The ADS is at the forefront of developing national and trans-national resource discovery, data aggregation and data dissemination technologies in archaeology. The ADS has extensive expertise in the fields of digital curation, heterogeneous data set mapping (including geospatial data) and digital data standard development and application in the arts and humanities sector.

BIG DATA PROJECT: PRESERVATION AND MANAGEMENT STRATEGIES FOR EXCEPTIONALLY LARGE DATA FORMATS: 'BIG DATA'

In 2004 the ADS was commissioned by English Heritage to examine the particular issues and problems associated with the preservation and management of large data formats in, culminating in a report on ‘Big Data’ published in 2007[20]. The project design set out a programme for investigating preservation (storage methods), reuse (usability) and dissemination (delivery mechanism) strategies for exceptionally large data files generated by archaeologists, researchers and cultural resource managers undertaking fieldwork and other research.

The data in question is typified by large formats that have exceptionally large file sizes and in particular the technologies associated with their storage and delivery. The generation and use of such data for research is increasing in specific fields of archaeological and cultural resource management activity (maritime archaeology and surveying, laser scanning, LiDAR, computer modelling and other scientific research applications). Yet there is little understanding of the implications for cost and good practice in data preservation, dissemination, reuse and access. This lack of understanding is potentially exacerbated by the proprietary nature of formats generally used by the new research technologies now being used in archaeology and cultural resource management.

The project seeks to answer immediate questions regarding cost and to develop recommendations and strategies for archaeologists, researchers, cultural resource managers and archivists dealing with 'Big Data'. The project recognises that computing capacity, both to create and to archive data, will continue to rise. The aims of the project consequently address generic and strategic issues as well as the immediate questions posed by 'Big Data' today.

OAIS - REFERENCE MODEL FOR AN OPEN ARCHIVAL INFORMATION SYSTEM (OAIS)

The Consultative Committee for Space Data Systems (CCSDS) was formed in 1982 by the major space agencies of the world, including NASA, to provide a discussion forum for common problems in the development and operation of space data systems. One outcome has been the recommendation of standards for the preservation of space related data through the OAIS reference model. It defines the basic functional components of an archive and provides a comprehensive framework for describing and analysing preservation issues.

“An OAIS is an archive, consisting of an organization of people and systems that has accepted the responsibility to preserve information and make it available for a Designated Community” (CCSDS 2002, 1-1)

In 2002 OAIS was approved as ISO standard 14721[21], to establish a system for archiving information, both digitalized and physical data.

Our discussions with ADS and subsequent research of their publications, in particular the ‘Big Data’ project emphasised the benefit in using the OAIS as an archiving model. This is particularly the case when collaborating with external organisations as it provides a language and a set of terms that can

aid communication. OAIS emphasises the requirement for ongoing management and administration in digital preservation, i.e. the need for life cycle management, a theme which is covered in more detail later in the chapter. The full OAIS 'blue book' [22] presents in detail the recommendations. For this report Figure 3-1 and the following table describing the key components of the framework give a brief introduction to how the system is designed.

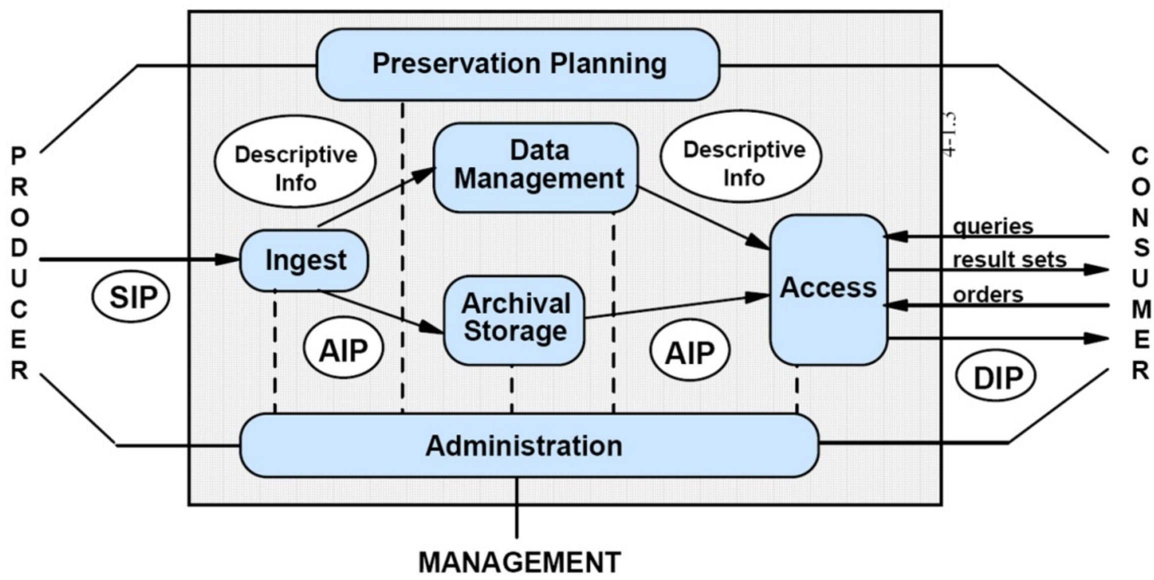

Figure 3-1 OAIS Functional Entities (taken from CCDSD 2002, 4-1)

OAIS TERMINOLOGY	Description
Producer	The role played by those persons or client systems, which provide the information to be preserved..
Submission Information Package (SIP)	An Information Package that is delivered by the Producer *(in this case the archaeologist)* to the OAIS for use in the construction of one or more AIPs.
Geophysics Example: raw DAT (.DAT), Geoplot software files (.geo), geoTIFF (.tif), pdf report	*Note: This is usually in a commercial proprietary software format*
Ingest	The OAIS entity that contains the services and functions that accept Submission Information Packages from Producers, prepares Archival Information Packages for storage, and ensures that Archival Information Packages and their supporting Descriptive Information become established within the OAIS.
Archival Information Package (AIP):	An Information Package, consisting of the Content Information and the associated Preservation Description Information (PDI), which is preserved within an OAIS.

Geophysics Example: ASCII comma delimited text file (.txt), metadata file (.xml)	*Note: This should be in a non-proprietary format, such as ASCII text*
Archival Storage	The OAIS entity that contains the services and functions used for the storage and retrieval of Archival Information Packages.
Data Management	The OAIS entity that contains the services and functions for populating, maintaining, and accessing a wide variety of information. Some examples of this information are catalogues and inventories on what may be retrieved from Archival Storage, processing algorithms that may be run on retrieved data, Consumer access statistics, Consumer billing, Event Based Orders, security controls, and OAIS schedules, policies, and procedures.
Access	The OAIS entity that contains the services and functions which make the archival information holdings and related services visible to Consumers.
Dissemination Information Package (DIP)	The Information Package, derived from one or more AIPs, received by the Consumer in response to a request to the OAIS. *Note: This is the content that users will see and will be supplied in a format they can currently use. This may evolve over time to suit evolving user requirements but the function of the OAIS will ensure it can be generated from the AIP*
Geophysics Example: GIS layers – shapefile (.shp), geodatabase (.mdb), raster grid, geoTIF	

Table 3-2 Selected OAIS terminology (taken from CCSDS-2002-1.7.2)

The OAIS standard identifies the following 6 mandatory responsibilities that an organization must discharge in order to be considered OAIS compliant:-

The OAIS must:

- Negotiate for and accept appropriate information from information Producers.
- Obtain sufficient control of the information provided to the level needed to ensure Long-Term Preservation.
- Determine, either by itself or in conjunction with other parties, which communities should become the Designated Community and, therefore, should be able to understand the information provided.
- Ensure that the information to be preserved is **Independently Understandable** to the Designated Community. In other words, the community should be able to understand the information without needing the assistance of the experts who produced the information.

- Follow documented policies and procedures which ensure that the information is preserved against all reasonable contingencies, and which enable the information to be disseminated as authenticated copies of the original, or as traceable to the original.
- Make the preserved information available to the Designated Community.[23]

DATA FORMATS

The data being considered by share-IT project is initially limited to three data types, LiDAR, orthoimagery, and geophysical survey. These datasets not only have a geo-spatial graphical component (i.e. a map) but also have associated underlying data files, and potentially a cultural interpretation component.

A key data preservation issue is which file format is selected as the archival version, and is critical to the longevity and future access to the data. Accepting the value of the OAIS reference model it is simplest to consider the appropriate file formats in terms of the three information packages defined in Table 3-2. The archival information package is the version which will be held in perpetuity, and as such need to be in a standard non-proprietary format such as ASCII. The choice of this format is critical as the submission format must be able to migrate into it, and the dissemination format be generated from it.

This section will consider the formats commonly associated with each of our data types defined in the stages of the OAIS model.

GEOPHYSICAL SURVEY DATA

DATA DESCRIPTION

Geophysical survey applies scientific techniques to remotely gather information about the location and characteristics of subsurface archaeological features. These techniques measure physical attributes (e.g. resistance to electrical current, or magnetic variations) and result in a matrix of data points.[24] The results from a geophysical survey will include the raw data files from proprietary software, a plot or image file (with geo-referencing information), and potentially an interpretative vector plan. It is important to distinguish between the processed results, such as the image file) which is the end result of a number of filtering and statistical calculations.

