



National Geospatial Data Framework

**Discovery Metadata
Transfer Format and
Communications
Protocol
Guidelines**

Version 1.2

www.ngdf.org.uk

RESPONSIBILITY FOR THIS DOCUMENT

National Geospatial Data Framework (NGDF) Management Board

DISTRIBUTION

This document is available free of charge to any persons seeking guidelines for documenting their georeferenced data resources provided due acknowledgement is made to the National Geospatial Data Framework (NGDF) Management Board. This document may be reproduced but only in its entirety and only if the publication details on this page are also produced.

COPYRIGHT

©National Geospatial Data Framework, 1999, 2000. All rights reserved.

DISCLAIMER

None of the statements made in this document should be considered an endorsement or otherwise of any of the commercial or non-commercial products which are cited. The statements made in this document are for guidance only and users should seek professional advice before acting on this document. The National Geospatial Data Framework Management Board assumes no responsibility for any errors or omissions or for damages resulting from the information contained in this document. The reliability and the currency of the metadata remain the responsibility of the supplier of the metadata.

TRADEMARKS

NGDF is a trademark of the National Geospatial Data Framework and may not be used without the written consent of National Geospatial Data Framework Management Board.

PUBLISHED BY

National Geospatial Data Framework (NGDF) Management Board
c/o Ordnance Survey
Maybush, Romsey Road
Southampton, UNITED KINGDOM, SO16 4GU

CHANGE HISTORY

Version	Date	Summary of change
1.2	April 2000	First issue as separate document extracted from NGDF Discovery Metadata Guidelines version 1.1 with the following modifications: Modification to the XML specifications in Section 3.11 Changes to XML tags to be compliant with GEO profile and existing Z39.50 software (Section 3.11) Detailing of issues for implementing NGDF Guidelines in the GEO Profile (Section 4.7) Modified the NGDF mapping to the GEO Profile (Section 4.7)

Note: this document is liable to change at any time without notice and should always check the NGDF website: www.ngdf.org.uk to ensure that the version is current.



CONTENTS

1. Introduction	5
1.1 Purpose of Guidelines	5
1.2 About this document	5
1.3 The National Geospatial Data Framework	6
1.4 Metadata	6
1.5 Comments on these Guidelines	9
1.6 Glossary of Terms & Abbreviations	9
1.7 References	9
2. Scope and Application of the Guidelines	10
3. NGDF METADATA Transfer Format	11
3.1 Introduction	11
3.2 The Purpose of a Transfer Format	11
3.3 Additional Resources	11
3.4 XML Document Type Definition	11
3.5 Defining Metadata Elements in the DTD	12
3.6 Specifying the XML Version	12
3.7 Referencing the NGDF Metadata DTDs	12
3.8 An Example of XML Encoding	13
3.9 Specifying Controlled Attributes	13
3.10 Entity Declarations	14
3.11 DTD for NGDF Discovery Metadata	14
4. NGDF Communications Protocol	17
4.1 Introduction	17
4.2 Use of Z39.50	17
4.3 How it works	17
4.4 Z39.50 Terminology	18
4.5 The Geospatial Profile	18
4.6 The ‘NGDF Profile’ – A subset of GEO	20
4.7 NGDF Discovery Metadata mapped to GEO	21
5. CONFORMANCE	24
5.1 Conformance Requirements	24
5.2 Conformance	24

5.3	Abstract Tests	24
5.4	Compatibility of Data Transfer Format	25
5.5	Compatibility of Z39.50 Protocol	25
<i>APPENDIX A - MAPPING TO XML</i>		27
<i>APPENDIX B - XML EXAMPLE</i>		29

1. INTRODUCTION

1.1 Purpose of Guidelines

These Guidelines need to be read in conjunction with the NGDF Discovery Metadata Guidelines. The two documents represent the first stage in the development of metadata services for the discovery of data resources that have a geographic component. They have been developed as part of the National Geospatial Data Framework (NGDF). Their aim is to provide a consistent and simple method of documenting any data resources that are referenced in some way to the earth's surface whether by coordinates or geographic identifiers (addresses, administrative area, component of postcode). The consistent recording of metadata or information about data resources and their presentation in catalogues accessible to the user community via the Internet considerably facilitates the discovery of such data resources.

It is estimated that 80% of information has some relationship to the earth's surface or, to put it another way, it is *geospatially referenced*. These Guidelines have been designed to be usable with any data resources that have some sort of geospatial reference whether these are maps, imagery, textual or numeric.

There are a number of metadata standards in existence or under development. None of those existing were found to meet with the requirements of being simple and applicable to the full range of data resources that are geospatially referenced. In developing these Guidelines metadata standards which are not specifically for geospatial data such as the "Dublin Core" [1] have also been examined.

These Guidelines have been produced for use in the United Kingdom. However, with the evolution of global information infrastructures, developments outside the UK cannot be ignored. Therefore the Guidelines must be regarded as interim pending the development of standards at the international level such as those being developed by Technical Committee 211 of the International Organisation for Standardisation (ISO).

The Guidelines will be reviewed and enhanced in the light of comments received from interested parties and developments in international standards.

1.2 About this document

Working Group 2 of the NGDF Task Force has prepared this document. The Working Group is responsible for the facilitation of NGDF compliant metadata services. It is made up of representatives from major vendors, national mapping agencies, government departments and members of the academic community.

The document is in three Sections and has two Appendices. The remainder of the introduction provides a brief background to the NGDF and describes what metadata and the proposed NGDF metadata services mean. A transfer format for the NGDF Discovery metadata is defined in section 3 and a communications protocol in section 4. Conformance requirements are given in section 5.

Appendix A provides a mapping to XML and Appendix B gives a worked example in XML.

1.3 The National Geospatial Data Framework

More than ten years ago, the Report of the Committee of Enquiry into the Handling of Geographic Information [3], chaired by Lord Chorley, established that the use of different geospatial data for the same area was crucial but difficult because of:

- data inconsistencies
- poorly documented datasets
- incompatible computer systems

In the UK today, over 40 government departments and other organisations produce geospatial data for their own needs, spending approximately £400 million annually on data collection and assembly.

Much of this data is not available for reuse or is collected in ways that make it difficult to use with other datasets. Most datasets remain poorly documented. There are still inconsistencies in the ways different computer systems treat sophisticated geospatial data.

