

Title: **Review of Metadata Activities**

Synopsis: This document provides an overview of metadata activities and achievements that occurred during the phase 2 demonstrator. This document discusses metadata standards for the Geo-data Browser. In particular, it reviews discusses the suitability of the NGDF Discovery Metadata Guidelines and proposes extensions to the standard for UK tertiary education and research.

Author: Tony Mathys, History Data Service and UK Data Archive

Date: 14 July 2003

Version: 1.0

Status: Final

Authorised:



1 Introduction

1.1 Purpose of document

The purpose of this document is to provide an overview and assessment of metadata activities and accomplishments associated with the current Phase 2 geo-data portal demonstrator project. A summary of activities includes the following:

- a choice was made between previously selected thesauri for the controlled subject vocabulary;
- a final version of FE/HE Metadata Application Profile of the NGDF Discovery Metadata Guidelines standard was adopted;
- a simple metadata tool was created to produce FE/HE compliant metadata records for the geo-data portal demonstrator;
- a comprehensive FE/HE Metadata Application Profile guidelines document was written;
- a sample of 100 metadata records were created for populating the geo-data portal demonstrator. Several organisations contributed their metadata records for this effort including the Archaeology Data Service (ADS), the History Data Service (HDS) MIMAS and EDINA;
- the FE/HE Metadata Application Profile was mapped to NGDF and FGDC geo-spatial standards to assure correct matches between that metadata records and portal results pages. Another mapping was done between the FE/HE Metadata Application Profile and the recently adopted ISO 19115 geo-spatial standard to facilitate any future transition likely to occur with the adoption of this standard in the United Kingdom;
- contacted academic departments and organisations to request metadata for the geo-data portal. Also received commitments from some to organise future metadata and portal workshops;
- distributed metadata and dataset inventory questionnaire to 16 organisations as an effort to assess dataset and metadata creation activities for a portion of UK academia;
- established contacts with the Association of Geographic Information (AGI) and the Office of e-Envoy as part of an effort to promote future collaborative efforts.

1.2 Target audience

This document is intended for those interested in the Geo-data Browser project including:

- Stakeholders invited to the geo-data portal project workshop on 28 January 2003
- JISC Geo-spatial Working Group
- JCEI Committee
- DNER Office
- JISC Secretariat

- JISC data centres and services, in particular, project associated partners, MIMAS and Archaeology Data Service
- National Geo-spatial Data Framework
- Development teams working on the Geo-data Browser project at EDINA within the Edinburgh University Data Library and at the History Data Service and UK Data Archive at the University of Essex

1.3 Status

This document is the final version.

1.4 Change summary

Version	Date	Author	Notes
0.1		AJM	Draft of document – submitted 21/07/03
0.2			
1.0			Final corrections.

1.5 Related documents

Title	Author	Date	Version
A geo-spatial data resource discovery tool for UK Further and Higher Education - Project Overview and Recommendations	David Medyckyj-Scott, Cressida Chappell, Anup Pradhan, Chris O'Hanlon	12/07/01	1.1
Geo-data Browser – Requirements Specification and Solution Strategy	Anup Pradhan and David Medyckyj-Scott	12/07/01	1.0
Geo-data Browser – Review of Controlled Vocabularies	Chris O'Hanlon	12/07/01	1.0
Geo-data Browser – Architecture	Anup Pradhan and David Medyckyj-Scott	12/07/01	1.0
Geo-data Browser – Geo-spatial Resources and Audit	Chris O'Hanlon, Anne Etheridge	12/07/01	1.0
Review of Metadata Standards	Chris O'Hanlon	12/07/01	1.0
Geo-spatial Projects – Glossary	Anup Pradhan, Anne Etheridge and Nat Evans	16/05/01	1.0

These documents are available on-line at:

Follow link to EDINA Projects Activity, <http://edina.ed.ac.uk>

Follow link to HDS Projects Activity, <http://www.data-archive.ac.uk/home/projects.asp>

1.6 Distribution

Open.

1.7 Contact details

Dr David Medyckyj-Scott

Hilary Beedham

EDINA
Data Library
Main Library Building
University of Edinburgh
George Square
Edinburgh EH8 9LJ
E-mail: d.medyckyj-scott@ed.ac.uk
Tel: +44 (0)131 651 1308

UKDA Projects Manager
The Data Archive,
University of Essex,
Wivenhoe Park,
Colchester CO4 3SQ
beedh@essex.ac.uk
Tel: +44 (0)1206 872 570

2 Overview of objectives

2.1 Review of deferred metadata initiatives from the phase 1 feasibility stage

Support for metadata creation is paramount to a successful portal service. This premise is well known and represents the catalyst for several metadata initiatives that were proposed during the Phase 1 feasibility stage and advanced for further development and completion during this current stage (Demonstrator Phase 2) of the project.

One of the feasibility studies entailed the draft of the FE/HE Metadata Application Profile of the (NGDF) Metadata Guidelines for UK academia (O'Hanlon, 2001a). The profile was based on the assessment and adoption of 42 elements and entities from the National Geo-spatial Data Framework (NGDF) Discovery Metadata Guidelines. Another 32 elements were also added to tailor the application profile for use among the further and higher education and the academic research communities. The draft FE/HE Metadata Application Profile of the (NGDF) Metadata Guidelines was subsequently forwarded to the current demonstrator phase for further evaluation and final approval.

Another focus during the phase 1 feasibility study (O'Hanlon, 2001a) was directed at the identification and selection of alternative controlled subject vocabulary candidates for the geo-data portal. The controlled subject vocabulary would also be employed to supplement the NGDF's Glgateway® Data Locator subject keywords, which provide limited coverage of the environmental and social sciences.