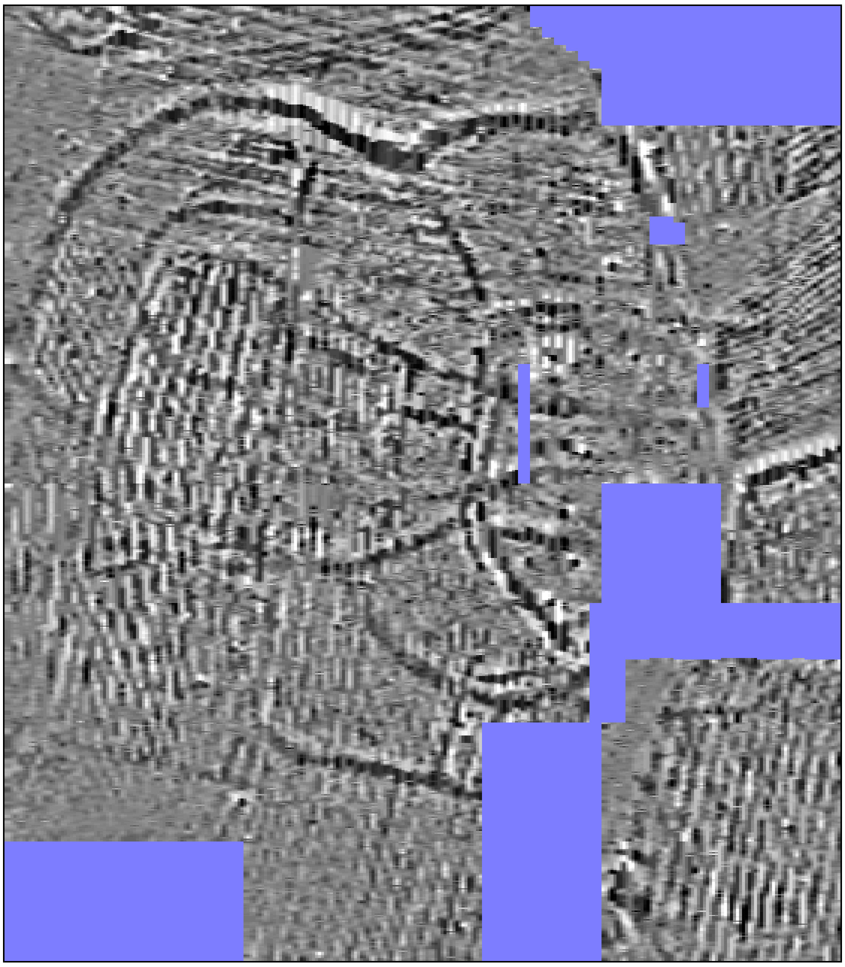

Figure 3-2 Magnetic Gradiometry survey from Killukin, Co. Roscommon (The Discovery Programme, MRS project)

The OAIS reference model can be used to identify the data formats appropriate at the three information package stages.

SIP POTENTIAL DATA FORMATS

Dat files, Geoplot files – proprietary file formats from the datalogger and initial processing stages of software such as Geoplot.

ASCII x,y,z – As with LiDAR (above) this is the standard for raw data files and is the common approach (e.g. ADS).

GeoTIFF – see orthoimagery (above).

JPEG format – see orthoimagery (above).

ESRI Shape files - an industry standard geospatial data format, and compatible with most GIS software. This is currently the format used to store vector interpretations of geophysical survey at the Discovery Programme.

DXF files –AutoCAD drawing exchange format, a proprietary format developed for enabling data interoperability. This format is suitable for storing the vector interpretations of geophysical survey.

AIP POTENTIAL DATA FORMATS

ASCII x,y,z – As with LiDAR (above) this is the standard for raw data files and is the common approach. The AIP should include the processed and unprocessed raw data.

Note: The AIP for all data formats must include the xml file containing the metadata, a concept which will be covered in detail later in the chapter.

DIP POTENTIAL DATA FORMATS

GeoTIFF – (see orthoimagery) this is often the basic dissemination format for geophysical survey data but it is limited in use in that it denies access to the actual surveyed values, it is simply a depiction. It has serious constraints such as it must be a full rectangle and has no inherent transparency. This creates s serious problem when viewing multiple geophysical images with overlapping polygons.

ESRI GRID – this raster format is a more powerful dissemination format as it contains the data values recorded in the survey, from which a raster image can be created.

ESRI Shape files – (see SIP above), used for interpretation and displaying gridlines.

DXF files – (see SIP above), used for interpretation and displaying gridlines.

ORTHOIMAGERY

DATA DESCRIPTION

Orthoimages are the geometrically corrected and geographically referenced images which result from photogrammetric processing. They are high resolution digital images, normally created as TIFF data files. The storage of these data types conforms to the standards applied to image files in general, however they contain a spatial component which needs to be accommodated within the format. The photogrammetric processing also generates a DSM (see previous section), which is required to produce the orthoimage.

The OAIS reference model distinction between the SIP, AIP and DIP is less clearly defined with this image data. The preserved archive format may well turn out to be the appropriate dissemination format partly due to the widespread adoption of and familiarity with standard image formats.

Figure 3-3 An orthoimage tile (1 km^2) from the Roscommon aerial project, part of the Discovery Programme's Medieval Rural Settlement project.

AIP - POTENTIAL DATA FORMATS

GeoTIFF - TIFF files which have geographic (or cartographic) data embedded as tags within the TIFF file. The geographic data can then be used to position the image in the correct location and geometry on the screen of a geographic information display. GeoTIFF is a metadata format, which provides geographic information to associate with the image data. But the TIFF file structure allows both the basic metadata and the image data to be encoded into the same file.[25] This is the currently used format at the Discovery Programme for orthoimagery created from PCI Geomatica 10 photgrammetric software. Although widely adopted by the community at large, this format is owned by Adobe Inc. and as such is deemed proprietary, adversely affecting its suitability as an archive format.

JPEG2000 format – the JPEG (Joint Photographic Experts Group) committee has addressed many of the limitations of the original JPEG format and its latest format, JPEG2000 has emerged as a new standard for the effective preservation of digital image data. The format is published as International Standard ISO/IEC 15444 Part 1[26]

The particular advantages from an archiving perspective are:-[27]

- Metadata - the format embeds metadata within the file in a standard XML compliant environment. This allows for the possibility to incorporate descriptive information within the file.
- Lossy and lossless compression (with high quality lossless decompression available naturally through all types of progression)
- Progressive transmission by quality, resolution, component, or spatial locality
- Multiple resolution representation (images are decomposed into multiple resolutions in the compression process). This will dramatically increase the speed of display for large images, particularly important for high resolution data.
- No limit on file size, significant as image resolution increases.

(JPEG was the original JPEG committee standard for images (IS 10918-1) developed more than 15 years ago. It is generally not considered as an archive quality format primarily due to loss of quality on compression, and generation loss issues[28])

The Open Geospatial Consortium (OGS) has adopted this format and defined the means by which the OpenGIS® Geography Markup Language (GML) can be used within the JPEG2000 format, GMLJP2[29] . GML is an xml schema used to describe geographic information, including elements such as coordinate system, coverage, unit of measure, and also vector based objects (e.g. points, lines, and polygons).

GMLJP2 is intended to handle a variety of imaging use cases including the following:

- *Single geo-referenced images. GML describes the geometry and the radiometry.*
- *Multiple geo-referenced images of the same type. GML describes the geometry and the radiometry of the constituent images. Examples include a stereo photographic pair, a triangulation block of images, or image mosaics.*
- *Multiple geo-referenced images of various types. GML describes the geometry and the radiometry of the constituent images. Examples include combinations of images such an optical image, FLIR and SAR images for target identification.*
- *Ortho-rectified images with or without associated digital elevation models.*
- *Digital Elevation Models that incorporate terrain-based constraints.*[30]

With support at this level in the GIS community this format is rapidly being established as an industry standard for image archiving. This should be monitored with a view to its adoption as an archiving standard.

LIDAR (LIGHT DETECTION AND RANGING)

DATA DESCRIPTION

LiDAR is an optical remote sensing technology that measures properties of scattered light to find range and/or other information of a distant target from an aerial platform, usually a plane or helicopter. Millions of height points are gathered, known as point clouds, from which detailed 3D surface models can be constructed. The laser returns more than one signal from the surface it strikes, and by filtering routines a data set for the first and last return can be generated, defining the DEM (surface model including trees and buildings) and a DTM (the ground surface or ‘bare earth’ model.

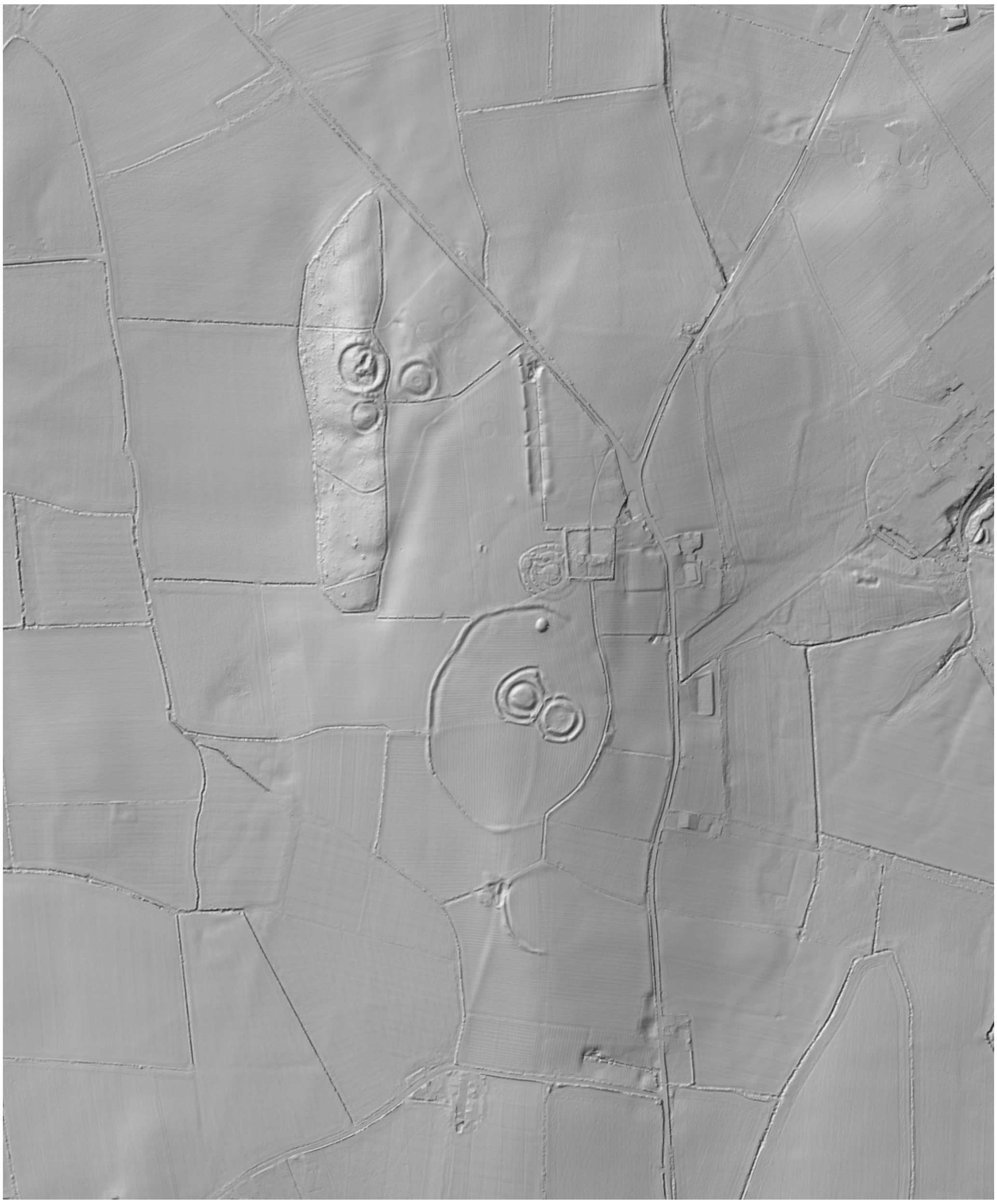

Figure 3-4 Hillshade model of the DTM for the Hill of Tara high resolution LiDAR survey, The Discovery Programme