The National Geospatial Data Framework is designed to reduce these problems and ultimately to:

- increase the use of geospatial data
- facilitate development of markets for data and services
- “future proof” investments and enhance decision-making through use of better information

1.4 Metadata

1.4.1 The Requirement for Metadata

If geospatial datasets are to become more widely available and their true value realised by the community, then methods must be found to enable the public to discover their existence and determine if a particular dataset is appropriate for their needs.

This means that information about the data must be made available. Metadata is the term used to define this “data about data”. The metadata itself has to be easily accessible and provided in such a way that the public can obtain information that will allow them to compare the suitability of data from different sources. It is therefore necessary for there to be standard ways of presenting metadata that will allow these comparisons to be made.

With the exponential growth in the use of the Internet and its adoption as a service within organisations (Intranets) a means is now available to use World Wide Web based search tools to locate sources of geospatial information. This increases the importance of providing standard methods for describing data that can be analysed by computer based systems.

Metadata content should vary according to purpose, for example:

- discovery of data resources
- access to data resources
- transfer of data

Discovery metadata provides sufficient information to enable an inquirer to ascertain the existence of data that is fit for purpose and to reference some point of contact for more information. If after discovery more detail is needed about individual datasets, then more

comprehensive and more specific metadata is required. If the data is transferred as a single dataset then quite specific and detailed metadata is needed possibly down to the feature, object or record level.

Thus, not only can metadata content vary according to purpose; it can also vary according to scope of the data being defined. Discovery metadata usually, but not exclusively, relates to collections of data resources or dataset series that have similar characteristics but relate to different geographic extents or times. A map series is the commonest example but it can equally be applied to statistical surveys. More detailed metadata may be applied to a collection or series but may apply to an individual dataset (e.g. one map tile). Transfer metadata applies exclusively to that transfer.

For successful discovery of data resources which are geospatially referenced the following are also essential:

- thesauri as a source of keywords and synonyms
- look-up tables of standard geographic units e.g. counties, districts, post-code districts preferably with a cross-reference to coordinates

The Guidelines set out in this document relate to discovery metadata only. The use of a thesaurus is recommended.

The rest of this section describes how NGDF metadata services might look and then goes on to explain discovery metadata, detailed metadata and thesauri in more detail.

1.4.2 NGDF Metadata Services

These Guidelines have been developed on the assumption that metadata services will be implemented as Internet/Intranet applications and that a gateway infrastructure will be supported to provide a national focus and to encourage compliance to the guidelines. A diagrammatic illustration of the way the services is to be structured is given in Figure 1 below.

The NGDF Gateway supports routing facilities to other service providers and Data Producers and a Metadata repository for those data providers unwilling or unable to support their own service. Service providers are free to support enquires from sources other than the gateway and may not even be NGDF compliant. The Central Gateway, will itself provide support and encouragement to adhere to the standards and provide a central focus, where minor data providers may chose to register their metadata. Helpdesk facilities and impartial advisory services may also be developed.

It is the intention of NGDF that all data flow through this system should be achieved using XML as a transfer file format and Z39.50 as the basis of a communications protocol. The application of XML and Z39.50 by NGDF is described in sections 3 and 4.

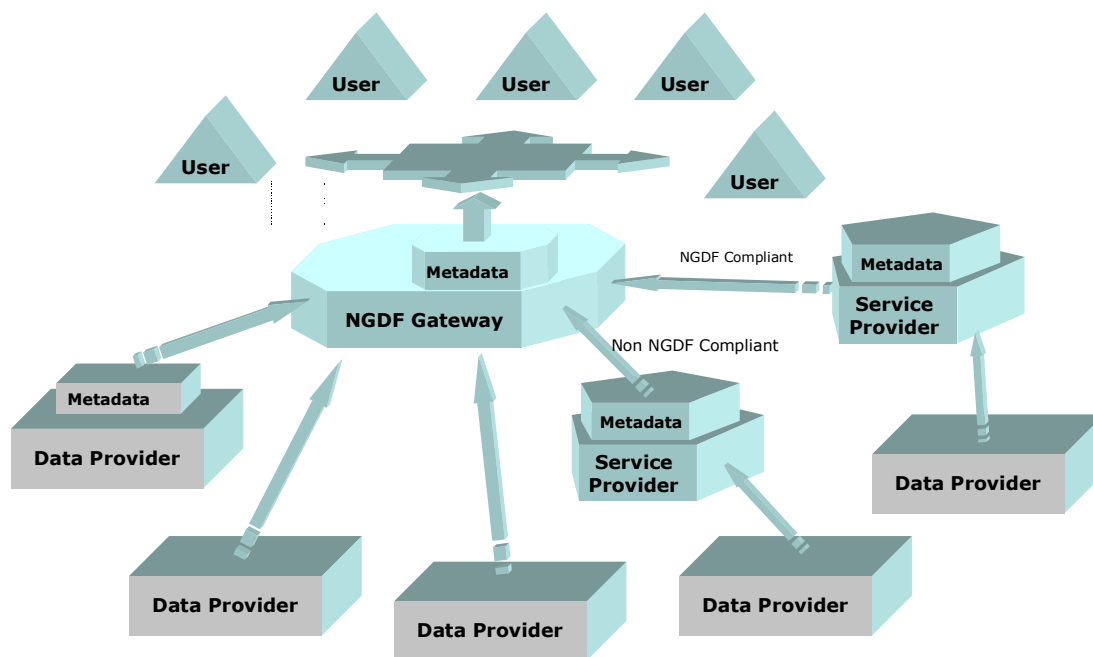


Figure 1 - NGDF Gateway Metadata Infrastructure

1.4.3 Discovery Metadata

Discovery Metadata is the minimum amount of information that needs to be provided to convey to the inquirer the nature and content of the data resource. The requisite information falls into the following broad categories;

- identification
- subject matter
- originator
- access use and restrictions
- time of capture
- geographic extent
- data supplier

Computer databases accessible on the Internet that contain metadata entries will be the first point of call for people searching for specific datasets. Indeed for some datasets the Discovery Metadata may be all that is needed for the inquirer to locate and obtain the data needed.