Controlled subject vocabularies provide a mechanism for supporting and improving information retrieval accuracy. A controlled subject vocabulary is a set of standard terms used to describe a data resource. They can be a simple list of defined terms from which a metadata creator chooses a suitable term. Alternatively, a controlled vocabulary might be a

complex thesaurus made up of hierarchical relationships and synonyms. Controlled vocabularies can be used by query engines to help users define their subject area as closely as possible, and in a way that is as consistent as possible. Query engines can also use an array of advanced techniques to improve searches on uncontrolled vocabularies by altering the way they look for different search terms.

The thesauri selected for further consideration were the GEMET, NASA, HASSET/ELSST and the UNESCO thesauri. The former two provide good representation of terms for environmental data resources and the latter two are noted for their representation of social science terms. The HASSET/ELSST thesaurus, developed by the UK Data Archive, was originally derived from UNESCO and has been modified to include terms relevant for the United Kingdom. It was also proposed during the feasibility study that a comparative assessment be carried out during the demonstrator phase of the project to select one of these thesauri for the controlled subject vocabulary.

2.2 Phase 2 demonstrator phase metadata objectives

As mentioned, the feasibility study found that work was required to identify which controlled subject vocabulary should be employed to be used for the geo-data portal during its demonstrator phase, and subsequently serve as a supplement to the NGDF's Glgateway® Data Locator subject keywords. A short comparison study would be carried out at the start of the project indexing a set of metadata records using GEMET, NASA, HASSET/ELSST and the UNESCO Thesauri, and comparing how well they describe subject terms used in content. The Archaeology Data Service (ADS), History Data Service (HDS) and MIMAS sites would be used as sources for selecting subject terms for the comparative study, and the thesaurus with the best representation of environmental and social science subject terms would be promoted for use to index the metadata.

Another priority carried over from the feasibility study for completion during the demonstrator phase is the adoption of a final version of FE/HE Metadata Application Profile of the NGDF Discovery Metadata Guidelines standard. This would entail a final review of the elements with the option to make additional modifications to improve mapping between standards and structure the profile format to provide greater clarity for users. Once a final version of the FE/HE Application Profile was adopted, a further step was taken to create a mock-up or template of the FE/HE Metadata Application Profile for the geo-data portal's metadata results pages.

New metadata-related initiatives were also introduced for and during the demonstrator phase of the geo-data portal in conjunction with the two aforementioned outputs from the phase 1 feasibility studies. In short, these new initiatives include the following:

- creation of comprehensive guidelines for the FE/HE Metadata Application Profile;
- cross-mapping between the FE/HE Metadata Application Profile and the NGDF Discovery Metadata Guidelines standard, the FGDC's Content Standard for Digital Geo-spatial Metadata (CSDGM) and the recently approved ISO 19115 geo-spatial metadata standard;
- compilation of a sufficient number of FE/HE compliant metadata records to populate the geo-data demonstrator portal. The metadata would describe and represent geo-spatial datasets taken from the FE/HE community. Existing resources within the FE/HE sector have been identified in the feasibility study and assessed for inclusion in the demonstrator (O'Hanlon & Etheridge, 2001). The metadata records would be used for portal demonstration purposes at workshops;

- creation of a simple metadata tool for converting FE/HE community geo-spatial metadata into the FE/HE Metadata Application Profile format for use on the geo-data demonstrator portal;
- work with the owners of data resources to agree access conditions for the distribution of both the metadata and the resources themselves, and to ensure the quality of the metadata created. For data which have no metadata, data holders would be assisted in the production of FE/HE Application Profile of the NGDF Discovery Metadata Guidelines compliant metadata;
- investigate and report on how users can be encouraged to provide metadata, how metadata can be kept up-to date and what kind of metadata quality assurance is required consider quality assurance and review policies for metadata records. This includes existing metadata records that are unlikely to have been created to the FE/HE Metadata Application Profile of the NGDF Discovery Metadata Guidelines, and given that a key objective of the project is to provide high quality and consistent information, it is imperative that proper quality control checks are put in place. Converted records will be checked against the standard to ensure that they meet the criteria for effective searching and retrieval and be improved where necessary.

3 Overview of activities

3.1 Finalisation of the FE/HE Metadata Application Profile of the NGDF Discovery Metadata Guidelines standard

The FE/HE Metadata Application Profile represents a template containing all the fields, or elements, required for each of geo-data demonstrator portal's metadata records. It is based on the UK National Geographic Data Framework (NGDF), the UK metadata standard for geo-spatial data. The FE/HE Profile borrows 36 elements present in the NGDF standard and augments these with an additional 34 elements to provide a total of 70 elements. These additions were introduced in order to address the needs of the academic community and improve mapping between the profile and metadata standards. The 70 elements are categorised and grouped under eight metadata major entities or headings (Table 1). These elements include 27 that are mandatory and represent information such as contact details or selections from lists. A complete version of the FE/HE Metadata Application Profile, which includes entity and element groupings, identifiers, definitions, obligations, data type and domain can be found in Appendix A.

Combined, these elements provide the academic user with the information necessary to assess the dataset type, its location, quality and completeness, temporal aspects in terms of data collection and content, and the necessary contact details to inquire further about access and use.