SIP- POTENTIAL DATA FORMATS

Data is unlikely to be gathered directly by archaeological organizations but will be supplied from specialist providers. The relationship between the data provider and the user will often determine the data format options available. Licensed data providers such as national mapping agencies often provide existing data simply in the form of processed DEM and DTM grids. These are continuous height surfaces generated from processing the point cloud data. Commissioning new data will enable the user to specify the formats required, and result in access to ascii data files - one from the first return, and one for the last return. This data will contain additional information:-

Intensity value - the strength of the laser return signal, determined by the surface characteristics

RGB value - the 3 spectral channel values which make up a digital image data

The strength of data in this format is that it enables the user to control the generation of DEM and DTM surfaces and presents the data in a raw form, albeit with a considerable amount of processing and computation to arrive at these data sets – the laser component, GPS position, INS calculations have all been processed to deliver the point cloud data in WGS84coordinates (or transformed to a local system such as Irish Grid).

Large file size, and the impact on submission mechanisms, is a consideration with LiDAR data. Even when tiled into smaller blocks ascii files will be in the order of 0.5GB, with similarly sized grid files. A project could easily be supplied with 50 – 100GB data. Submission has generally been made via portable media, such as DVD or Hard Drives.

There may also be an argument for storing the pre-processed raw data in the event that improvements in algorithms allow better models to be achieved from the data. This should be considered after consultation with the data provider.

AIP- POTENTIAL DATA FORMATS

The preserved data must include the full data set before the creation of the DTM to ensure that improvements in processing algorithms over time can be applied to the data.

LAS format - The LAS format is a public file format for the interchange of LIDAR data between vendors and customers. This binary file format is an alternative to proprietary systems or a generic ASCII file interchange system used by many companies[31]

ASCII xyz – This is considered the standard format for text files. LiDAR data supplied by BKS to the Discovery Programme is in this format. Unlike the LAS format, ASCII can be easily understood by other software and opened easily to view and read by users.

Xml – metadata file, to be discussed later in the chapter.

DIP - POTENTIAL DATA FORMATS

The dissemination of LiDAR data is often an extracted visual representation of the data rather than the data itself. A common example would be disseminating a prepared hillshade model, providing a more accessible and useful product to the user. The range of dissemination format options could be extended to GIS geodatabase files if the user has the expertise to exploit them.

ESRI ASCII raster file format - a simple format that can be used to transfer raster data between various applications. It is basically a few lines of header data followed by lists of cell values. This is the format used by the UK Environment Agency.

DML - the Doppler Markup Language. This format is based on the XML- Markup-Language which is commonly used to describe data formats in the World-Wide-Web. The format has the advantages that it is ASCII and can be edited with any standard editor under Unix or Windows[32]

ADDITIONAL INFO AND DATA PROVIDED

LiDAR data providers offer a range of additional services, orthoimagery (which will be covered in the next section) is available as standard and video footage from the flight may also be available.

It should be noted that LiDAR is a developing technology and currently only the discrete returns of the laser pulse are analysed. Research into exploiting the entire full-waveform signal will result in new data format requirements in the future.

ARCHIVE PROCESSES

A core component of an OAIS compliant archive is Archival Storage. This represents the part of the archival system that manages the storage and maintenance of digital objects entrusted to the archive. The Archival Storage function is responsible for ensuring appropriate types of storage, the appropriate structure of the file system, the necessary amount of storage available and other issues related to the physical management of data storage.

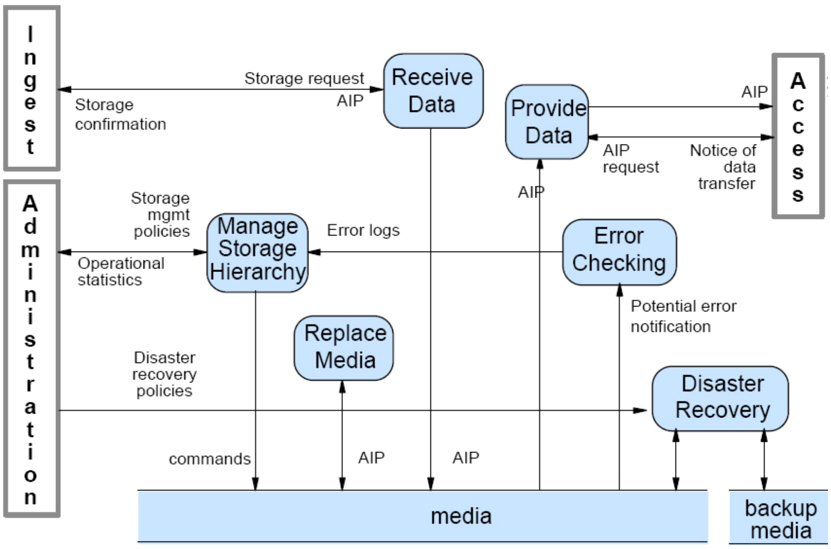

Figure 3-5 Functions of Archival Storage (taken from CCDSD 2002, 4-3)

DATA INGEST

Data ingest, or data acquisition is an important component of an OAIS system, however the OAIS model does not cover pre-ingest activities and assumes that agreements are in place before Submission Information Package (SIPs) are ready for transfer.

PRE-INGEST ACTIVITIES

The following information comes from[33] and emphasises the important of a coherent pre-ingest strategy:-

“....the pre-ingest phases for producer-archive interaction that lead to new material being accepted to the archive are:

- *the Preliminary Phase, also known as a pre-ingest or pre-accessioning phase, includes the initial contacts between the Producer and the Archive and any resulting feasibility studies, preliminary definition of the scope of the project, a draft of the submission information package (SIP) definition and finally a draft Submission Agreement;*
- *the Formal Definition Phase includes completing the SIP design with precise definitions of the digital objects to be delivered, completing the Submission Agreement with precise contractual transfer conditions such as restrictions on access and establishing the delivery schedule;*
- *the Transfer Phase performs the actual Transfer of the SIP from the Producer to the Archive and the preliminary processing of the SIP by the Archive, as it is defined in the agreement. The transfer and validation phases are often carried out partially in parallel, as there is iteration when all the information to be submitted is not submitted at once;*
- *the Validation Phase includes the actual validation processing of the SIP by the Archive and any required follow-up action with the Producer”.*[34]

They conclude pre-ingest functions are an essential component of an efficient and effective system.

ENCOURAGING DATA SUBMISSION

Pre-ingest and SIP activities will clearly have cost implication to organizations considering submitting data to any potential Irish digital archive. This raises the question of how will we encourage people to submit data to the system, and what are the appropriate leverage mechanisms?

One possible approach could be to insert a contractual obligation to recipients of grants or state funded contracts that data must be submitted on completion of the project. This would embed the cost of delivery with the project, and ensure high standards of quality control. In terms of geophysical survey current legislation requires that to fulfil the terms of a licence a report must be submitted to the DoEHLG on completion of the work. This may present the opportunity for the legislation to be adjusted to include the requirement to also supply the data files along with the report. Legal obligations to supply data may also result from initiatives such as the EU INSPIRE directive, discussed later in this chapter.

An alternative approach is to allow the momentum of data and the quality of the system to encourage participation without recourse to legislation or contractual obligation. A good argument can been made that if a webGIS as proposed by this project is developed, promoted and seeded with quality data then organisations will see the value in using it, wish to be associated with it, and be prepared to take on the task (and cost) of preparing data to the required standard.

ACCESS AND RIGHTS POLICY

One of the aims of the share-IT project is the dissemination of geo-spatial data, therefore our objective has to be to maximise use of the system. However we need to consider the intellectual rights and copyright implications of making data available via a webGIS system.

COPYRIGHT

A range of approaches to the issue of copyright were observed noted during the review of best practice. MIDA (the Marine Irish Digital Atlas) confronts this issue by way of a *'Memorandum of Understanding (MOU) for its data contributors.* The precise specification is adjusted to meet the needs of the supplier, creating a document in which the conditions that govern data supply, access and exploitation are fully laid out. Typical principles include:-

- The spatial dataset provided by a data owner may be displayed in the web-based GIS. This will be displayed as the data owner provides them, or generalised in a way that the data owner and the Coastal & Marine Resource Centre (CMRC) agree upon.
- Spatial dataset cannot be downloaded from the web-based GIS unless the owner has given prior consent.
- The contact details of the data owner will be provided in the metadata and therefore will be available over the Internet to atlas users who are interested in acquiring a copy of the spatial dataset.[35]

The ADS requires users to accept both a *Copyright and Liability Statement* and a *Common Access Agreement (see appendix I)* before accessing its *ArchSearch + Data* resource.

The OAIS model recognises the importance of copyright, *"An archive will honour all applicable legal restrictions. These issues occur when the OAIS acts as a custodian. An OAIS should understand the copyright concepts and applicable laws prior to accepting copyright materials into the OAIS. It can establish guidelines for ingestion of information and rules for dissemination and duplication of the information when necessary*."

The large part of archaeological activity in Ireland is undertaken under licenses issued by the Archaeological Licensing Section of the National Monuments Service at the Diehl. The submission of the results in the form of a report is a condition of the license, and as such the results are in the public domain. Whether this system could be extended to include the data files which support the published reports is something which needs further consideration.

Currently a vast amount of archaeological work is being undertaken in advance of infrastructural projects, commissioned by state bodies such as the National Roads Authority. This data is being paid for by the state and it would seem appropriate that it be made available once the planning process has been passed, and the project completed.