The provision of a Discovery Metadata component also recognises the considerable amount of work that some data producers may be required to undertake in order to prepare a detailed metadata definition. It is intended that there will always be a direct mapping between Discovery Metadata and Detailed Metadata. This will ensure that no discrepancies exist between the two definitions and that the Discovery Metadata can be automatically extracted from the Detailed Metadata thus saving time and effort.

A concise listing of Discovery Metadata is given in The NGDF Discovery Metadata Guidelines.

1.5 Comments on these Guidelines

Users of the Guidelines are encouraged to send any comments concerning their scope and applicability to the NGDF Working Group 2 at ngdfwg2@ordsvy.gov.uk. Or through the NGDF web site (www.ngdf.org.uk).

1.6 Glossary of Terms & Abbreviations

Data collection	Data captured to a common specification or to a common standard
Dataset	Identifiable collection of data
Dataset series	Data captured to a common specification or to a common standard
Discovery Metadata	A subset of a dataset's metadata as defined by these Guidelines
Geospatial data Metadata	Data defined by reference to a position on the Earth's surface
Metadata element	Concise descriptions of a data resource
NGDF	An individual characteristic of a data resource
NTF	National Geospatial Data Framework
Tag	BS7567 National Transfer Format
URL	XML notation to signify the beginning or end of an element data entry. The syntax for a tag is <tagname> for a start tag and </tagname> for an end tag.
W3C	Uniform Resource Locator
www	The World Wide Web Consortium
XML	World Wide Web
Z39.50	Extensible Markup Language, a subset of SGML (Standardised General Markup language)
	Protocol to enable querying of multiple remote heterogeneous databases.

1.7 References

- [1] Dublin Core, Online Computer Library Centre (OCLC) (URL: purl.org/metadata/dublin_core). (Note: there is an underscore “_” between dublin and core.)
- [2] ISO/TC 211 Working Draft 15046-5 Geographic Information – Conformance and Testing (URL: www.statkart.no/isotc211/).
- [3] Handling Geographic Information, Report of the Committee of Enquiry chaired by Lord Chorley, 1987, published by Her Majesty's Stationary Office, London. ISBN 0 11 752015 2.

2. SCOPE AND APPLICATION OF THE GUIDELINES

These guidelines define an XML based transfer format and a communications protocol based on Z39.50 to be used in conjunction with the NGDF Discovery Metadata Guidelines.

Only datasets that are geospatially referenced to the United Kingdom may be documented using these Guidelines although the principles are applicable world-wide. Although these Guidelines are primarily intended for use with digital data resources they may be applied to other types of data resources such as those that are paper-based.

The guidelines do not constrain the manner in which a user interface may be designed or data stored. For example the guidelines specify that dates should conform to the format CCYYMMDD for transfer purposes, a user interface may wish to enable a user to enter the date in a more readable form such as DD-MMM-CCYY and hold the date in a commercial database format, neither of these options is excluded by the guidelines. The guidelines merely state that when date information is transferred between parties it should be converted to the format CCYYMMDD.

Although these Guidelines contain certain information about implementation, they provide no specific guidance on metadata creation, metadata quality, metadata maintenance or the use of Internet search engines.

3. NGDF METADATA TRANSFER FORMAT

3.1 Introduction

The NGDF Metadata transfer format sets out a definition for how the encoding of the 42 metadata elements in the Draft Guidelines for NGDF Discovery Metadata should take place. The transfer format utilises the syntax and features of Extensible Markup Language (XML) Version 1.0 as defined by the World Wide Web Consortium (W3C). Extensible Markup Language is a restricted form of SGML, the Standardised General Markup Language. SGML is an ISO standard for marking up documents that has been widely used in government and industry. XML documents are conforming SGML documents.

3.2 The Purpose of a Transfer Format

The purpose of the transfer format is to enable the conveyance of NGDF compliant metadata to take place between data suppliers and service providers. The transfer format may also be used for the onward transmission of metadata between any other parties and could be used in certain circumstances as a means of transferring metadata between service providers and end users, although the normal protocol here will be through Z39.50. The transfer format is largely intended for the transmission of whole, or substantial portions of, metadata sets.

3.3 Additional Resources

The transfer format is predicated on a Document Type Definition (DTD) that is available via the World Wide Web. Users of the format should note from the example that they will be required to reference this document in any discovery metadata descriptions that are coded using the transfer format. Details of how this should be done are given in Appendix A.

3.4 XML Document Type Definition

This section defines the XML Document Type Definition (usually abbreviated to DTD) for the NGDF Metadata transfer format. A Document Type Definition is a formal explanation of how the components that make up an XML document are defined and structured. It is expressed in a machine-readable form as a series of declarative statements using a simple syntax.

The DTD defines how the elements in a NGDF Metadata document fit together, which ones are compulsory and which ones are not, how many times different elements can be referenced and what data can be entered against each metadata element. A more comprehensive explanation of DTD formatting is given below. The DTD is read and used by indexing programs to check that all the required elements are present and correctly ordered. It can also be used by display programs like web browsers to present the XML information to the user.

3.5 Defining Metadata Elements in the DTD

Each of the 42 NGDF metadata elements is formally defined as an ELEMENT in the DTD. An XML element has a name and a formal specification of the type of information that it can hold. For example, the following line shows the XML DTD element definition for the Title element of NGDF Metadata.

```
<!ELEMENT title (#PCDATA)>
```

This statement informs us that the name of the element is “title”, and that the value it can take is any set of free text. The term “PCDATA” stands for “Parsed Character Data”.

In some cases, The NGDF Discovery Metadata Guidelines specify that the possible values that an element can take are restricted to a controlled set. The System of Spatial Referencing by Coordinates (Element 17) can take only three possible values : “British National Grid”, “Irish National Grid” and “Latitude and Longitude”.

In the DTD, this restriction is encoded using an attribute list. Attribute lists define a prescribed set of values that the XML element can take. The following example shows the prescribed set of attributes for NGDF Metadata Element 12, Access Constraints.

```
<!ATTLIST access  
    accrestrict (Financial|Legal|Other|Not Known|None) >
```

This statement tells us that the ELEMENT called “access” will be assigned one of five possible values from an attribute list called “accrestrict” instead of being defined in its own right.

