
Biodiversity and Conservation **13:** 1009–1022, 2004.

The establishment of China's core biodiversity metadata standard

HAIGEN XU[1], YIMIN LI[2,*] and HUI DING[1]

*[1]Nanjing Institute of Environmental Sciences, State Environment Protection Administration, Nanjing 210042, China; [2]College of Automation, Nanjing University of Aeronautics and Astronautics, Nanjing, China; *Author for correspondence*

Received 6 May 2002; accepted in revised form 28 March 2003

Key words: Biodiversity, China, Information sharing, Metadata, Standard

Abstract. Biodiversity metadata provide service to query, management and use of actual data sets. The progress of the development of metadata standards in China was analyzed, and metadata required and/or produced based on the Convention on Biological Diversity were reviewed. A biodiversity metadata standard was developed based on the characteristics of biodiversity data and in line with the framework of international metadata standards. The content of biodiversity metadata is divided into two levels. The first level consists of metadata entities and elements that are necessary to exclusively identify a biodiversity data set, and is named as Core Metadata. The second level comprises metadata entities and elements that are necessary to describe all aspects of a biodiversity data set. The standard for core biodiversity metadata is presented in this paper, which is composed of 51 elements belonging to 6 categories (entities), i.e. inventory information, collection information, information on the content of the data set, management information, access information, and metadata management information. The name, definition, condition, data type, and field length of metadata elements in these six categories (entities) are also described.

Introduction

Metadata standards are receiving a lot of attention worldwide, as their use will facilitate queries from, management, and use of actual data. China is a country with one of the richest biodiversities in the world (State Environmental Protection Administration of China 1998), and is therefore rich in biodiversity information. Much work has been done on creating biodiversity inventory, monitoring and information systems (Xu et al. 1999, 2000; Xu and Wu 1999). The bottleneck in sharing all this information may be not knowing where to find useful information and how to use it. Metadata form the basis for making informed decisions regarding the fitness of a particular data source for a specific use (Chrisman 1994).

During the implementation of the project entitled Biodiversity Data Management and Information Networking Capacity Building in China, financed by UNEP through GEF between 1995 and 1998, the contents of metadata (database fields) of institutions, data sets, databases, books and maps related to biodiversity were presented, and a national biodiversity information query system was designed (Xu et

al. 1999). The content of metadata does not bring the function of metadata into full play. There is a lack of information on spatial data and data quality and of electronic access to these. The lack of unified metadata standards also made it difficult to load data into the metadatabase and to query and use metadata by users. This national biodiversity information query system cannot run on computer networks and limits access by users. With the fast development of the Internet and lots of biodiversity databases accessible through the Internet, there is a common hope that a more comprehensive and integrated metadata standard be developed, so as to guide the combination, sharing and management of biodiversity data sets in a range of institutions in China, and to make data input and data enquiry more convenient for users. In order to promote information-sharing based on the Internet, a project named the Demonstration of Information-Sharing for Sustainable Development was financed under the framework of the National R & D Program in the Ninth Five-Year Plan (1996 to 2000). A biodiversity metadatabase was needed to guide the query, use and management of vast amounts of biodiversity data across numerous thematic information networking systems, such as the networking systems of animal species (http://sdinfo.ioz.ac.cn), microorganisms (http:// www.im.ac.cn), crop germplasm resources (http://icgr.caas.net.cn), forest ecosystems (http://sdinfo.forestry.ac.cn), marine ecosystems (http://sdinfo. coi.gov.cn), and nature reserves (http://www.biodiv.org.cn). An integrated biodiversity metadata standard was required to guide such services.

The Clearing-House Mechanism (CHM) of the Convention on Biological Diversity (CBD), which aims to promote and facilitate technical and scientific cooperation between contracting Parties and participating partners, functions on metadata level. It plays an important role in the implementation of the objectives of the Convention. Information required and/or produced based on Articles 5 to 20 and 26 of the Convention are involved in metadata (Table 1). The Informal Advisory Committee of the CHM has organized many meetings to discuss metadata standards. Metadata standards were also discussed by other international organizations, such as the International Organization for Standardization, and by many countries including the USA, the UK, Australia, and New Zealand. This metadata standard was developed according to the requirements of China and in line with international metadata standards.