As will be discussed later in this chapter, the metadata must clearly state the conditions attached to access and copyright, and must deal with the issue of quality assurance.

USER COMMUNITY

The OAIS model identifies the 'Designated Community' as the set of consumers who should be able to understand the preserved information. It also emphasises that this community will evolve or change over time.

Archives can allow different access to information or data depending on the user status. It may be that general open access is only given to basic levels of data and simple viewing tools, with different access and functions such as downloading facilities available to those registered or even paying subscription. This was noted with the SAFER-Data web-based interface of the EPA. They identified three categories of user, controlled by a registration and login system:-

- ***Public Users*** - those interested in finding out about environmental research, exploring data, and possibly downloading data and/or reports to their own computers for further studies
- ***Researchers*** - both researchers looking for data and information about other research projects, and also researchers uploading their environmental data and information for archival on the Secure Archive For Environmental Research Data System
- ***EPA Users*** - interested in exploring information about environmental research currently being carried out and results of research projects which have concluded

PROMOTION OF DIGITAL ARCHIVING

The success of a webGIS such as that proposed by the share-IT project is dependent on high volumes of data being submitted. This is a reason in itself for the share-IT project to actively promote the value of digital archiving.

The ADS express the view '*that there is little point in preserving data unless it is reused'* (Mitcham and Richards, 2008), and activity promote the dissemination of data through its web interface. Options range from pages with downloadable files to interactive maps and searchable online interfaces.

HOW MUCH DOES IT COST?

It should be clear from this chapter that archiving data involves costs from the data preparation and ingest stage, through to the long term costs of the digital archive lifecycle.

Two costing models from separate organisations were investigated:-

ADS COST MODEL

Archiving costs are calculated on the basis of 4 key elements:-

- **Management and Administration** – i.e. negotiating with depositor, processing the deposit, licences. This normally involves 2 - 3 days of effort
- **Ingest –** migrating data to preferred formats, creation of metadata, and entry of data to system. Cost dependent on number and complexity of files
- **Dissemination –** basic data delivery via simple file download is included in the price of data ingest, but special interfaces such as searchable databases or interactive maps may cost up to €15000 depending on functionality.
- **Storage –** (this includes the ongoing periodic process of data refreshments) Archives have to periodically upgrade systems - hardware and software - to take advantage of technological advances. (ADS have progressed through 3 generations of equipment during 10 years). [36]

The ADS has developed formulae to estimate the cost of archiving data over variable time periods, which include the costs of refreshing data, costs of physical equipment, and factor in decreases in these costs over time, shown in Table 3-3.

Retention Period	Cost	Cost (pence per MB)	Cumulative (pence per MB)
5 years	R + E	9 + 4 = 13	13
10 years	R – DR + E – DE	9 -3 + 4 – 1 = 9	22
15 years	R – DR + E – DE	9 -6 + 4 – 2 = 5	22
20 years	R – 3DR + E – DE	9 – 9 + 4 – 3 = 4	27
ongoing			30

Table 3-3 Retention cost model where R= refreshment cost, DR = decreasing cost of refreshment, E = cost of physical equipment, DE = decreasing cost of equipment (adapted from ADS)[37]

As the table shows, the conclusion from the ADS project was that a cost of (applying figures from the Big Data project) a one of charge of 30p per megabyte would cover ongoing preservation costs beyond 20 years. However, no account is taken of the number of files to be archived; e.g. 1 large file of 1GB size would involve significantly less effort than archiving 1000 smaller files of 1MB, although the total file size would be the same. Some adjustment to this model to account balance volume and number of files would be an improvement.

Applied to a small geophysical survey undertaken recently by the Discovery Programme which has generated 97MB of archiving data, the cost for preservation is around €30.

LIFE (LIFE CYCLE INFORMATION FOR E-LITERATURE

The LIFE Project has developed a methodology to model the digital lifecycle and calculate the costs of preserving digital information for the next 5, 10 or 100 years.[38]

There are 6 main lifecycle elements which are broken down further into lifecycle elements, similar to the OAIS functions, as shown in the following table.

LIFECYCLE CATEGORIES	LIFECYCLE ELEMENTS
Acquisition (Aq)	Selection
(or pre-ingest)	IPR
	Licensing
	Ordering and invoicing
	Obtaining
	Check-in
Ingest (I)	Quality assurance
	Deposit
	Holdings Update
Metadata (M)	Characterization
	Descriptive
	Administrative

Access (Ac)	Adding / maintaining links
	User support
	Access mechanism
Storage (S)	Bit-stream storage costs
Preservation (P)	Technology watch
	Preservation tool cost
	Preservation metadata
	Preservation action
	Quality assurance

Table 3-4 Breakdown of the elements in the LIFE model

The LIFE model elements defined are not compulsory, but rather provide a framework within which to work that will be applicable to most situations. The accuracy of the output however is dependent on the sub layers and customisation added alongside the amount of real data that you have to put into the calculator. The more data you collect or have, the more accurate the model becomes.

$$L_T = Aq + I_T + M_T + Ac_T + S_T + P_T$$

Figure 3-6 L is the complete lifecycle cost over time 0 to T. (from Lifecycle Information for E-literature [39])

From Figure 3-6 it can be seen that apart from the data acquisition costs all the other categories involve ongoing costs throughout the complete lifecycle. In terms of the share-IT project this is an important observation, which has to be understood in the context of identifying an appropriate hosting organisation.

Table 3-5 presents an example of costing using the LIFE model from the National Archives and Royal Library, Denmark. It gives a comparison of the costs of archiving 20,000 original film negatives (13TB) by either creating and archiving digital copies, or new film copies.

LIFE Costing Model 20,000 copies (13TB)	TIFF Uncompressed	105mm film
Production (Digitisation)	€134,886	€134,886
Production (Film Output)	€0	€180,201
Acquisition	€1,889	€1,889
Ingest	€2,283	€1,194
Archival storage	€35,910	€326
Preservation Planning	€922	€0
TOTAL	€175,890	€318,496

Table 3-5 LIFE Model costs from Danish National Archive / Royal Library example[40]

This LIFE example equates at slightly less than €9 per image. Other examples taken from case studies on the LIFE website quote costs between as little as £3 per image from the example of simple archiving of newspaper images to £30 for complex digital visual media.[41] There is clearly an element of cost associated with a project set up, and economies of scale will apply.

This improved understanding of the cost of archiving, and the models to help calculate the costs suggest it may be appropriate for projects to include this as a component in future grant applications. This would see the digital archiving of research assets become an integral part of overall project design and budget.

METADATA & ISO STANDARDS

GENERAL INTRODUCTION TO METADATA

Metadata is often described as 'data about data'[42]. Its purpose is to provide context for data and to facilitate the understanding and management of a specific dataset. This is a similar function to that of a legend, north arrow and scale bar on a map. It provides the 'who, what, where, why, when and how' information which allow users to judge the quality or reliability of the data.

Metadata is an integral part of the OAIS model. It is the 'Descriptive Info' component shown in Figure 1 at both Ingest and Dissemination sides of the model.

Metadata contains different levels of information which are all contained in the final schema. Three broad levels of metadata can be identified:-

- **Discovery** – the minimum information to convey the nature and content of the resource.
- **Exploration** – the information to ensure data is appropriate for purpose.
- **Exploitation** – the information required to access, transfer, and apply the data in an end application.

DUBLIN CORE

The Dublin Core metadata element set is a standard for cross-domain information resource description. It provides a simple and standardised set of conventions for describing things online in ways that make them easier to find. Dublin Core is widely used to describe digital materials such as video, sound, image, text, and composite media like web pages. Implementations of Dublin Core typically make use of XML and are Resource Description Framework (RDF) based.[43]

The Dublin Core Metadata Element Set is a vocabulary of fifteen properties for use in resource description.[44] The 15 metadata elements are,

1. Title
2. Creator
3. Subject
4. Description
5. Publisher
6. Contributor

7. Date
8. Type
9. Format
10. Identifier
11. Source
12. Language
13. Relation
14. Coverage
15. Rights

ISO 15836:2003 defines the Dublin Core metadata element set which deals with cross-domain information resource description.

Qualified Dublin Core enables the extension of the core metadata element set to include additional schema such as controlled vocabularies. The ADS recommend the use of Dublin Core and have refined and defined how the elements should be created. [45]In particular they defined the schema for Subject as the Thesaurus of Monument Types (RCHME 1995). However they note the flexibility of Dublin Core allows elements to be repeated, so to increase potential interoperability additional Subject element(s) could be added, possibly the Getty AAT controlled vocabulary to give an international dimension. The following fictional example (adapted from the ADS) shows the way in which Dublin Core might be used to describe a typical resource from the humanities. [46]

DC.title.main		Excavations at 2 Ordnance Terrace, Chatham
DC.creator.corporateName.1		Canterbury Archaeological Trust
DC.creator.postal.1		92a Broad Street
DC.creator.town.1		Canterbury
DC.creator.phone.1		+44 227 462062
DC.creator.role.1		contact
DC. subject	TMT	dwelling, house, detached house
DC.description.short		Excavations undertaken in advance of development examined the remains of a Victorian domestic dwelling, and uncovered previously unexpected Roman remains preserved beneath the cellar. The house is reputed to be the childhood home of author Charles Dickens.
DC.publisher.corporateName		Canterbury Archaeological Trust
DC.date.accessioned	ISO8601	1997-06-24
DC.type	ADS	process
DC.format.fileSize.1		720
DC.identifier	ADS	150

DC.identifier	*CAT*	*1984.11*
DC.identifier.url		http://blah/data.dbf
DC.identifier.url		http://blah/drawing1.dxf
DC.language	ISO639	en
DC.coverage.placeName		Chatham, Kent
DC.rights	AHDS	free

XML SCHEMA

As noted, metadata is usually presented as an extensible markup language **(**XML) document. An XML schema is a description of a type of XML document with constraints on the structure and content beyond the basics imposed by XML itself. As the word extensible implies an xml schema has the flexibility to be extended or altered to suit the specific needs of particular user communities. Not surprisingly standard schema has been adopted for geospatial datasets and has been adopted by the International Standards Organisation (ISO).