3.6 Specifying the XML Version

All XML documents must include a statement defining which version of XML has been used to encode the document. Presently, there is only one version of XML, release 1.0. The statement to specify this XML version would be encoded thus:

```
<?xml version="1.0"?>
```

3.7 Referencing the NGDF Metadata DTDs

XML documents that are compliant with this transfer format will be required to utilise the customised tag names, attribute definitions and syntax that are defined in the DTD. Users will therefore need to include a reference to the DTD in each of their Discovery Metadata XML documents. This can be achieved two ways :

By including a pointer in the Discovery Metadata XML document to the URL for the online version of the DTD. This is available on the NGDF web server at www.ngdf.org.uk/ngdfdm2.dtd. An XML capable browser will then reference this document and use it to read the Discovery Metadata.

By downloading the DTD and storing it on a local server. The local copy of the DTD can then be used to read the discovery metadata.

3.8 An Example of XML Encoding

Figure 1, below, gives a simple illustration of how an XML document is constructed. It shows how the Document Type Definition reference and XML version are encoded, and the basic syntax that NGDF Discovery Metadata XML documents use.

```
<?xml version="1.0"?>
<!DOCTYPE NGDFDiscoveryMetadata SYSTEM "www.ngdf.org.uk/ngdfdm2.dtd">
<NGDFDiscoveryMetadata>
  <abstract>
    large scale digital topographic mapping for site location, planning
    and land and asset management. Land-Line contains 30 feature
    codes and six text categories. Land-Line.Plus has additional
    landscape features represented by a further 27 feature codes.
    Standard map tile coverage 500 x 500 m in urban areas, 1 km x
    1 km in rural areas and 5km x 5km in moorland areas.
  </abstract>
  <syscoord system="britishNationalGrid" />
</NGDFDiscoveryMetadata>
```

Figure 1

All NGDF Metadata documents must include two lines at the beginning specifying the XML version and the source DTD that has been used for encoding. In Figure 3, line one specifies the XML version that has been used while line two specifies the Document Type Definition that this document uses and gives its location in a machine and human-parseable manner.

The metadata itself consists of a list of elements, which may be composed of other elements. An NGDF Discovery Metadata XML document is defined as including a range of 42 possible elements, which occur in a specific order. Some of these elements are compulsory. Some can occur more than once. Each element is represented in the XML document by a start tag, the data contained in the element, and the end tag. The format for this is as follows:

```
<element name> (The start tag)
... Information contained in the element ....
</element name> (The end tag)
```

Where an element is composed of other elements (like the Discovery Metadata as a whole), the format for encoding is as follows :

```
<element name> (Start Tag)
<sub element name> (Start Tag)
... Sub element data ...
</sub element name> (End Tag)
<sub element name> (Start Tag)
... Sub element data ...
</sub element name> (End Tag)
</element name> (End Tag)
```

Figure 1 shows how this structure is applied. In a valid NGDF Metadata document, all of the necessary sub elements would have to be included for the document to be acceptable.

3.9 Specifying Controlled Attributes

In some cases, the Guidelines state that the possible entries relating to an element are restricted to a particular set of responses. The Document Type Definition uses an Attribute List to define what these are. In the NGDF Metadata Guidelines, the System of Spatial

Referencing by Coordinates (Element 17) can take only three possible values : “British National Grid”, “Irish National Grid” and “Latitude and Longitude”.

It should be noted that in an attribute definition, there is no requirement for start and end tags for the element that the attribute describes. This is because the element itself is declared as EMPTY. In an XML document, the value of the attribute is defined as follows :

```
<element_name attribute_name="attribute_value" />
```

Here, the start and end tags are replaced by the syntax <element_name ... />. The ‘/’ symbol effectively replaces the end tag.

Figure 1 shows how a value of “British National Grid” for Element 17 would be encoded according to the Discovery Metadata DTD.

3.10 Entity Declarations

The DTD contains formatting information for symbols that might occur in an NGDF Discovery Metadata document. This information is encoded using ENTITY references which occur at the beginning of the DTD prior to the encoding of elements and attributes.

3.11 DTD for NGDF Discovery Metadata

This part of the document presents a formal XML 1.0 Document Type Definition for NGDF Discovery Metadata. This DTD forms the core of the transfer standard.

```
<!DOCTYPE NGDFDiscoveryMetadata [  
<!-- NGDF Discovery Metadata DTD 1.2 20000209
```

This is the Document Type Definition for discovery metadata conforming to the UK National Geospatial Data Framework Discovery Metadata Guidelines. This DTD corresponds to Version 1.1 (March 1999) and 1.2 (October 1999) of the Standard. The DTD is for use with XML 1.0 compliant viewers and word processors.

Three special characters in the DTD are defined as follows :

+ - This element must be present and can occur one or more times.

? - This element can be absent, and can occur only once if present.

* - This element can be absent, or can occur one or more times.

Where these special characters are absent, the element must be specified and can occur only once.

The values of metadata elements that are not compound are here declared as PCDATA following the example of the US Federal Geographic Data Committee so that parsers will recognise and support entities representing special characters.

Reference : NGDF Working Group 2 - NGDF Discovery Metadata Guidelines, March 1999.

DTD Ref : ngdfdm2.dtd

DTD Document URL : www.ngdf.org.uk/ngdfdm2.dtd

Authors : Martin Ralphs and Paul Miller, for NGDF Working Group 2

Changes:



2000-02-09 v2 modified by NGDF Metadata Working Group in line with NGDF Discovery Metadata Guidelines version 1.2

Amended westext, eastext, northext, southext to mandatory rather than optional. The boundrec element is optional but the sub elements must be included if boundrec is used.