Table 1. FE / HE metadata application profile structure

METADATA ENTITY ELEMENT GROUP HEADING (Level 1)	METADATA ENTITY ELEMENT SUBGROUP HEADING (Level 2)	METADATA ENTITY ELEMENT SUBGROUP HEADING (Level 3)	METADATA ELEMENTS	METADATA ELEMENTS MANDATORY	METADATA ELEMENTS CONDITIONAL MANDATORY

Citation			1) Title 2) Alternative Title 3) Creator 4) Identifier* 5) Edition*	1) Title 3) Creator 5) Edition	
Identification Information (WHAT)			6) Topic* 7) Controlled Vocabulary* 8) Controlled Keywords 9) Other Keywords* 10) Controlled Place Name Vocabulary* 11) Controlled Place Name Keywords* 12) Description 13) Quality* 14) Language* 15) Further Information 16) Related Datasets	6) Topic 7) Controlled Vocabulary 8) Controlled Keyword 12) Description 14) Language	
	Spatial Reference System		17) Co-ordinate System** 18) Geographic Identifiers**		17) Co-ordinate System 18) Geographic Identifiers
	Level of Spatial Detail		19) Source Scale Denominator** 20) Imagery or Raster Cell/Pixel Size (X- Value)** 21) Imagery or Raster Cell/Pixel Size (X- Value)** 22) Smallest		

			Administrative Unit**		
Dataset Capture Period (WHEN)			23) Status of the Start Date for Data Capture 24) Start Date of Data Capture Process 25) Status of Completion Date for Data Capture 26) Completion Date of Data Capture Process 27) Update Frequency		
Time Period of Content (WHEN)			28) Start Date for Time Period Covered by Dataset* 29) End Date for Time Period Covered by Dataset*		
Spatial Extent of Dataset (WHERE)					
	Spatial Referencing Using Geographic Co-ordinates		30) System of Spatial Referencing by Co-ordinates	30) System of Spatial Referencing by Co-ordinates	
		Spatial Referencing Using Co-ordinates of a Bounding Rectangle	31) West Bounding Co-ordinate 32) East Bounding Co-ordinate 33) North Bounding Co-ordinate 34) South Bounding Co-ordinate	31) West Bounding Co-ordinate 32) East Bounding Co-ordinate 33) North Bounding Co-ordinate 34) South Bounding Co-ordinate	

		Spatial Referencing Using Co-ordinates of a Bounding Polygon	35) Spatial Referencing Using the Co-ordinates of a Bounding Polygon*		
	Referencing by Geographic Identifiers		36) Nations 37) Administrative Areas 38) Postcode Districts		
Custodian			39) Name of Custodian* 40) Postal Address of Custodian* 41) Postal City of Custodian* 42) Postal County of Custodian* 43) Postal Code of Custodian* 44) Postal Country of Custodian* 45) Telephone Number of Custodian* 46) Facsimile Number of Custodian* 47) Email Address of Custodian* 48) Web Address of Custodian*	39) Name of Custodian 40) Postal Address of Custodian 41) Postal City of Custodian 43) Postal Code of Custodian 44) Postal Country of Custodian	
Distributor (ACCESS)			49) Name of Distributor 50) Postal Address of Distributor 51) Postal Code of Distributor 52) Telephone Number of Distributor	49) Name of Distributor 50) Postal Address of Distributor 51) Postal Code of Distributor	

			Distributor 53) Facsimile Number of Distributor 54) Email Address of Distributor 55) Web Address of Distributor 56) Presentation Type 57) Data Format 58) Supply Media 59) Sample		
	Access and Use Constraints		60) Access Constraints 61) Access Details* 62) Use Constraints		
Metadata Creator			63) Name of Metadata Creator* 64) Postal Address of Metadata Creator* 65) Postal Code of Metadata Creator* 66) Telephone Number of Metadata Creator* 67) facsimile Number of Metadata Creator* 68) Email Address of Metadata Creator* 69) Web Address of Metadata Creator* 70) Metadata Last Update	63) Name of Metadata Creator 64) Postal Address of Metadata Creator 65) Postal Code of Metadata 70) Metadata Last Update	

* Denotes FE/HE element.

**Denotes modified NGDF elements.

3.1.1 Review of modifications made to the FE / HE Metadata Application Profile during the demonstrator phase of the project

As presented in Table 1, the Profile's structure is a departure from the NGDF Standard's format. The NGDF does not take full advantage of metadata entities or compounds as referred to under the FGDC's standard (CSDGM). It seems advantageous to use entities to establish structure and order to the profile to improve clarity and access for the user. The use of these metadata entity headings and subheadings provides the structure to arrange the elements into logical and user friendly groupings. This in turn directs users to navigate more quickly to the respective entity group and elements. This structure, in turn, also corresponds to the layout of the supporting guidelines documentation, thus making it easier and faster for the user to find relevant information and examples for each of the entities and elements.

The FE/HE Metadata Application Profile uses eight metadata entity headings to categorise and group the 70 elements. The metadata entity headings (level 1) and identifiers are as follows:

- **G1 Citation;**
- **G2 Identification Information (What);**
- **G3 Data Capture Period (When);**
- **G4 Time Period Covered by Dataset (When);**
- **G5 Spatial Extent of Dataset (Where);**
- **G6 Custodian;**
- **G7 Distributor (Access);**
- **G8 Metadata Creator.**

As a means to provide further clarification for metadata creators and editors, metadata entity subheadings were also created to separate and categorise elements into subgroups and place them under metadata entity headings and subheadings (levels 2 and 3).

- **G2 Identification Information (What)**
 - G2.sg1 Spatial Reference System
 - G2.sg2 Level of Spatial Detail
- **G5 Spatial Extent of Dataset (Where)**
 - G5.sg1 Spatial Referencing Using Geographic Co-ordinates
 - G5.sg1-a Spatial Referencing Using Co-ordinates of a Bounding Rectangle
 - G5.sg1-b Spatial Referencing Using Co-ordinates of a Bounding Polygon

- **G7 Distributor (Access)**

G7.sg1 Access and Use Constraints

In addition to this modification in the structure and layout of the FE / HE Metadata Application Profile, changes were also made to some of the elements that were first introduced during the phase 1 feasibility study (O'Hanlon, 2001a). These included several inclusions and exclusions of elements and modifications made to two NGDF elements and a FE/HE element.