GEOSPATIAL METADATA

Geospatial metadata; a specific form of metadata, is applicable when objects have an explicit or implicit geographic extent. The Federal Geographic Data Committee (FGDC), USA has a good definition:-

'A metadata record is a file of information, usually presented as an XML document, which captures the basic characteristics of a data or information resource. It represents the who, what, when, where, why and how of the resource. Geospatial metadata are used to document geographic digital resources such as Geographic Information System (GIS) files, geospatial databases, and earth imagery. A geospatial metadata record includes core library catalog elements such as Title, Abstract, and Publication Data; geographic elements such as Geographic Extent and Projection Information; and database elements such as Attribute Label Definitions and Attribute Domain Values.'[47]

ISO

Many different metadata schemas exist specifically designed for digital objects. They can be general such as Dublin Core, or more specialised, but they are normally extensions to the Dublin Core schema.

Our review of best practice revealed a strong emphasis on geospatial metadata standards and the adoption of particular ISO standards to achieve this. By adopting an ISO standard users are able to know what to look for in the schema and are then better able to use the data, understanding its suitability and possible restriction.

The ISO 19100 is a series of standards for defining, describing, and managing geographic information. Standardization of geographic information can best be served by a set of standards that integrates a detailed description of the concepts of geographic information with the concepts of information technology. A goal of this standardisation effort is to facilitate interoperability of geographic information systems, including interoperability in distributed computing environments. From this

series one particular ISO metadata standard appeared to be almost universally recognized and adopted

ISO 19115 defines the schema for describing geographical information and associated services, including contents, spatial-temporal purchases, data quality, access and rights to use. The standard defines more than 400 metadata elements, 20 core elements. The ISO standards are revised and modified on a regular basis, ISO 19115:2003 is the current version.

The UK GEMINI project [representing the Office of e-Envoy, the UK data archive, and the Association of Geographic Information (AGI)] defined a metadata schema compliant with ISO 19115. It defined the following 29 elements to record metadata for geographic datasets:-

- *1 Title – name given to the data set*
- *2 Alternative title –Short name, acronym or alternative language title*
- *3 Originator – person or organization having primary responsibility for intellectual content*
- *4 Abstract – brief free text narrative summarising the dataset*
- *5 Date – date and time for the content of the dataset*
- *6 Frequency of update – frequency with which modifications / deletions are made*
- *7 Presentation type – mode in which the data is represented*
- *8 Access constraints – restrictions and legal prerequisites for the access of the data*
- *9 Use constraints- restrictions and legal restraints on using the data*
- *10 Topic category – main themes of the dataset*
- *11 West bounding coordinate – west limit of dataset, longitude decimal degrees, -ve west*
- *12 East bounding coordinate – east limit of dataset, longitude decimal degrees, +ve east*
- *13 North bounding coordinate – north limit of dataset, latitude decimal degrees, +ve north*
- *14 South bounding coordinate – south limit of dataset, latitude decimal degrees, -ve south*
- *15 Extent – extent of the dataset by subdivision of country e.g. admin area, postcode*
- *16 Spatial reference system – name or description of spatial referencing used in the dataset*
- *17 Spatial resolution- description of the spatial granularity of the data*
- *18 Supply media – type of media in which the data can be supplied*
- *19 Data format – format in which the data can be provided*
- *20 Additional information source – source of further information about the dataset*
- *21 Supplier – details of the organisation from which resource can be obtained*
- *22 Date of update of metadata – date on which metadata last changed*
- *23 Sample of dataset – location where a sample of the dataset is stored*
- *24 Dataset reference date – reference date for the dataset*
- *25 Dataset reference language – language used in dataset*
- *26 Vertical extent information – vertical domain of the dataset*
- *27 Spatial representation type – method used to represent the spatial aspect of the data*
- *28 Lineage – info about the events or source data used in the construction of the dataset*
- *29 Online resource – info about online resources from which resource can be obtained*[48]

For each of these elements precise details of what and how each element of the schema is to be recorded is listed in a sequence of annex tables, with the relationship to the ISO element shown, Table 3-6 shows the information for element 17, Spatial Resolution.

Metadata Element Name	*Spatial Resolution*
Definition	*Description of the granularity of the data*
Equivalent ISO 19115 Element number : name	*61 : MD_DataIdentification.spatialresolution>spatial resolution.distance*
Comparison with ISO 19115 element	*Equivalent*
Obligation (Mandatory or Optional)	*Optional*
Occurrence	*1*
Data Type	*INTEGER*
Domain	*Integer > 0*
Comment	*Equivalent to ground sample distance*

Table 3-6 Extract from UK GEMINI Annex A, 17 Spatial Reference details

Where applicable, information recorded into a schema should adopt the appropriate ISO standard to enable full integration with international standards. An example of this would be how we enter date information into the Gemini schema should be based on:-

ISO 8601 advises numeric representation of dates and times on an internationally agreed basis. It represents elements from the largest to the smallest element: year-month-day: YYYY-MM-DD[49]

INSPIRE (INFRASTRUCTURE FOR SPATIAL INFORMATION IN EUROPE)

Accepting the value of and necessity for ISO 19115 compliant metadata has become more significant following the implementation of the European Union INSPIRE directive.

THE NEED FOR INSPIRE

The general situation on spatial information in Europe is one of fragmentation of datasets and sources, gaps in availability, lack of harmonisation between datasets at different geographical scales and duplication of information collection. These problems make it difficult to identify, access and use data that is available, which in turn leads to ill-informed decision making.

Fortunately, awareness is growing at national and at EU level about the need for quality geo-referenced information to support understanding of the complexity and interactions between human activities and environmental pressures and impacts The INSPIRE initiative is therefore timely and relevant but also a major challenge given the general situation outlined above and the many stakeholder interests to be addressed.

INSPIRE is complementary to related policy initiatives, such as the Commission proposal for a Directive on the re-use and commercial exploitation of Public Sector Information.[50]

WHAT IS INSPIRE?

The INSPIRE Directive sets out to improve the efficiency and efficacy of public services – those associated with European environmental policy in the first instance – through the provision of a European spatial data infrastructure. INSPIRE is a Directive which mandates Member States to provide their public authority datasets and services so that they can more easily be used by other public organisations in the country concerned, in adjacent countries if required, and by the EC itself for policy making, reporting and monitoring. It is a set of principles and rules that each country must now choose how to implement - it will not necessarily need legislation.[51]

WHAT ORGANIZATIONS ARE EFFECTED?

The directive is specifically targeted towards public bodies (national and regional agencies) involved in environmental data. The INSPIRE website contains a database of Legally Mandated Organisations (LMO's), and this lists 6 Republic of Ireland organisations (160 are listed Europe wide):-

- Marine Institute *(http://www.marine.ie)*
- Environmental Protection Agency Ireland *(http://www.epa.ie)*
- Coastal & Marine Resources Centre *(http://cmrc.ucc.ie/)*
- Ordnance Survey Ireland (OSI) *(http://www.osi.ie)*
- Property Registration Authority of Ireland
- Department of Environment, Heritage and Local Government *(http://www.environ.ie/en/)*[52]

Although INSPIRE does not directly mandate commercial companies or non-governmental organizations, if OSi and DoEHLG adopt the standards of the directive then it will have become the de facto standard to which everyone should aspire. As stated the directive is focused primarily on environmental datasets but this can be extended and adapted to encompass cultural heritage data in the future.

INSPIRE IMPLEMENTATION

REGULATION../…/EC implementing Directive 2007/2/EC of the European Parliament and of the Council as regards metadata[53] specifies, in detail, the metadata requirement for INSPIRE compliance.

Category	Element	Short description
IDENTIFICATION	Resource title	characteristic and often unique name
	Resource abstract	brief summary of the content of the resource
	Resource type	type of resource being described
	Resource locator	link to additional information
	Unique resource identifier	value uniquely identifying resource
	Coupled resource	Identifies the target spatial data sets of the service
	Resource language	the language(s) used within the resource
CLASSIFICATION	Topic category	high level to assist in grouping and topic based searching
	Spatial data service type	to assist in the search of spatial data services

KEYWORD	Keyword value	commonly used word to describe the subject
	Originating controlled vocabulary	the citation for the controlled vocabulary
GEOGRAPHIC LOCATION	Geographic bounding box	the extent of the resource in geographic space
TEMPORAL REFERENCE	Temporal extent	time period covered by resource
	Date of publication	publication or entry date – could be both
	Date of last revision	date resource last revised, if ever
	Date of creation	date of creation of the resource
QUALITY & VALIDITY	Lineage	statement on process history / quality of data set
	Spatial resolution	level of detail of the data set
CONFORMITY	Specification	citation of implementing rules to which data conforms
	Degree	degree of conformity of the resource
CONSTRAINTS	Conditions of access & use	free text description
	Limitations on public access	free text – if none then entered as text anyway
ORGANIZATIONS	Responsible party	organisation responsible for establishment, management etc
	Responsible party role	the role of the responsible organization
METADATA	Metadata point of contact	Organization responsible for creating/ maintaining metadata
	Metadata date	when the metadata record was created or updated
	Metadata language	language in which the metadata elements are expressed

Table 3-7 The INSPIRE metadata elements, grouped by category

The INSPIRE metadata schema is compliant with ISO 19115 / 19119 containing 27 elements grouped into 10 broad categories, see Table 3-7

METADATA GEOPORTAL

INSPIRE provides a metadata editor (http://www.inspire-geoportal.eu/InspireEditor/) which allows users to either create or validate metadata records and then save them as xml files. The editor has a validation function which will which will display errors if mandatory elements are missing.