Modified keyword to separate individual keywords

Modified coordref so that multiple bounding rectangles defined are enclosed in one coordinate definition

modified code list terms to capitalise multiple word descriptions (e.g. britishNationalGrid, notKnown)

1999-03-10 v1 Initial release (ngdfdm.dtd) in line with NGDF Discovery Metadata Guidelines versions 1.0 and 1.1

```
-->
<!ENTITY lt "&#38;#60;">
<!ENTITY gt "&#62;">
<!ENTITY amp "&#38;#38;">
<!ENTITY apos "&#39;">
<!ENTITY quot "&#34;">
<!ENTITY trd "&#174;"> <!-- trademark symbol -->
<!ENTITY lbs "&#163;"> <!-- pound sterling sign -->
<!ENTITY cpy "&#169;"> <!-- copyright symbol -->
<!ENTITY deg "&#176;"> <!-- degree symbol -->
<!ENTITY plm "&#177;"> <!-- plus or minus symbol -->
<!ENTITY s2 "&#178;"> <!-- superscript two -->
<!ENTITY at "&#64;"> <!-- commercial at symbol, used in email -->

<!ELEMENT metadata (title , alttitle* , originat* , abstract , capture? , present* , access* , usecon? , keywords+ ,
geogext , spatrefsys+ , spatdet? , media* , dataform* , addinfo* , associat* , supplier+ , update , sample*)>

<!ELEMENT title (#PCDATA)>
<!ELEMENT alttitle (#PCDATA)>
<!ELEMENT originat (#PCDATA)>
<!ELEMENT abstract (#PCDATA)>
<!ELEMENT capture (startstat , begdate? , endstat , enddate? , frequpdt)>

<!ELEMENT present EMPTY>
<!ATTLIST present
presentation (image|graphic|map|numeric|text) #REQUIRED>

<!ELEMENT access EMPTY>
<!ATTLIST access
accrestrict (financial|legal|other|notKnown|none) #REQUIRED>

<!ELEMENT usecon (#PCDATA)>
<!ELEMENT keywords (#PCDATA)>
<!ELEMENT geogext (coordref* , geogref* )>
<!ELEMENT spatrefsys (#PCDATA)>
<!ELEMENT spatdet (#PCDATA)>
<!ELEMENT media EMPTY>
<!ATTLIST media
mediatype (paper|magnetic|optical|online|other) #REQUIRED >
<!ELEMENT dataform (#PCDATA)>
<!ELEMENT addinfo (#PCDATA)>
<!ELEMENT associat (#PCDATA)>
```

```
<!ELEMENT supplier (cname , postaddr , postcode , telno? , faxno? , email? , webaddr? )>
<!ELEMENT update (#PCDATA)>
<!ELEMENT sample (#PCDATA)>

<!ELEMENT coordref (syscoord , boundrec+)>
<!ELEMENT geogref (natext* , adminext* , pcodeext*)>

<!ELEMENT startstat EMPTY>
<!ATTLIST startstat
  sstat (known|notKnown|noApplicable) #REQUIRED>

<!ELEMENT stdate (#PCDATA)>

<!ELEMENT endstat EMPTY>
<!ATTLIST endstat
  edstat (known|notKnown|notApplicable|ongoing) #REQUIRED>

<!ELEMENT enddate (#PCDATA)>

<!ELEMENT frequpdt EMPTY>
<!ATTLIST frequpdt
  frequency
  (hourly|daily|weekly|fortnightly|monthly|quarterly|biannually|annually|biennially|triennially|quinquennially|decennially|continuous|irregular|never|other) #REQUIRED>

<!ELEMENT syscoord EMPTY>
<!ATTLIST syscoord
  system (britishNationalGrid|irishGrid|latitudeAndLongitude) #REQUIRED>

<!ELEMENT boundrec (westext , eastext , northext , southext)>
<!ELEMENT westext (#PCDATA)>
<!ELEMENT eastext (#PCDATA)>
<!ELEMENT northext (#PCDATA)>
<!ELEMENT southext (#PCDATA)>

<!ELEMENT natext EMPTY>
<!ATTLIST natext
  natextent (england|northernIreland|scotland|wales|isleOfMan|channelIslands|offshore|unitedKingdom)
  #REQUIRED>

<!ELEMENT adminext (#PCDATA)>
<!ELEMENT pcodeext (#PCDATA)>
<!ELEMENT cname (#PCDATA)>
<!ELEMENT postaddr (#PCDATA)>
<!ELEMENT postcode (#PCDATA)>
<!ELEMENT telno (#PCDATA)>
<!ELEMENT faxno (#PCDATA)>
<!ELEMENT email (#PCDATA)>
<!ELEMENT webaddr (#PCDATA)>
]>
```

4. NGDF COMMUNICATIONS PROTOCOL

4.1 Introduction

The NGDF Discovery Metadata communications protocol sets out a definition for how applications servicing the needs of an end user browser may interrogate many different NGDF Metadata Service Providers, either directly or indirectly through the Gateway and other Service Providers, and, in turn, how the Gateway and Service Providers will interrogate other Service Providers. The NGDF Discovery Metadata Guidelines provide standardised data formats, what cannot be standardised is the infrastructures (hardware, software, databases) being used by the service providers. Thus a communications protocol is required to enable the exchange of data across this diverse environment.

At the time of writing, the only means of facilitating this vision in a Web-based environment is Z39.50 (Z39.50 Maintenance Agency, 1995). Other protocols and approaches continue to evolve, but only Z39.50 can currently demonstrate an established base of interoperating geospatial services such as those which NGDF envisions.

4.2 Use of Z39.50

ANSI standard Z39.50–1995, captured internationally as ISO 23950, is a powerful yet complex protocol which has evolved to enable the near-transparent query of multiple, remote, heterogeneous databases via one simplified user interface.

Z39.50 is principally used in the library community, with Z39.50 add-ons available for virtually all of the commercially available library systems. Within this community, Z39.50 is used to provide near-transparent access to collections data held in multiple databases, whether on a local level or more widely through services such as COPAC.

In the geospatial community, adoption of Z39.50 has been most apparent in government-related initiatives such as the United States' Geospatial Clearinghouse network <<http://www.fgdc.gov>> and the Australian Spatial Data Directory <<http://www.environment.gov.au/net/asdd>>. In both these government examples, numerous agencies offer Z39.50 access to their databases, either individually for visitors to agency- and subject-specific interfaces, or collectively through the clearinghouse URLs given above.

4.3 How it works

A major benefit of Z39.50 is its neutrality. Z39.50 servers will effectively run on any hardware platform and on top of any standards-compliant DBMS. Importantly, collaborating partners in a Z39.50 project do not need to be using the same hardware, software or underlying data structure, so long as both can agree a common mapping into one or more Z39.50 'profiles'.

Although the underlying complexities of Z39.50 are important (Z39.50 Maintenance Agency, 1998a; Evans 1998a; Evans 1998b), it is sufficient at this stage to state that selecting the appropriate Z39.50 'profile' is a crucial step in designing a Z39.50-based system.

Fortunately, a profile has already been developed for geospatial data and this is suitable for NGDF requirements of both Discovery and Detailed metadata.