The inclusion of new elements, in part, reflects the Association for Geographic Information's (AGI) intentions of modifying the NGDF Discovery Metadata Guidelines standard to conform with the recently adopted (March, 2003) ISO 19115 geo-spatial metadata standard. Other changes are presented and summarised in the table (Table 2) on the following page.

Table 2. Changes made to the proposed feasibility study's FE/HE Metadata Application Profile

Element or Entity Name	Action	Result	Reason
Topic	Added		To make FE/HE Profile ISO 19115 compliant.
Controlled Vocabulary	Added		Lists the thesaurus source used for controlled keywords.
Controlled Keywords	Added		Added to supplement the NGDF's Glgateway® Data Locator subject keywords, which provide limited coverage of the environmental and social sciences.
Other Keywords	Added		Allow additional subject keywords not found in controlled vocabulary.
Controlled	Added		Lists the gazetteer source

Place Name Vocabulary			used for controlled place name keywords.
Controlled Place Name Keywords	Added		Represents an additional source for identifying location of study area.
Postal County of Custodian	Added		Added as an option for users wishing to include county name for contact address.
Access Details	Added		Added for users to include text statements.
Audience	Dropped		Difficult to define audience types. Would need to establish a list.
<i>Bounding Rectangle</i>	Dropped		Not necessary as an entity under FE/HE profile structure.
<i>Dataset G-Polygon</i>	Dropped		Not necessary as an entity under FE/HE profile structure.
<i>Dataset G-Polygon Outer G-Ring</i>	Dropped		Not necessary as an entity under FE/HE profile structure.
<i>G-Point Ring</i>	Dropped		Not necessary as an entity under FE/HE profile structure.
G-Ring Latitude	Dropped and Replaced	Spatial Referencing using the Co-ordinates of a Bounding Polygon	Substituted it with G-Ring (Co-ordinates of a Bounding Polygon) to simplify for user.
G-Ring Longitude	Dropped and Replaced	Spatial Referencing using the Co-ordinates of a Bounding Polygon	Substituted it with G-Ring (Co-ordinates of a Bounding Polygon) to simplify for user.
<i>Dataset G-Polygon Exclusion Ring</i>	Dropped		Exclusion Ring not represented in ISO 19115.
<i>G-Ring Point</i>	Dropped		Exclusion Ring not represented in ISO 19115.
G-Ring Latitude	Dropped		Exclusion Ring not represented in ISO 19115.
G-Ring Longitude	Dropped		Exclusion Ring not represented in ISO 19115.
G-Ring	Dropped		Exclusion Ring not represented in ISO 19115.

Time Period Covered by Resource	Modified	- Start Date for Time Period Covered by Dataset - End Date for Time Period Covered by Dataset	Modified to include start and end dates to account for intermittent periods.
Spatial Reference System	Modified	- Co-ordinate System - Geographic Identifiers	Modified to separate one element into two elements for purposes of presentation simplification and more options for users. Also improved interoperability between standards.
Level of Spatial Detail	Modified	- Source Scale Denominator - Imagery or Grid Raster Cell/Pixel Size (x-value) - Imagery or Grid Raster Cell/Pixel Size (y-value) - Smallest Administrative Unit	Modified to separate one element into four elements for purposes of presentation simplification and more options for users. Also improved interoperability between standards.

Italicised words represent metadata entities.

Several entity names were also changed to provide greater clarity for users.

- Archival Responsibility* was changed to ***Custodian***
- Supplier* was changed to ***Distributor***
- Person(s) Responsible for Update of Metadata* was changed to ***Metadata Creator***

The latter entity faced some scrutiny and additional suggestions. Some other titles included Meta -Metadata, Metadata Editor and Metadata.

In summary, this FE/HE Metadata Application Profile introduces the necessary structure and element representation to prompt and enable metadata creators and editors to produce quality metadata. The profile will also benefit geo-data portal users because it provides the necessary information to answer their questions about the content, context and structure of a dataset.

3.2 Selection of thesaurus for controlled subject vocabulary

A two-stage assessment procedure was employed to select a thesaurus for the controlled keyword vocabulary source. First, a spreadsheet was created with a list of 93 subject terms

(see Appendix B). There were 46 social science terms selected from the Archaeology Data Service (ADS), English Heritage, History Data Service and MIMAS web sites. The web sites for British Geological Survey, British Oceanographic Data Centre, Centre for Ecology and Hydrology and Environmental Information Centre represented sources for 30 of the terms used for the environmental terms. Several additional environmental terms were also taken from the ADS and MIMAS web sites. These UK-based web sites seemed appropriate sources for finding relevant terms for a representative assessment of terms because each one houses metadata or dataset information for the social sciences, environmental and geo-spatial disciplines. In addition, 17 general terms were also included with the 30 environmental terms. Many of these pertained to geo-spatial terminology. The first stage of this assessment procedure would lead to the selection of two thesauri for the second stage procedure, which entailed a comparative review of thesauri and relevant UK census terms. There were 208 census terms selected for this final assessment and the sources for these terms were ONS and HDS.

The first stage of assessment revealed that the NASA thesaurus was inadequate as a controlled vocabulary source. There were fewer terms found in the NASA thesaurus (Table 3) and it did not contain British spellings for any of the terms.

The HASSET thesaurus fell short. Its social science and UK representation were quite strong, but selecting HASSET alone would have been inadequate for the environmental and physical sciences.

GEMET, as expected, covered environmental and physical sciences well. It also had an adequate coverage of social science and geo-spatial terms as well. Its representation qualified it for further consideration.