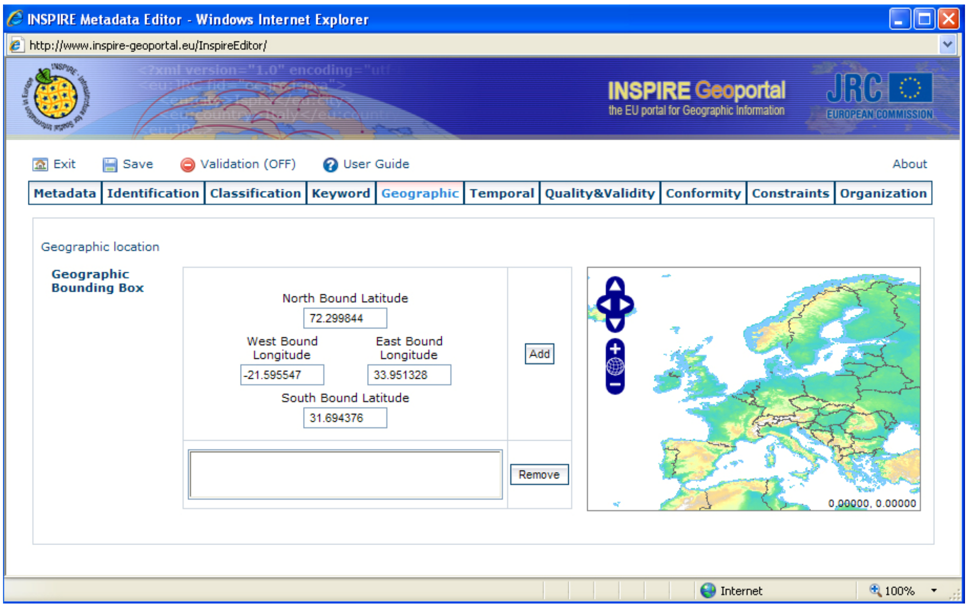

Figure 3-7 - The INSPIRE Geoportal showing the input TAB for Geographic metadata. The 10 tabs contain the other prescribed elements

INSPIRE DATA THEMES – GEMET THESAURUS

The INSIPRE spatial data themes are controlled by the GEneral Multilingual Environmental Thesaurus (GEMET) thesaurus, which defines core general terminology for the environment.[54] The current list of themes either includes our share-IT datasets directly, as in the case of orthoimagery, or indirectly in the case geophysical survey, or LiDAR digital elevation models.

CULTURAL HERITAGE INCLUSION IN METADATA

The adoption of the INSPIRE directive metadata standard, compliant with ISO 19115, will ensure the geographical description of our datasets is completed to an international standard. However, as INSPIRE is environmentally focused the schema has to be extended to acknowledge the cultural component which may accompany our data. Interpretations and classifications are often an integral part of our data.

Thesauri, or controlled vocabularies can be added to the Keyword component of the metadata schema. Controlling how the cultural component is described using these resources enhances the ability of users to search and retrieve our data in intelligent ways. More than one thesauri can be defined within a schema and our research identified a number which could be adopted.

EXAMPLES OF CULTURAL HERITAGE THESAURI / CONTROLLED VOCABULARIES

THE GETTY INSTITUTE

The Getty Research Institute is dedicated to furthering knowledge and advancing understanding of the visual arts. Its Research Library with special collections of rare materials and digital resources serves an international community of scholars and the interested public.[55]

It provides a number of thesauri which provide controlled vocabulary necessary in developing standardised languages.

1. Getty Thesaurus of Geographic Names Online (TGN)[56]
   This identifies 'place' based on hierarchal relationships, with the superordinate 'whole' and its subordinate 'members' or 'parts'. Table 3-8 shows an example of this hierarchy to define the geographic description of Newgrange.

Top of the TGN hierarchy (hierarchy root)

.... World (facet)

.... Europe (continent)

.... Ireland (nation)

.... Leinster (province)

.... Meath (county)

.... Newgrange (historic site)

Table 3-8 TGN hierarchal data structure, example of Newgrange (from http://www.getty.edu/research/conducting_research/vocabularies/tgn/)

The relationships in TGN include hierarchical (as Table 3-8), but also equivalence and associative relationships

2. Art & Architecture Thesaurus (AAT)
   This is a controlled vocabulary used for describing items of art, architecture, and material culture. This thesaurus is compliant with two further ISO standards:-
   ISO 2788 & ISO 5964 – both provide guidelines for establishing and developing monolingual thesauri. Table 3-9 shows the AAT hierarchical structure defining the term 'hillforts'.

Terms:

hillforts (preferred, C,U,D,American English-P)

hillfort (C,U,AD,American English)

forts, hill (C,U,UF,American English)

hill-forts (C,U,UF,American English)

hill forts (C,U,UF,American English)

Facet / Hierarchy Code: V.RK

Hierarchical Position:

Objects Facet

.... Built Environment (Hierarchy Name)

.... Single Built Works (Hierarchy Name)

.... <single built works (Built Environment)>

.... <single built works by specific type>

.... <single built works by function>

.... fortifications

.... forts

.... hillforts

Table 3-9 AAT extract for the term Hillfort (from: http://www.getty.edu/research/conducting_research/vocabularies/aat/)

CIDOC CONCEPTUAL REFERENCE MODEL CRM

CRM provides definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation...to promote a shared understanding of cultural heritage information by providing a common and extensible semantic framework that any cultural heritage information can be mapped to. It is intended to be a common language for domain experts and implementers to formulate requirements for information systems and to serve as a guide for good practice of conceptual modelling. In this way, it can provide the "semantic glue" needed to mediate between different sources of cultural heritage information, such as that published by museums, libraries and archives.[57]

CRM has been accepted as ISO standard 21127, guidelines for the exchange of information between cultural heritage institutions. In simple terms this can be defined as the curated knowledge of museums.[58]

CIDOC CRM is an extremely complex model for cultural objects and maybe something which could be adopted at a later stage. Initially this would be too complex to incorporate into a proposed metadata schema, which should be kept relatively simple, if we want to ensure it is completed by users.

ARCHAEOML

This is this XML schema of the University of Chicago Online Cultural Heritage Research Environment (OCHRE) user interface[59]. OCHRE is a web database system for research on cultural heritage making information accessible and searchable. The XML element hierarchies defined in ArchaeoML include archaeological descriptions consisting of observations about ancient landscapes (roads, canals, fields), settlement sites (architecture, stratigraphy, botanical and faunal remains), and artifacts (including the physical properties and contexts of inscribed artifacts).

MONUMENT INVENTORY DATA STANDARD (MIDAS) HERITAGE

MIDAS Heritage is the UK data standard for information about the historic environment, developed for and on behalf of the Forum on Information Standards in Heritage (FISH). It states what information should be recorded to support effective sharing of the knowledge of the historic environment, and the long-term preservation of those records. The structure of MIDAS is shown below, and its objective is to complement existing standards such as CIDOC CRM and UK GEMINI. The MIDAS data standard has a three level structure,

- Themes
- Information Groups
- Units of Information.

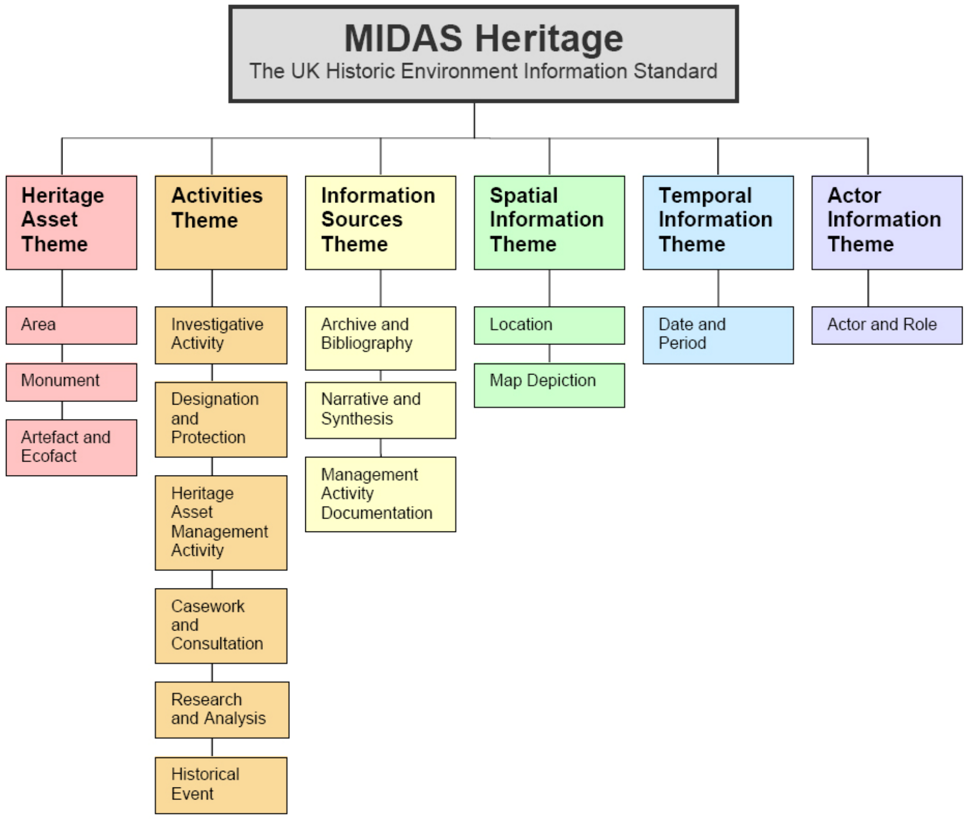

Figure 3-8 An overview of the structure of MIDAS heritage (from http://www.english-heritage.org.uk/upload/pdf/MIDAS_Heritage_Part_Two.pdf)

Figure 3-8 shows the six main themes with the associated information groups. Of particular relevance to the share-IT project is the Monument Information Group, which is part of the Heritage Asset Theme. The units of information designated mandatory for this group are defined in Figure 6.

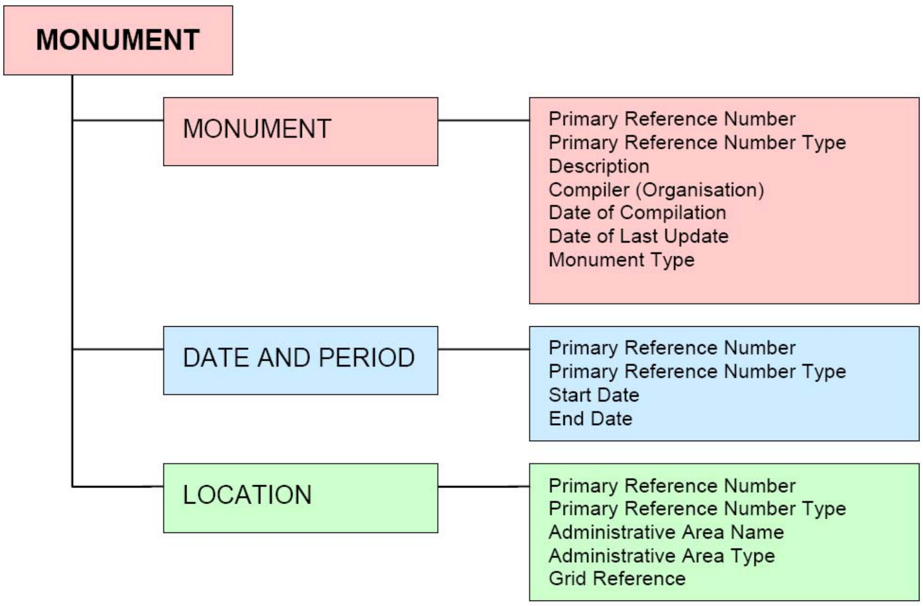

Figure 3-9 The mandatory units of information which are requires for the monument information group (from: http://www.english-heritage.org.uk/upload/pdf/MIDAS_Heritage_Part_Two.pdf)

FISH maintains an online collection of wordlists called INSCRIPTION[60] and they are organised by the MIDAS unit of information they relate to. For the example of *monument* four online wordlists are prescribed,

- Defence of Britain Thesaurus Operational
- Thesaurus of Monument Types Operational
- English Heritage Thesaurus of Maritime Craft Types Operational
- English Heritage Historic Aircraft Thesaurus

Incorporating wordlists from INSCRIPTION into an information system will improve standards of indexing and data retrieval.