4.4 Z39.50 Terminology

attribute set	a collection of ‘attributes’ used both to declare any mapping between native database structures and to handle the query process. The ‘attribute set’ may well include ‘use attributes’ (field or element names), as well as attributes intended to describe how queries should be handled, such as ‘relation attributes’ (<, =<, etc.), ‘completeness attributes’ (whether or not a query maps to the entire contents of a field, or only to a subset), etc. Bib-1 is the most commonly utilised attribute set at present
client	the application responsible for initiating a request which ultimately results in a Z39.50 brokered query
gateway	an application capable of submitting user queries to a number of ‘targets’ and able in some manner to translate, manipulate, and pool returning result sets
profile	a structured statement of functions and the context of their use. Where applicable, this includes a declaration of those subsets of the broader Z39.50 standard and its relevant ‘profiles’ which are relevant in this case
record syntax	a ‘record syntax’ is the set of definitions used by both ‘client’ and ‘server’ in exchanging information
schema	a commonly understood representation of data held in the resource to be queried. Both client and server must understand the same schema
server	the application responsible for responding to requests from a ‘client’. Often, this ‘server’ is synonymous with the ‘target’
service	Z39.50 comprises a number of ‘services’ which make up the whole. Normal applications of Z39.50 will declare their requirement of one or more of these ‘services’, such that any compliant application must support those which are required. These ‘services’ include; Explain, Init, Present, Query, Retrieve, and Search target the part of a ‘server’ application responsible for handling Z39.50-based requests

4.5 The Geospatial Profile

Version 2.2 of the Z39.50 Application Profile for Geospatial Metadata (GEO), <<http://www.blueangeltech.com/Standards/GeoProfile/geo22.htm>>, has been developed by the United States’ Federal Geographic Data Committee (FGDC) and others as a means of enabling cross-searching across distributed collections of geospatial metadata complying with FGDC’s *Content Standards for Digital Geospatial Metadata* (FGDC, 1998). This specifically geospatial profile is closely based upon earlier work which resulted in the Government Information Locator Service’s (GILS) application profile (GILS 1997).

Given its tailoring towards geospatial requirements, and the close relationship between the *Content Standards for Digital Geospatial Metadata* to which it was designed to provide access and the Metadata work of ISO's Technical Committee 211 (1998), GEO is well placed to meet the requirements of NGDF.

The inner workings of both Z39.50 and the GEO Profile are, as with all applications of such potential, complex. It is not intended to discuss these details here, as they are only truly relevant to those building Z39.50 applications from scratch. Rather, sufficient information is given for geospatially cognisant individuals and organisations to make informed decisions about Z39.50 and the Profile. It is anticipated that implementers will either make use of existing Z39.50-capable applications or else sub-contract such work to Z39.50-experienced programmers, such that overloading of this report with technical detail would serve only to confuse most readers, whilst offering nothing new to the Z39.50-literate. Numerous references are offered for those in search of further detail.

The GEO Profile is a fully conformant application of ANSI/NISO Z39.50 version 2, and additionally provides requirements for the construction of a GEO-compliant service capable of serving Content Standards for Digital Geospatial Metadata-compliant data across the Internet in an interoperable fashion.

4.5.1 Z39.50 Services

Services are the basic structural blocks into which the Z39.50 standard is divided. In each case, the service defines a two-way exchange of information between client and target. Simplifying hugely, the Initialisation (or Init) service may be seen as a greeting from the client ("Hello. Do you speak English?") and a response from the target ("Hello. Yes I do, let's talk"). Without this positive two-way dialogue, the Z39.50 session cannot proceed.

The Init, Search, and Present services are required for conformance with GEO. It is normally assumed that TCP/IP will be used as the transport service.

Following the simplistic language-based analogy, above, the remaining two core services for GEO may be explained, thus;

A Search request is transmitted from the client ("OK – can I have everything you've got about a place called 'York'?"), and is responded to by the target ("I've got 23 records matching your request").

And finally, the client uses the Present service to ask for the data ("23, eh? Can I have the first ten, please?"), resulting in transmission of the records themselves from the target.

4.5.2 Attribute Sets

GEO requires that applications recognise the Bib-1 (Z39.50 Maintenance Agency 1998b) and GILS (1997) attribute sets and support the GEO attribute set. The GEO attribute set is outlined as part of the Profile, and the other two are often supported in existing Z39.50 applications.

4.5.3 Diagnostic Set

The GEO Profile utilises the commonly understood Bib-1 diagnostic set (Z39.50 Maintenance Agency 1998c) for transfer of diagnostic details between applications.

4.5.4 Record Syntax

Record syntaxes are used in Z39.50 to define the manner in which data is passed from target to client in response to a Present request.

The GEO Profile requires support of the HTML and SUTRS (Simple Unstructured Text Record Syntax) record syntaxes, and optionally allows use of SGML (ISO 1986), GRS–1 and USMARC (Library of Congress 1998) record syntaxes within conformant applications.

4.5.5 Schemas

Schemas may be used within Z39.50 to reflect dependencies and relationships between individual data elements. A schema might, for example, declare that where an ‘Easting’ is used, a ‘Northing’ must also be present.

The GEO Profile defines a GEO Schema which all conformant applications must support. This Schema is expressed as a Document Type Definition (DTD), and is similar in structure to the NGDF Transfer Format

4.6 The ‘NGDF Profile’ – A subset of GEO

Although only a subset of full Z39.50 functionality, the GEO Profile is over–large for NGDF Discovery Metadata requirements. As such, this section is intended to identify the set of elements and attributes required for an NGDF Discovery Metadata Service, mapping these to NGDF’s Discovery Metadata elements.

In the short term, NGDF–conforming applications should be designed to support the GEO profile in its unaltered state. There is scope, however, for a degree of streamlining in the future, resulting in a more tightly defined Profile specifically tailored to NGDF. This work might also usefully tailor GEO’s Summary Record Format to map exactly with NGDF’s Discovery Metadata elements. This is not currently the case, meaning that NGDF records can only reliably be presented using GEO’s Full Record Format. Until an NGDF Profile is deemed sufficiently desirable for the work to be done, all NGDF applications will utilise a subset of GEO. Any reference to an ‘NGDF Profile’ is therefore to a virtual subset of GEO; applications, clients and targets have no knowledge of any NGDF modifications to the Profile, nor any understanding of why a large number of GEO elements are always left empty. There are also two limitations that need to be addressed in adopting this subset approach.