Table 3. Summary of representative comparison of terms with thesauri

	Social Science (yes)	Environment (yes)	Total	Social Science (no)	Environment (no)	Total	Difference
GEMET *(8)	31	36	67	11	7	18	+49
HASSET *(9)	36	16	52	4	28	32	+20
NASA *(9)	20	26	46	24	14	38	+8
UNESCO *(1)	40	28	68	6	18	24	+45
	Census (yes)	Census (no)		Census, Environment and Social Science (yes)	Census, Environment and Social Science (no)	Difference (all terms)	
GEMET *(23)	38	147		105	165	-60	

UNESCO *(23)	79	100		147	124	+23
------------------------	----	-----	--	------------	-----	------------

* denotes the number of similar terms associated with that thesaurus.

The UNESCO thesaurus probably offered the best balance in terms of representation for both the social sciences and environment. This thesaurus was also considered for further evaluation against GEMET.

Census terms were used as a final measure between the two thesauri. A considerable number of UK geo-spatial datasets are derived from census data. Based on the comparisons made in the first and second stages of this exercise, the UNESCO thesaurus offered better term representation and selected on this basis and the results of the first stage to represent the source for the FE/HE Metadata Application Profile's controlled keywords.

In summary, this assessment also suggests that a combination of two thesauri might be more suitable in the future for the FE/HE Metadata Application Profile and the UK academic geo-spatial community. As seen with this comparative analysis, GEMET, HASSET and UNESCO provide good overall representation; however, HASSET and UNESCO provide excellent coverage for the social sciences and census data and GEMET's terms serve the environmental and physical sciences very well. A combination of GEMET with HASSET or UNESCO might be the best possible option, with HASSET's representation of the UK giving it an advantage over UNESCO.

Other options would be for the geo-spatial community to develop its own thesaurus through a licensing arrangement between the aforementioned thesauri intellectual property owners, or create a thesaurus as an independent effort. Either approach would improve the overall coverage of terms across most disciplines and eliminate the need to evaluate and select or re-evaluate thesauri for their controlled vocabularies, especially as editors update and edit their respective thesauri.

3.3 Guidelines for the FE/HE Metadata Application Profile

The Go-Geo! FE/HE Metadata Application Profile Guidelines (Appendix C) have been written to serve as a reference resource for individuals and organisations wishing to document and create metadata records for their geo-referenced datasets. Use of the guidelines can also assure that compliant metadata can be created for the Go-Geo! portal site or to aid in the management of internal datasets. The Go-Geo! FE/HE Metadata Application Profile Guidelines also represent a critical resource for maintaining metadata quality assurances and currency.

The document comprises over 140 pages of detailed reference information for the metadata creators, editors and Go-Geo! portal users. It is organised in a manner that provides a one-stop source of reference for users. The document holds detailed information and examples for each of the FE/HE Metadata Application Profile elements and entities. Some complex elements are also linked to supplemental example pages which contain comprehensive explanations and example graphics to provide greater clarity for the user. Many of the examples are presented in a manner that can be understood across various disciplines such as Archaeology and Anthropology, Biological Sciences, Civil Engineering, Earth Sciences, Environmental Sciences, Geography, Health Sciences, History and Sociology.

3.4 Template for the Go-Geo! geo-data demonstrator portal's metadata results page

As part of the web design effort for the Go-Geo! portal, a mock-up metadata records web page (Appendix D) was created and sent to EDINA. The design layout presents the metadata record using a folder format (Figure 1). A tab on each folder is labelled with a FE/HE Metadata Application Profile's metadata entity name or generic term that corresponds to each entity. Each folder contains elements that are associated with that FE/HE Metadata Application Profile entity and the elements are revealed when a folder tab is mouse clicked.

- **What** corresponds to the **Identification Information** entity
- **Where** corresponds to the **Spatial Extent of Dataset** entity
- **When** refers to the **Data Capture Period** and **Time Period Covered by Dataset** entities
- **Access** corresponds to the **Distributor** entity
- **Citation** corresponds to the **Citation** entity
- **Creator** corresponds to the **Metadata Creator** entity

Figure 1. Screen shot of the metadata record display screen

What	Where	When	Access	Citation	Metadata
soil survey natmap100					
Access					
Distributor: soil survey and land research centre (sslrc)					
Address of Distributor: cranfield university, barton rd, silsoe, bedfordshire					
Postcode of Distributor: mk45 4dt					
Tel: 01525 863263					
Fax: 01525 863253					
Email: sslrc@cranfield.ac.uk					
URL: http://www.cranfield.ac.uk/sslrc/					
Access Constraints: legal					
Access Details:					
Use Constraints: use of natmap100 is subject to a data lease agreement which regulates the use and reproduction of the data.					
Presentation Type: map numeric					
Data Format:					
Supply Media: url magnetic optical					

Record 8 of 82

Given space limitations on the metadata record's display web page, the *Custodian* entity and its elements are not available for view on the screen. The *Custodian* information can be viewed when the metadata record is printed to hardcopy or sent as an email.

3.5 Metadata acquisition and creation for the geo-data demonstrator portal

Metadata acquisition was established as a priority during the demonstrator phase to provide content for the Go-Geo! portal. The metadata records would serve to demonstrate the search capabilities of the Go-Geo! portal to stakeholders. The metadata records would also be created in the FE/HE Metadata Application Profile format.

A form letter requesting metadata was written and sent to 13 selected individuals. The letter also drew attention to the project web pages and supporting information that would assist them with metadata creation. All were members of academia across various disciplines, institutions and data centres. The disciplines included Archaeology, Biological and Environmental Sciences, Geography and the Geosciences and Health. This list also included individuals from England, Northern Ireland, Scotland and Wales. The data centres included the Archaeology Data Service (ADS) and MIMAS. Internal requests were also made to the History Data Service (HDS) and EDINA.

One source used to locate these individuals was the Geo-data Browser Geo-spatial Resources and Audit report (O'Hanlon and Etheridge, 2001), which was conducted during the phase 1 feasibility study. Because some of the contact information in the report was dated and no longer valid, additional searches were conducted to find additional contacts.