Methods

Although the national biodiversity information query system (Xu et al. 1999) promotes the enquiry and management of biodiversity data sources in China, some disadvantages do exist with respect to the fast increase of biodiversity information and development of the Internet. Different metadata entities were addressed with different databases. For instance, metadata of data sets was deposited in the metadatabase of data sets, while metadata of databases was organized by the metadatabase of databases. This may cause some inconvenience for users in locating data

Table 1. Metadata required and/or produced based on the Convention on Biological Diversity.

Articles of the Convention	Metadata of data sets on
5 Cooperation	International cooperation projects
6 General Measures for Conservation and Sustainable Use	National/departmental strategies, plans or programs
7 Identification and Monitoring	Ecosystems and habitats
	Species diversity
	Gene diversity
	Monitoring activities
8 *In situ* Conservation	Protected areas
	Natural resources
	Rare and endangered species
	Degraded ecosystems
	Living Modified Organisms
	Alien Invasive Species
	Indigenous knowledge, innovation and practices
	Legislation for protection of rare and endangered species
	Infrastructure and natural resources exploitation projects
9 *Ex situ* Conservation	Botanical gardens and zoos
10 Sustainable Use	Traditional cultural practices
	Remedial actions in degraded areas
	Expertise for sustainable use of biological resources
11 Incentive Measures	Incentives for the conservation and sustainable use of biodiversity
12 Research and Training	Research and training projects
	Expertise for the conservation and sustainable use of biodiversity
13 Public Education and Awareness	Public education materials
14 Impact Assessment and Minimizing Adverse Impacts	Projects that may have adverse impacts on biodiversity
	Environmental Impact Assessment reports

Table 1. Continued.

Articles of the Convention	Metadata of data sets on
15 Access to Genetic Resources	National legislation on access to and benefit sharing of genetic resources
	Reports on access to genetic resources granted
	R & D projects based on genetic resources provided by other Contracting Parties
16 Access to and Transfer of Technology	National legislation on intellectual property and patents
	Available technology
	Demand for technology
17 Exchange of Information	Results of technical, scientific and socio-economic research
	Indigenous and traditional knowledge
18 Technical and Scientific Cooperation	Research projects and training programs
19 Handling of Biotechnology and Distribution of its Benefits	Projects based on genetic resources provided by other Contracting Parties
	National legislation on biosafety
	Environmental release, commercial production and transboundary movement of LMOs and risk assessment reports
20 Financial Resources	Multilateral and bilateral cooperation projects
	Reports on funds granted
26 Reports	National reports and thematic reports on measures taken and assessment of their effectiveness

sources. Meanwhile, the national biodiversity information query system needs to be updated to meet information enquiry and sharing based on the Internet. Therefore, the content of metadata should be integrated into one unified metadata standard, so that the earlier different metadatabases could be modified and a unified metadatabase could be established to integrate all metadata entities into one metadatabase.

Since the 1990s, many international organizations and countries have been working on metadata standards, e.g. the Standard for Digital Geospatial Metadata released by the US Federal Geographic Data Committee (FGDC) (FGDC 1994, 1997), the Metadata Standard of the Center for International Earth Science Information Network (CIESIN), the Metadata Standard of US National Biological Information Infrastructure (NBII), the Core Metadata Standard on Australia Terrestrial Creature Survey, etc. These metadata standards provide a framework and guidance for the development of China's biodiversity metadata standard. Meanwhile, biodiversity data have the characteristics of extensive coverage, coexistence of time sequence with space sequence, coexistence of attribute data with spatial data, and coexistence of electronic data with written data. The biodiversity metadata standard should reflect the characteristics of biodiversity data.

The biodiversity metadata standard developed was based on the characteristics of biodiversity data and in line with the framework of international metadata standards, including that of ISO/TC 211. In addition, succinctness and applicability should be observed. Because metadata are provided by data set supervising institutions (authors), and no law stipulates the obligation of the institutions (authors) to provide metadata, institutions (authors) should be encouraged to provide, supplement and update metadata.

The content of biodiversity metadata is divided into two levels. At the first level there are metadata entities and elements that are necessary to exclusively identify a biodiversity data set, and this is named Core Metadata. The second level has metadata entities and elements that are necessary to describe all aspects of the biodiversity data set. The standard for core biodiversity metadata is only discussed in this context. A preliminary core biodiversity metadata standard was formulated according to metadata standards home and abroad. After consulting more than 20 experts who are engaged in biology, ecology, nature conservation, and information management, a draft of a core biodiversity metadata standard was obtained.