There are a few NGDF metadata attributes (i.e. Status of Start Date of Capture and Status of End Date of Capture, Alternative Title and level of Spatial Detail) that cannot be supported as there is no appropriate mapping to the GEO attribute sets. Until an NGDF Profile is defined these optional attributes will be excluded from the protocol.

As the GEO profile does not provide a method to search the status of start or end dates of capture, any record that does not include a valid date will not be found when a date search is undertaken. When providing metadata for searching through the NGDF Gateway, users will need to decide if a date can be provided. Where the status of the enddate is “ongoing”, the metadata providers could provide a date concurrent with the metadata update date.

To utilise the spatial searching capability of the NGDF Gateway, the geographic extent of each dataset will need to be described as a bounding rectangle based on latitude and longitude co-ordinates. While the NGDF DTD does not specify that bounding rectangle

coordinates are mandatory, it will be a mandatory requirement for data served by the Gateway. The NGDF Board recognise that organisations will describe data extents in any of the methods described by the guidelines and that the latitude and longitude extents can be determined from these other descriptions (e.g. postcode district or administrative area extents or British National Grid coordinates).

Further details of implementing NGDF metadata into the NGDF Gateway service are provided on the NGDF web site.

4.7 NGDF Discovery Metadata mapped to GEO

GEO Use Attribute	GEO Attribute Name	NGDF Discovery Metadata Element Name (NGDF Working Group 2, 1998)
4	Title	Title
	Alternate Title	Alternative Title
1005	Originator	Originator
62	Abstract	Abstract
		Status of Start Date of Capture
2072	Beginning Date	Start Date of Capture
		Status of End Date of Capture
2073	Ending Date	End Date of Capture
3109	Maintenance and Update Frequency	Frequency of Update
3805	Geospatial Data Presentation Form	Presentation Type
2004	Access Constraints	Access Constraint
2005	Use Constraints	Use Constraint

GEO Use Attribute	GEO Attribute Name	NGDF Discovery Metadata Element Name (NGDF Working Group 2, 1998)
3121	Keywords	Keywords
		System of Spatial Referencing by Co-ordinates
2038	West Bounding Co-ordinate	West Bounding Co-ordinate
2039	East Bounding Co-ordinate	East Bounding Co-ordinate
2040	North Bounding Co-ordinate	North Bounding Co-ordinate
2041	South Bounding Co-ordinate	South Bounding Co-ordinate
2042 (in conjunction with 2043)	Place Keyword	National Extent
2042 (in conjunction with 2043)	Place Keyword	Administrative Area Extent
2042 (in conjunction with 2043)	Place Keyword	Postcode District Extent
2050	Supplemental Information	Spatial Reference System
		Level of Spatial Detail
3632	Offline Media	Supply Media
3608	Format Name	Data Format
3812	Other Citation Details	Additional Information

GEO Use Attribute	GEO Attribute Name	NGDF Discovery Metadata Element Name (NGDF Working Group 2, 1998)
		Source
2068	Cross Reference	Dataset Association
2023/2024	Contact Person/ Contact Organisation	Contact Name or Title
2023/2024	Contact Person/ Contact Organisation	Contact Name or Title
2025	Address	Full Postal Address of Supplier
2028	Postal Code	Postcode of Supplier
2032	Contact Voice Telephone	Telephone Number of Supplier
2033	Contact Facsimile Telephone	Facsimile Number of Supplier
2030	Contact Electronic Mail Address	E-mail Address of Supplier
3618	Network Address	Web Address of Supplier
1012	Metadata Date	Date of Update of Metadata
3137	Browse Graphic	Sample

5. CONFORMANCE

5.1 Conformance Requirements

Until such time as the NGDF Management Board establishes a process of accreditation and certification, including software validation, conformance to these NGDF Guidelines must rely on self-certification. If users wish to claim conformance then they should, as a minimum, comply with the conformance clause given at 0. The approach adopted here follows that given in draft international standard ISO 15046 - Geographic Information - Part 5: Conformance and Testing [2].

5.2 Conformance

5.2.1 Statement of conformance

Any metadata claiming conformance to these NGDF Guidelines shall pass all the requirements described in the following abstract test suite.

5.2.2 Abstract Test Suite

The abstract test suite contains two sub-clauses:

(1) that relating to the applicability of the data resource described in the metadata (0);

(2) that relating to the content of the metadata (0)

The requirements of both these sub-clauses shall be met in full for a product to claim conformance to the NGDF Guidelines.

5.3 Abstract Tests

5.3.1 Applicability of data resource to the NGDF Guidelines

Purpose of test: to check whether the data resource is within scope of the NGDF Guidelines.

Test method: ascertain that the data resource:

- is *either*:
 - a dataset i.e. an identified collection of data
- or*,
- a dataset series or a collection i.e. datasets captured to a common specification and relating to different time periods or geographic extents or both.

- contains data which is geospatially referenced either by coordinates or geographic identifiers, or both, to the United Kingdom or bounding seas.

The dataset may be in a digital, textual or graphical form.

5.3.2 Completeness of metadata content

Purpose of test: to check completeness of metadata relating to a valid data resource.

Test method: verify that the metadata conforms to the requirements set out in sub-clause 3.2 by ascertaining that:

- i) values for all metadata elements which are mandatory are present;
- ii) values for all metadata elements which are conditional are present where that condition applies;
- iii) all repeating groups which are mandatory are present and the values for the metadata elements within the repeating groups which are (1) mandatory or (2) conditional (where the condition applies), are present;
- iv) all repeating groups which are conditional are present, where the condition applies, and the values for the metadata elements within the repeating groups which are (1) mandatory or (2) conditional (where the condition applies), are present;
- v) for each completed metadata element there is no more than one value where the maximum occurrence is stated as "1";
- vi) for each completed metadata element there is at least one value where the maximum occurrence is stated as "N";
- vii) the data type for each value of a completed metadata element is as stated;
- viii) the value for each completed metadata element is within the domain stated;
- ix) where the value is "Not Known" or "Not Applicable" or "None" this is recorded according to the stated requirement.