The email letter was sent with a subsequent follow-up for those who didn't respond, of which there were eight. The five that did respond offered various levels of interests and support. One of those interested expressed concern about distribution and access and unwilling to create or provide metadata until these issues were resolved. These concerns pertain licensing issues surrounding data derived from the Ordnance Survey. Other responses acknowledged interest in providing metadata, but at a latter date. The time wasn't right because of other pressing obligations.

Cooperation was received from ADS, EDINA, HDS and MIMAS. ADS, HDS and MIMAS would allow the conversion and use of their metadata for purposes of demonstrating the Go-Geo! portals capabilities to stakeholders. It was agreed that these metadata records would be removed if the portal were to become a service. EDINA agreed to provide 75 metadata records for their in-house datasets. In total, 100 metadata records have been created or converted for use in the Go-Geo! portal.

3.6 Tool for creating and converting metadata into a FE/HE profile compliant format to populate the Go-Geo! demonstrator portal

A simple metadata tool (Appendix E) was developed for creating and converting metadata records into a FE/HE compliant format for the Go-Geo! demonstrator portal. The tool was developed using spreadsheet and word processing software. Basic instructions were also provided for metadata creators.

Subsequently, the tool was used for converting metadata records that ADS, HDS and MIMAS had provided for the portal. EDINA used a metadata tool developed within database software and made some modifications to make the tool FE/HE Profile compliant format. The EDINA software engineers also created an import tool to convert the files into a portal compatible format.

3.7 Cross-mapping between the FE/HE Metadata Application Profile and the NGDF Discovery Metadata Guidelines standard, the FGDC's Content Standard for Digital Geo-spatial Metadata (CSDGM), the ISO 19115 geo-spatial metadata standard and the History Data Services (HDS) DDI profile

Cross-mapping was conducted between the FE/HE Metadata Application Profile and the NGDF Discovery Metadata Guidelines standard, the FGDC's Content Standard for Digital Geo-spatial Metadata (CSDGM), the ISO 19115 geo-spatial metadata standard and the History Data Services (HDS) DDI. These were done to assess compatibility, improve interoperability and provide cross-searching capabilities.

The anticipation of the AGI's plans to modify the NGDF Discovery Metadata Guidelines standard to conform to the ISO 19115 geo-spatial metadata standard led to a cross-mapping between the FE/HE Metadata Application Profile and ISO. Because the profile is derived in

part from the NGDF, this entailed mapping both to the ISO. At this point, it was also established that FE/HE Metadata Application Profile would be modified to conform to the relevant elements of the ISO standard.

In addition to this exercise, a mapping was done between the History Data Service's DDI profile and the FE/HE Metadata Application Profile. This was done so that the Go=Geo! portal could cross-search the HDS collection's database for metadata records.

Another mapping had to be conducted between the FE/HE Metadata Application Profile and the FGDC's Content Standard for Digital Geo-spatial Metadata (CSDGM) as well. The EDINA software engineers encountered problems with the earlier mappings between NGDF and the FE/HE Metadata Application Profile. The NGDF was originally mapped to a GEO Profile of the FGDC's standard. This mapping was problematic, so the alternative and successful solution to improving interoperability was to provide a direct mapping between the FE/HE Metadata Application Profile and the FGDC's standard and bypass the NGDF's mapping.

3.8 Metadata and dataset inventory questionnaire

As part of an update process to the geo-spatial resources audit conducted during the phase 1 feasibility study (O'Hanlon and Etheridge, 2001), a metadata and dataset inventory questionnaire was produced for distribution at the Geographical Information Systems Research-UK (GISRUK) conference that was held in April 2003 at City University, London. The idea behind the questionnaire was to collect current information about metadata and dataset creation activities at various universities across the UK. Because most of the conference attendees were affiliated with academia, this venue provided a good opportunity to distribute the questionnaire, and project brochure, plus provide individuals a brief overview of the geo-data portal project.

In all 16 individuals were approached, each affiliated with a different university. The reactions were of general enthusiasm and considerable awareness of the project, but unfortunately only two questionnaires were returned. Cartographic theme cards and follow-up emails were sent out to the people contacted at the conference, but only one response was received from the remaining 14 contacts.

The initial effort made to contact people about metadata and this follow-up exercise involving the questionnaire reveal the many challenges that lie ahead. Rather than considering these efforts as shortcomings, it is important to recognise that there are issues and problems associated with metadata creation. Understanding this now can pave the way towards identifying the concerns and constraints that affect people's attitudes towards metadata and addressing these with innovative and motivational solutions.

3.9 Metadata creation strategies, acquisition, currency and quality assurances

As stated, it needs to be emphasised that geo-spatial metadata creation and acquisition represent challenges that need to be addressed with a range of strategies that support and motivate academia's geo-spatial community. As it stands, this has been achieved with only limited success and is not a ubiquitous practice. Thus, it remains a fact that numerous undocumented geo-spatial datasets exist within the academic community and this poses the risk of undocumented dataset files becoming misplaced and made redundant or requiring considerable time and cost for reassessments to determine the content, context and structure of a dataset.

The absence of a formal documentation system also means that datasets cannot be acquired and revealed to the wider geo-spatial community; therefore, limiting other

researchers' access to the original dataset developers as well as to the datasets themselves. It also prevents other potential users from augmenting or complementing their own datasets for various applications. The absence of information about existing datasets can lead other organisations to expend considerable time and costs in producing data that are already in existence, but at an undisclosed location.

3.8.1 Importance of supporting metadata creation

These possible conditions make it necessary to establish a support system that reaches out to the academic geo-spatial community. The support system must use a range of strategies that make dataset developers and users aware of the importance of dataset documentation. First, these audiences must be introduced to metadata, metadata standards and guidelines. Next, the case must be made to motivate data developers to create dataset documentation or metadata for existing material where possible and for new material as routine. Motivation represents the greatest challenge and it requires various approaches to address this attitude.