For an information system to be said to be MIDAS Heritage compliant it has to have the functionality to store and export the mandatory units of information. However in the documentation relating to compliance it is made clear that it is an adaptable approach that is anticipated. MIDAS Heritage is seen as a set of closely integrated data standards rather than one single standard, and it is not expected that an information system would cover all the information groups.

FISH have developed an interoperability toolkit[61] to help with attaining MIDAS compliance, and attaining the objectives of sharing, archiving and migrating data between systems. The toolkit has three main components,

- **MIDAS XML** - The heart of the Toolkit is a W3C XML schema which provides a common format for the storage, processing and exchange of historic environment information.
- **Data Validator** - This online application validates the content of MIDASXML files.
- **Historic Environment Exchange Protocol (HEEP)** – A web services protocol that supports the querying of the MIDAS Heritage compliant information systems using the internet.

HUMANITIES AND SOCIAL SCIENCE ELECTRONIC THESAURUS (HASSET)

HASSET is a subject thesaurus which has been developed by the UK Data Archive (UKDA) over the past 20 years. Coverage is fuller in the core subject areas of social science disciplines: politics, sociology, economics, education, law, crime, demography, health, employment, and, increasingly, technology. These continue to be developed and are subject to addition and change as the holdings grow.[62]

IRISH CULTURAL HERITAGE CONTENT

The use of international thesauri provides a good standardised approach but this need to be supplemented to take account of the Irish context of the datasets. For this some de facto standards do exist which could be adopted such as the DoEHLG monuments database which contains terms for describing archaeological monuments.

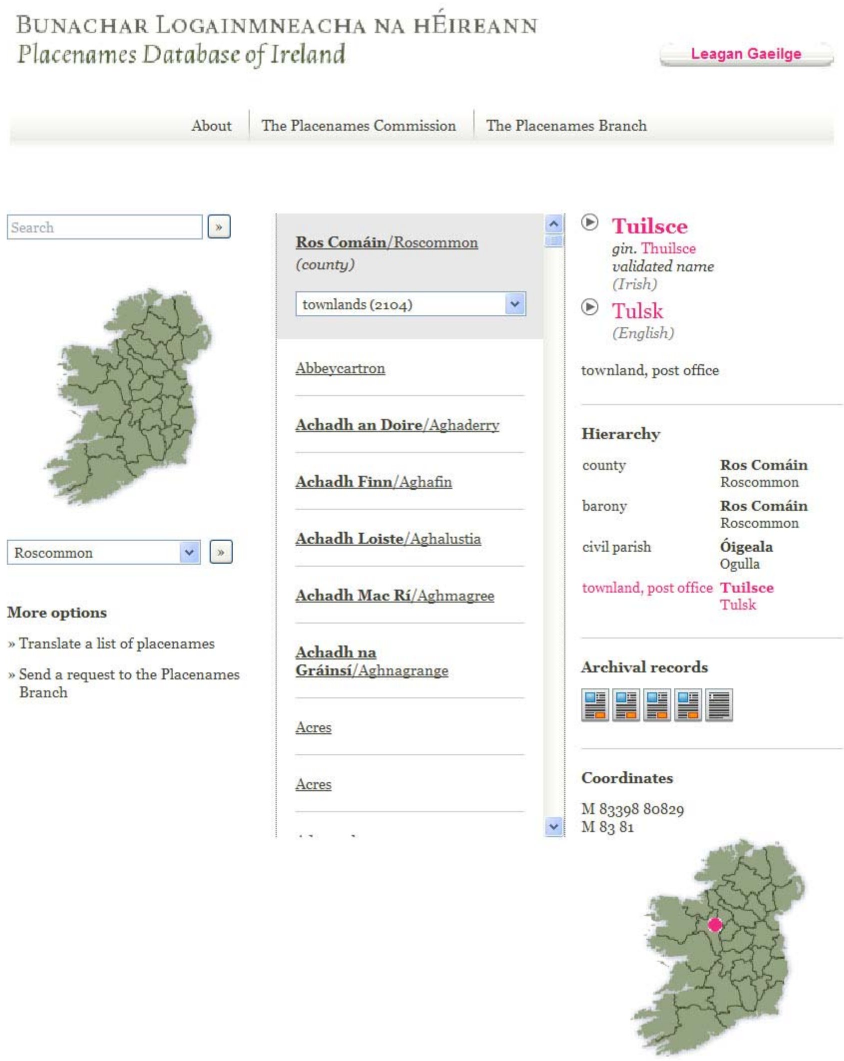

Figure 3-10 Placenames Database of Ireland showing the information available for the townland of Tulsk, Co Roscommon

For a controlled list of Irish place names the Placenames Database of Ireland [63] provides an excellent resource. This joint initiative between FIONTAR (DCU) and An Brainse Logainmneacha (Department of Community, Rural and Gaeltacht Affairs) has made available through the Internet, a database of Irish placenames that have been approved by the Placenames Branch, searchable under both Irish and English versions, see Figure 3-10.

It is important to remember that more than one controlled thesauri can be incorporated into the final schema, and that doing so will greatly improve the interoperability of the data. These provide a controlled environment to enable better access, searching and interrogation of the resource.

TOOLS FOR METADATA

There are a variety of free and commercially available tools to support metadata creation editing and validation.

INSPIRE GEOPORTAL

See section 6

ESRI

The ArcCatalog component of ArcGIS has a flexible metadata creation and viewing application. Data is input into the fields of a tab-based interface with mandatory fields indicated. Once created the metadata can then be viewed in the following range of international standard formats by selecting the appropriate stylesheet,

- Federal Geographic Data Committee (FGDC)[64]
- FGDC Classic
- FGDC ESRI
- FGDC FAQ
- FGDC Geography Network
- ISO
- ISO 19139
- ISO Geography Network
- xml

ISO METADATA EDITOR (IME) TOOLS

A number of IME application's can be freely downloaded from the internet. A good example is that available from INTA (National Institute for Aerospace Technology), Earth Observation Department (remote sensing area) in Spain. [65]

IME is an application focused on making it easier to understand and work with ISO19115 and ISO19139 standards, and validate the interoperability of xml files metadata.

It defines four steps to geographic Information metadata creation,

1. Profile definition
2. Metadata editor – according to the data type defined by ISO 19115

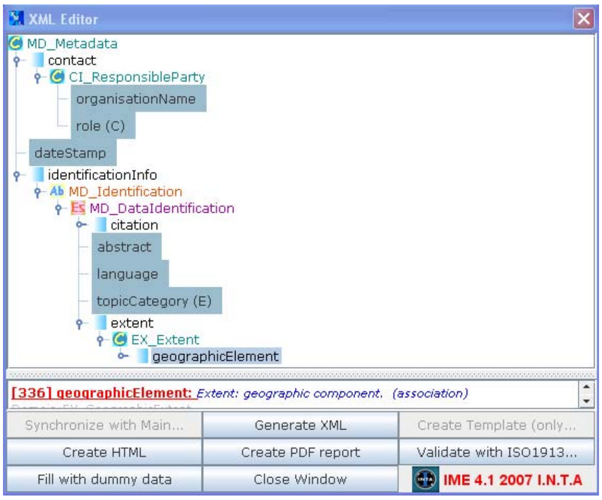

Figure 3-11 - IME 4.1 2007 I.N.T.A xml editor

3. XML generation – file generation according to the ISO 19139 schema
4. HTML generation – to facilitate the data visualization

METASCRIBE, NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION (NOAA)

This tool is designed to reduce significantly the effort required to produce metadata compliant to FGDC standards. It works on a template basis, taking advantage of the fact that records are generally very similar in content with only a few fields changing from one record to the next.

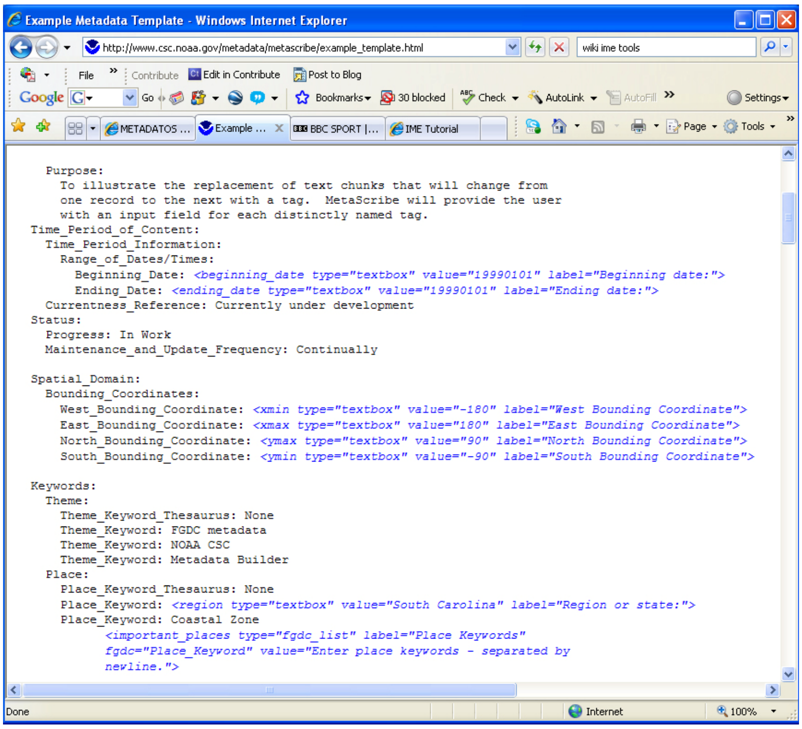

Figure 3-12 - Extract from a MetaScribe template (from, http://www.csc.noaa.gov/metadata/metascribe/example_template.html)

The user must create a metadata template which the MetaScribe website describes as 'not a trivial task'. The template is then proofed or validated by MetaScribe, with any errors reported. Once a valid template is created for a given data type, the user can create multiple records quickly and easily.