5.4 Compatibility of Data Transfer Format

Purpose of test: To test that data may be exported and imported using the NGDF transfer protocol.

It is the intention to develop a test data set that any conforming software must be able to ingest and export using the NGDF transfer format such that the ingest file set is identical to the export file. At the time of writing this work has yet to be done.

5.5 Compatibility of Z39.50 Protocol

This section is subdivided into two compatibility tests, a software developer may chose to develop software that is compatible with either or both tests depending on the software's purpose. Purpose of test: To test that data may be queried from an external database and that data may be supplied in response to an external request using the Z39.50 based protocol defined for NGDF.

It is the intention to develop a test data set and transaction sequences and expected outcomes that any conforming software should be able to support and complete. At the time of writing this work has yet to be done.

Blank Page

APPENDIX A - MAPPING TO XML

This appendix explains how the NGDF Discovery Metadata elements presented in the Guidelines map to the element definitions in the XML DTD presented here.

NGDF Discovery Metadata Guidelines		XML DTD Specification
Identifier	Element Name	Element Name
1	Title	title
2	Alternative Title	alttitle
3	Originator	originat
4	Abstract	abstract
5	Data Capture Period	capture
6	Status of Start Date of Capture	startstat
7	Start Date of Capture	startdate
8	Status of End Date of Capture	endstat
9	End Date of Capture	enddate
10	Frequency of Update	frequpdt
11	Presentation Type	present
12	Access Constraint	access
13	Use Constraint	usecon
14	Keywords	keywords
15	Geographic Extent	geogext
16	Spatial Referencing by Coordinates	coordext
17	System of Spatial Referencing by Coordinates	syscoord
18	Bounding Rectangle	boundrec
19	West Bounding Coordinate	westext
20	East Bounding Coordinate	eastext
21	North Bounding Coordinate	northext
22	South Bounding Coordinate	southext
23	Spatial Referencing by Geographic Identifiers	geogref
24	National Extent	natext
25	Administrative Area Extent	adminext
26	Postcode District Extent	pcodeext
27	Spatial Reference System	spatrefsys
28	Level of Spatial Detail	spatdet
29	Supply Media	media
30	Data Format	dataform
31	Additional Information Source	addinfo
32	Dataset Association	associat
33	Supplier	supplier
34	Contact Name or Title	cname
35	Full postal address of supplier	postaddr
36	Postcode of Supplier	postcode
37	Telephone number of supplier	telno

NGDF Discovery Metadata Guidelines		XML DTD Specification
Identifier	Element Name	Element Name
38	Facsimile number of supplier	faxno
39	Email address of supplier	email
40	Web address of supplier	webaddr
41	Date of update of metadata	update
42	Sample	sample

APPENDIX B - XML EXAMPLE

This section of the document contains a complete example of encoding using the Discovery Metadata XML Document Type Description. The information presented here is taken from first example in the NGDF Discovery Metadata Guidelines Appendix A – Worked Examples and refers to OS Land-Line.

```
<?xml version="1.0" ?>
<!DOCTYPE NGDFDiscoveryMetadata SYSTEM "www.ngdf
.org.uk/ngdfdm2.dtd">
<NGDFDiscoveryMetadata>
<title>
Land-Line
</title>
<altitle>
Not applicable
</altitle>
<originat>
Ordnance Survey, Great Britain
</originat>
<abstract>
Large scale digital topographic mapping for site location, planning
and land and asset management. Land-Line contains 30 feature codes
and six text categories. Standard map tile coverage 500 x 500 m in
urban areas, 1 km x 1 km in rural areas and 5km x 5km in moorland
areas.
</abstract>
<capture>
<startstat ststat="notApplicable" />
<stdate>
</stdate>
<endstat edstat="ongoing" />
<enddate>
</enddate>
<frequpdt frequency="continuous" />
</capture>
<present presentation="map" />
<access accrestrict="financial" />
<usecon>
All Ordnance Survey digital map data are Crown Copyright. Full
details of the conditions under which Ordnance Survey digital map
data may be processed, manipulated or copied by a customer are
contained in the Ordnance Survey leaflet: Copyright 3 – Digital Map
Data. This is available from Copyright Licensing, Ordnance Survey,
Romsey Road, Southampton SO16 4GU, telephone +44 (0) 1703 792913, fax
+44 (0) 1703-792435
</usecon>
<keywords>
base maps
</keywords>
<keywords>
geoscientific information
</keywords>
<keywords>
roads
</keywords>
<keywords>
buildings
</keywords>
<keywords>
geographic names
</keywords>
```

```

<geogext>
<coordref>
<syscoord system="britishNationalGrid" />
<boundrec>
<westext>
0
</westext>
<eastext>
700
</eastext>
<northext>
1300
</northext>
<southext>
0
</southext>
</boundrec>
</coordref>
<geogref>
<natext natextent="england" />
<natext natextent="scotland" />
<natext natextent="wales" />
</geogref>
</geogext>
<spatrefsys>
British National Grid
</spatrefsys>
<spatdet>
Nominal Scale 1 : 1250 in urban areas, 1 : 2500 in rural areas and
1 : 10000 in moorland areas
</spatdet>
<media mediatype="magnetic" />
<media mediatype="optical" />
<dataform>
NTF v. 2.0 (BS7567)
</dataform>
<dataform>
DXF Release 12
</dataform>
<addinfo>
Ordnance Survey Customer Information Helpline : phone +44 (0) 345 33
00 11, fax +44 (0) 1703 792452, e-mail
custinfo@ordsvy.gov.uk
</addinfo>
<addinfo>
www.ordsvy.gov.uk
</addinfo>
<associat>
ADDRESS-POINT
</associat>
<associat>
OSCAR
</associat>
<supplier>
<cname>
Customer Services - Digital Helpdesk
</cname>
<postaddr>
Ordnance Survey, Romsey Road, SOUTHAMPTON, United Kingdom.
</postaddr>
<postcode>
S016 4GU
</postcode>
<telno>
+44 (0) 1703 792773
</telno>

```

<faxno>
+44 (0) 1703 792324
</faxno>
<email>
custinfo@ordsvy.gov.uk
</email>
<webaddr>
www.ordsvy.gov.uk
</webaddr>
</supplier>
<update>
19971117
</update>
<sample>
</sample>
</NGDFDiscoveryMetadata>