Three main arguments can be made for motivating individuals and organisations to create metadata:

1. metadata creation can provide security and structure for internal dataset management;
2. it can provide an opportunity to locate and access external datasets that can be used to augment or complement existing datasets;
3. it can be seen as a way of marketing the dataset, even to recover some costs incurred in creation.

The benefits of metadata creation can be introduced in a number of ways to the dataset developers and users. The case made for internal management reflects on several key points:

- metadata creation helps to establish an operational structure for managing and storing dataset information for departmental and intra-departmental use;
- implementing a metadata strategy assures the integrity of existing and new datasets as metadata can be used as a tracking mechanism to monitor changes and edits to datasets;
- it also, of itself, provides an inventory of datasets to reduce redundancy and the time required to reassess existing datasets for new and future applications;
- metadata creation eliminates or reduces the risk of redundancy in data collection or deletion of existing datasets;
- the disruptive effects of staff turnover are minimised, especially for those members working independently on data capture projects;
- the introduction of metadata ensures that a standard is in place so that the same information about every dataset is captured.

All these benefits of metadata creation ultimately protect the investments of time and cost dedicated to data development.

Another approach that can be considered deals with the release of metadata to the geo-spatial academic community. This is an important benefit that needs to be stressed:

- it helps to ensure that other organisations will not replicate data at added cost and time;
- it contributes to the research base by providing potential users with the means to select datasets to complement or augment their existing in-house datasets;
- it could contribute to the widening of the research agenda by providing potential users with a dataset catalogue from which to view the wider picture of datasets which already exist (the discovery of a dataset may instigate the user to develop an idea for a new application);
- the portal could serve as a repository for organisations to manage their metadata and as an internal resource to access and share datasets, saving an organisation the time and costs associated with establishing and maintaining internal servers to store, manage and update metadata;
- the metadata and portal will complement and augment other UK academic portals within the JISC IE as well as the national UK Glgateway® site.

3.8.2 Strategies for metadata motivation and metadata acquisition

Strategies must be employed to initiate and sustain metadata creation activities. They represent a combination of support mechanisms, ranging from indirect contact to outreach programmes that entail workshops, and collaborative efforts.

These are the following recommendations for metadata support and acquisition:

- if available, offer a portal as a resource for storing an organisation's metadata;
- organise travelling workshops around the UK, within universities, to promote metadata creation, metadata profiles and guidelines (the workshops would also be used to introduce attendees to portal technology: the attendees would be asked in advance to bring information about a dataset so that metadata could be created as part of the demonstration process);
- make proactive requests for metadata creation through letters, follow-up email requests and visits to academic departments, organisations and conferences;
- provide on-line assistance and guidance through metadata standards, expanded metadata guidelines on the web site, and, possibly, metadata creation tools, with other *ad hoc* assistance for the creation of metadata also offered. Metadata creation tools should provide XML export capabilities to facilitate transfer and acquisition of metadata files;

- provide interactive support for metadata creation through offers to supply metadata template forms¹;
- provide metadata quality assessments as part of creation support, with follow-up interviews to assure accuracy of records;
- set up meetings with key geo-spatial academics and decision makers to motivate and stimulate interest in metadata;
- promote the creation and support of metadata at key GIS-related conferences, such as, at least, GISRUUK, the Association for Geographic Information and BURISA;
- explore the options for using externally produced tools (for the purposes of this work, the working prototype developed during the demonstrator stage would be used, however; if a tool is created, it will be made to meet minimal data input requirements);
- encourage editors to publish metadata in relevant journals and magazines - this would recognise and acknowledge the metadata authors' contributions, thus providing an additional incentive for metadata creation;
- continue the work begun during the demonstration project to create as large an inventory as possible of the number and types of datasets that exist in academia, and the distribution of the datasets among the disciplines; a questionnaire would also include a dataset needs request to determine which datasets are in great demand;
- establish a mindset that encourages present and future metadata creation by encouraging the importance of metadata support among academics and students, especially those of the generation that has worked with geo-spatial systems over recent years.

These are the critical activities and recommendations that must be addressed and take place if metadata creation is to be sustainable in UK academia. Initial efforts to encourage metadata creation during the current demonstrator phase fell short and reveal a daunting challenge for the future, which will be addressed during the next phase of this project. Many of the aforementioned recommendations cited for motivating metadata creation will be employed and evaluated to assure that long term strategies can be put into place to continue any momentum gained during the phase 3 metadata initiative effort.

3.8.3 Metadata currency and quality assurances

The use of the FE/HE Metadata Application Profile and supporting guidelines provide important first steps towards establishing assurances and confidence in metadata currency, and content and structural accuracy. The FE/HE Metadata Application Profile introduces standardisation, which assures accuracy based on consistency. The FE/HE Metadata Application Profile can be used across various disciplines, while maintaining the same

¹ A metadata template typically includes contact information in a metadata file. Once created, this information can be transferred through to successive metadata files that are created for different datasets within the same organisation.

requirements and structures that assures continuity. The supporting metadata guidelines are written in a manner that provides users clarity regardless of their training and backgrounds. Combined, the use of these two resources can instil a considerable level confidence and reliability in the integrity of metadata.

Both the FE/HE Metadata Application Profile and supporting guidelines should be made available on a web site to facilitate access for potential metadata creators and editors. Ideally, access to these resources should be made available on the geo-data portal, with links between the metadata results web pages and the relevant profile and guideline document pages.

Development of a user-friendly metadata creation tool is also critical for addressing currency and accuracy assurances. Ideally, the FE/HE Metadata Application Profile's supporting guidelines would be made accessible to a user through an interface. Quick access to the guidelines should incline the tool's user to reference them frequently in a similar manner as might be expected with a "help" button on a menu.