TK METADATA EDITOR (TKME)

Tkme is an editor for formal metadata which, as with MetaScribe, aims to ensure conformance with FGDC standard. [66]Its aim is to verify the syntactical structure of a file and then to re-express the metadata in various useful formats such as indented mp compatible text documents, SGML or XML.

REFERENCES

[1] Archaeology in Ireland: A Vision for the Future, 2006. RIA (http://www.ria.ie/committees/pdfs/archaeology/archaeology_forum_recommendations.pdf)
[2] Review of Research Needs in Irish Archaeology, 2007. The Heritage Council (http://www.heritagecouncil.ie/publications/Research_Needs/Research_Needs_in_Irish_Archaeology.pdf)
[3] The Heritage Council Strategic Plan 2007 – 2013 Consultation Document , 2006. Heritage Council.
[4] Archaeology 2020. Repositioning Irish Archaeology in the Knowledge Society. 2006, UCD (http://www.ucd.ie/t4cms/archaeology_2020.pdf)
[5] Sands, R, O'Sulliavan, A, and Kelly, E. P. 2006, Envisioning a Landscape: Investigating Lough Derravarragh, Co. Westmeth, Final Report (Heritage Council Grant 14505).
[6] http://www.ria.ie/policy/pdfs/RIA%20letter%20aerial%20photography.pdf
[7] http://www.ec-gis.org/inspire/directive/l_10820070425en00010014.pdf

[8] Condron, F., et al., *Findings and recommendations from Digital Data in Archaeology: A Survey of User Needs*. 1999, Archaeology Data Service

[9] The Share-it project is a collaborative undertaking involving a consortium from UCD, DIT, The Discovery Programme and Margaret Gowan & Co. For more information see http://www.share-it.ie

[10] Voss, J., *Tagging, Folksonomy & Co - Renaissance of Manual Indexing*, in *10thInternational SymposiumforInformationScience*. 2007: Cologne

[11] http://www.dlib.org/dlib/january00/01hodge.html

[12]Duncan H. Brown, 2007, *Archaeological archives: a guide to best practice in creation, compilation, transfer and curation*, Archaeological Archives Forum

[13] Jenny Mitcham and Julian Richards, 2008, *Digital Archiving at the Archaeological Data Service: a quest for OAIS compliance*, Archaeological Data Service, University of York

[14] Kevin Ashley, 2008, *The MS Office 2003 format debacle*. University of London Computing Scuance Centre digital archives blog (online at: http://dablog.ulcc.ac.uk/2008/01/11/the-ms-office-2003-format-debacle/)

[15] Martin Waller and Robert Sharpe, 2006, Minf d the Gap: Assessing digital preservation needs in the UK, Digital Preservation Coalition, York, p8

[16] http://ahds.ac.uk/creating/case-studies/newham/

[17] http://mida.ucc.ie/contents.htm

[18] http://erc.epa.ie/safer/

[19] http://www.life.ac.uk/2/documentation.shtml

[20] http://ads.ahds.ac.uk/project/bigdata/

[21] http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=24683

[22] Consultative Committee for Space Data Systems (2002). *Reference Model for an Open Archival Information System (OAIS).* CCSDS 650.0-B-1 Blue Book, Washington DC

http://public.ccsds.org/publications/archive/650x0b1.pdf

[23] Consultative Committee for Space Data Systems (2002). *Reference Model for an Open Archival Information System (OAIS).* CCSDS 650.0-B-1 Blue Book, Washington DC Page 3-1

http://public.ccsds.org/publications/archive/650x0b1.pdf

[24] Armin Schmidt 1998, Geophysical Data in Archaeology, ADS. *Guide to Good Practice Available online at:-* http://ads.ahds.ac.uk/project/goodguides/geophys/sect21.html

[25]http://www.remotesensing.org/geotiff/faq.html#What%20is%20GeoTIFF%20and%20how%20is%20this%20different%20from%20TIFF?

[26] http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=37674

[27] http://rii.ricoh.com/~gormish/pdf/dcc2000_jpeg2000_note.pdf

[28] http://en.wikipedia.org/wiki/JPEG#Recommended_usage

[29] http://www.opengeospatial.org/standards/gmljp2#overview

[30] 2005, Ron Lake, David Burggraf, Martin Kyle, Sean Forde, 2005, *GML in JPEG 2000 for Geographic Imagery (GMLJP2)Implementation Specification,* Open Geospatial Consortium Inc. Page 5

[31] http://www.asprs.org/society/committees/lidar/lidar_format.html

[32] http://prswww.essex.ac.uk/lidar/DAViS_UserGuide.html#B

[33] Hilary Beedham, Julie Missen,Matt Palmer, Raivo Ruusalepp, 2004, *Assessment of UKDA and TNA compliance with oais and mets standards*, UK Data Archive /the national archives, page 26 – 35. Available online at:- http://www.data-archive.ac.uk/news/publications/oaismets.pdf

[34] Hilary Beedham, Julie Missen,Matt Palmer, Raivo Ruusalepp, 2004, *Assessment of UKDA and TNA compliance with oais and mets standards*, UK Data Archive /the national archives, page 26 – 35. Available online at:- http://www.data-archive.ac.uk/news/publications/oaismets.pdf

[35] http://mida.ucc.ie/pages/dataPrinciples.htm

[36] http://ads.ahds.ac.uk/project/userinfo/charging.html

[37] http://ads.ahds.ac.uk/project/bigdata/final_report/bigdata_final_report_1.3.pdf (pages 25 -26)

[38] http://www.life.ac.uk/

[39] Lifecycle Information for E-literature, Available online at:_

http://eprints.ucl.ac.uk/1855/1/LifeProjSummary.pdf

[40] http://eprints.ucl.ac.uk/9313/

[41] http://eprints.ucl.ac.uk/9032/

[42] http://en.wikipedia.org/wiki/Metadata

[43] http://en.wikipedia.org/wiki/Dublin_Core

[44] http://dublincore.org/documents/dces/

[45] http://ads.ahds.ac.uk/project/goodguides/gis/sect54.html

[46] http://ahds.ac.uk/public/metadata/disc_09.html

[47] http://www.fgdc.gov/metadata

[48] Rob Walker, 2003, *ISO19115: Metadata standards and proposed element set,* UK GEMINI, GIgateway, London

[49] http://www.iso.org/iso/date_and_time_format#what-iso-8601-covers

[50] http://www.ec-gis.org/inspire/whyinspire.cfm

[51] http://www.agi.org.uk/SITE/UPLOAD/DOCUMENT/Policy/INSPIRE_Vision.pdf

[52] http://inspire.jrc.ec.europa.eu/ir/lmo_search_action.cfm

[53]http://www.irlogi.ie/userfiles/File/Database/IR%20metadata%20draft%20for%20consideration%20on%2014%20May%202008.pdf

[54] http://www.eionet.europa.eu/gemet/inspire_themes?langcode=en

[55] http://www.getty.edu/research/institute/

[56] http://www.getty.edu/research/conducting_research/vocabularies/tgn/

[57] http://cidoc.ics.forth.gr/

[58] http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=34424

[59] http://ochre.lib.uchicago.edu/index.htm

[60] http://www.fish-forum.info/i_lists.htm

[61] http://www.heritage-standards.org.uk/

[62] http://www.data-archive.ac.uk/search/hassetAbout.asp

[63] http://www.logainm.ie

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

[65] http://www.crepad.rcanaria.es/metadata/en/index_en.htm

[66] http://geology.usgs.gov/tools/metadata/tools/doc/tkme.html

[67] Longley, P., Goodchild, M. F., Maguire, D. and Rhind, D., 2005. Geographical Information Systems: Principles, Techniques, Management and Applications, John Wiley & Sons, New Jersey

[68] Peters D., 2008, Building a GIS: System Architecture Design Strategies for Managers, ESRI Press, California

[69] Nebert, D. ed., 2004. *Developing Spatial Data Infrastructures: The SDI Cookbook*. Available at: http://www.gsdi.org/docs2004/Cookbook/cookbookV2.0.pdf /p8 [accessed 18 December 2007].

[70] Nebert, D. ed., 2004. *Developing Spatial Data Infrastructures: The SDI Cookbook*. Available at: http://www.gsdi.org/docs2004/Cookbook/cookbookV2.0.pdf /p4 [accessed 18 December 2007].

[71] Peris, M.L. et al., 2005. *Protocolo para la generación de Metadatos Espaciales*. *Versión 1.0*. [Online]. Available

at: http://www.gvsig.gva.es/fileadmin/conselleria/images/Documentacion/protocolo/IDEGV-protocolo-metadatos-v1.0-es.pdf / [accessed 20 December 2007].

[72] After Nebert, D., 2004. Geospatial Data Catalogue – making data discoverable in *Developing Spatial Data Infrastructures: The SDI Cookbook*. Available at: http://www.gsdi.org/docs2004/Cookbook/cookbookV2.0.pdf

[73] Cartwright W. 2008. Delivering geospatial information with Web 2.0 in Lecture Notes in Geoinformation and Cartography: International Perspectives on Maps and the Internet ed. Michael P. Peterson, Springer, New York p11-28

[74] http://gis.cityofboston.gov/solarboston/

[75] http://resources.esri.com/arcgisexplorer/index.cfm?fa=home

[76] http://www.virtual.alabama.gov/,

[77] Terms and conditions for Internet License (Business Location Extracts), Ordnance Survey Ireland, 2002 available for download http://www.osi.ie/en/alist/copyright.aspx

[78] Consultative Committee for Space Data Systems (2002). Reference Model foe an Open Archival Information System (OAIS). CCSDS 650.0-B-1 Blue Book, Washington DC

[79] Peters D., 2008, Building a GIS: System Architecture Design Strategies for Managers, ESRI Press, California, USA, p 141