Metadata workshops can provide important venues to introduce the FE/HE Metadata Application Profile, the supporting guideline documentation and a metadata creation tool. The introduction of the latter could also include the opportunity for workshop participants to use the tool to create metadata. Participants might be expected to collect relevant dataset information prior to the workshop and use it to create metadata as part of a metadata creation tool exercise. These types of workshop activities should establish important working practices that will improve the likelihood that quality metadata will be produced.

Technical solutions should also be employed to assure metadata integrity. A metadata creation tool offers the potential for providing accurate metadata through imposing standardisation and detection mechanisms. The structural design of the tool should make it possible for it to detect the presence or absence of information in element fields, especially those fields that are mandatory or conditional. Spelling and date format checkers and access to lists would also improve metadata output, which would pass through a metadata parser and subsequently filtered for accuracy and FE/HE Metadata Application Profile compliance. If the metadata output detects problems, the metadata creator or editor would receive a warning and error messages. The output process would also reveal which fields have problems and provide a navigation mechanism that would direct the user to each problem field.

An additional concern that should be addressed with a technological solution is the extraction of geographical co-ordinates for the bounding rectangles or polygons of a study area. This can also be a time-consuming process, so the creation of a co-ordinate extraction tool would be important not only in terms of saving a metadata creator or editor time, but also improve the accuracy and currency of the metadata records.

These aforementioned approaches represent important first steps and bases towards assuring metadata currency and accuracy, especially for individuals and organisations wishing to control maintenance and distribution of their metadata. These combinations of resources and support offer the best opportunities for metadata creators and editors to produce compliant, accurate metadata. Currency would also be addressed under this scheme because any time saved in the dataset documentation process would encourage individuals and organisations to create and update their metadata.

Centralised systems of metadata collection are evident in UK academia and other public institutions, and as structured and managed, offer resources for assuring metadata currency and quality. Assuming similar practices, if geo-spatial metadata were to be stored at a centralised site, then metadata review procedures and practices could be put in place to

assure quality control. Such practices might include tracking and reviewing metadata, verifying integrity or compliance of metadata and providing back-up and archival schemes for metadata. These review and maintenance procedures probably represent the most reliable measures to assure metadata accuracy and currency, but entail considerable resources to meet these objectives.

Another realm is a decentralised system based on node technology. The responsibilities of metadata quality and assurances rest in the hands of disaggregate groups of individuals and organisations, who collectively might represent a consortium of regional organisations or institutions. Under this scenario, an organised, co-ordinated effort might be employed to assure that the uniform interoperability structure and standardisation systems are in place, and relevant staff are given the necessary resources and training to create quality metadata. The former may rely on the type of system that is in place, whether it be a closed, intra-departmental network, or an external, public domain portal or gateway. Closed or open, metadata accuracy and currency play vital roles towards instilling confidence in the information and data that are exchanged between organisations associated with a decentralised system.

4.0 Summary and Recommendations

The phase 2 metadata activities were successfully completed and met the requirements of the stated objectives. A thesaurus was selected as a vocabulary source for controlled keywords, a final version of the FE/HE Metadata Application Profile was produced, a comprehensive FE/HE Metadata Application Profile Guidelines document was written and efforts were made to secure and provide metadata records for the Go-Geo! demonstrator portal. Though overall response to a request for metadata fell short, those that did provide metadata made it possible to successfully populate the portal with 100 records. A metadata tool was also developed for converting the metadata records into the FE/HE Metadata Application Profile.

In addition, a design template was created for the portal's metadata records display page. Mapping was also done between standards to address and improve interoperability, access future compatibility with the ISO 19115 and cross-search the History Data Service's catalogue. Finally, efforts were made to approach colleagues in the UK geo-spatial communities. Within academia, the approach entailed requests for information pertaining to their metadata and dataset creation activities. This effort falls under earlier attempts to encourage academics to create and submit metadata for the Go-Geo demonstrator portal. These efforts fell short, but the experience still provided an invaluable insight into challenges that lie ahead. One issue that comes to mind is the access and distribution rights for Ordnances Survey dataset users.

At this stage of the project, the resources have been developed and put into place as support mechanisms for metadata creation. Furthermore, these resources can be used to promote metadata and data management to individuals and organisations across the multiple disciplines within the UK academic community. This can be accomplished through the organisation of workshops, which can serve as venues for introducing these resources and explaining the benefits of metadata for both the data users and data developers.

As part of this approach, it is critical to develop a better, user-friendly metadata creator tool. The tool created for this phase of the project is limited in every respect and not a conducive solution to metadata creation. There are initiatives under way that may bring metadata creation tool technology to the UK geo-spatial community. One effort includes the UK Data Archive and EDINA, both of whom are involved with testing a beta version of a JAVA-based

tool that will be released later in 2003. A German group is also developing a JAVA-based metadata tool that will be public domain. If either tool can be modified to conform to the FE/HE Metadata Application Profile format, then this will be an important step towards metadata creation for the UK academic geo-spatial community.

Finally, it is recognised that collaboration with the broader UK geo-spatial community is critical for the FE/HE geo-data portal project. As more data become available, and can be discovered through portals and gateways, the apparent benefits, and peer pressure, start to influence all sectors of the community, including academia. The Association for Geographic Information (AGI) serves and works with the government and private sectors in the UK. There could certainly be times when collaborative projects and initiatives might benefit both the FE/HE portal project and the AGI. For these reasons, we've established contact with the AGI. This contact has included meetings and the current agreement to test the beta version of a metadata tool. It might also be possible to encourage or support other activities in the future. As mentioned earlier in this report, there is a need for a comprehensive controlled vocabulary source for the geo-spatial community.