Results

The Core Biodiversity Metadata Standard is presented in Table 2. It is composed of 6 categories (entities) and 51 elements. The entity of inventory information consists of elements including the title, version, series, and publication of the data set.

The entity of collection information consists of elements including the programme to support the production of the data set, the person responsible for the programme, and the quality of the data set.

Table 2. Draft of core biodiversity metadata standard.

Entity	Element	No. of element	Definition	Characteristics/ condition[a]	Largest occurrence	Data type	Field length	Remark
Inventory information	Chinese title of data set	101	Chinese full name of data set	M	1	Character	70	
	English title of data set	102	English full name of data set	O	1	Character	100	
	Version	103	Version of data set	O	1	Character	20	
	Name of series	104	General name of series production that data set belongs to	C/series publication	1	Character	50	
	Identification of publication series	105	Identification information or serial number of data set in series production	C/series publication	1	Character	50	
	Press	106	Name of publisher which represents main publication of data set	C/publication	1	Character	30	
	Published site	107	Address of publisher	C/publication	1	Character	20	
	Published date	108	Published or planned publishing date of data set	O	1	Date	8	
Collection information	Name of programme	201	Programme to support the research and production of data set	O	1	Character	50	
	Type of programme	202	Explaining the type of programme	O	1	Character	40	Such as National R&D programme, National Natural Science Foundation, departmental programme, company R&D programme, etc.
	Chief	203	Name of chief responsible for the programme	O	1	Character	20	
	Collection manner	204	Manner to collect data	O	1	Character	130	
	Quality explanation	205	General explanation on data quality	O	1	Character	180	

Table 2. Continued.

Entity	Element	No. of element	Definition	Characteristics/ condition[a]	Largest occurrence	Data type	Field length	Remark
Information on content of data set	Purpose	301	Main usefulness of data set	O	1	Character	100	
	Main content	302	Simple introduction to the content of data set	M	1	Memo		
	Keywords	303	Common words to generalize the main content of data set	O	N	Character	50	
	Coordinates of western border	304	Longitude of western border covered by data set	C/spatial data	1	Double precision	8.2	$-180° \leqslant$ Coordinates of western border $< 180°$
	Coordinates of eastern border	305	Longitude of eastern border covered by data set	C/spatial data	1	Double precision	8.2	$-180° \leqslant$ Coordinates of eastern border $< 180°$
	Coordinates of northern border	306	Latitude of northern border covered by data set	C/spatial data	1	Double precision	8.2	$-90° \leqslant$ Coordinates of northern border $< 90°$ Coordinates of northern border $\geqslant$ Coordinates of southern border
	Coordinates of southern border	307	Latitude of southern border covered by data set	C/spatial data	1	Double precision	8.2	$-90° \leqslant$ Coordinates of southern border $< 90°$ Coordinates of southern border $\leqslant$ Coordinates of northern border

Table 2. Continued.

Entity	Element	No. of element	Definition	Characteristics/ condition[a]	Largest occurrence	Data type	Field length	Remark
Information on content of data set	Coverage of data set	308	Areas or regions covered by data set	C/geographic coordinates are not employed	1	Character	100	
	Range of time	309	Type of time range on data set	M	1	Character	10	1. Single time, 2. Time range
	Time 1	310	Singular time or initial moment of time on data content	M	1	Character	40	
	Time 2	311	Terminative moment of time on data content	C/time range	1	Character	40	
	Scale	312	Vector representing denominator of scale on spatial data	C/vector	1	Character	44	
	Resolution	313	Parameter representing data density of grid or image	C/grid or image data	N	Numeric	20	Ground resolution of remote sensing data, mesh dimension of grid data
	Type of data representation	314	Method to represent information of data set	O	N	Character	50	1. Character, 2. Vector, 3. Grid, 4. Image, 5. Relational type
	Data item	315	Names of fields of database	C/relational data	N	Memo		
	Type of spatial reference system	316	Spatial reference system for spatial positioning of data set	O	2	Character	50	1. Geographic coordinate (longitude, latitude), 2. Projecting coordinate

Table 2. Continued.

Entity	Element	No. of element	Definition	Characteristics/ condition[a]	Largest occurrence	Data type	Field length	Remark
	Institution supervising data set	401	Name of institution supervising data set	C/ data set not owned by individual	1	Character	50	
	Chief of supervising institution	402	Name of chief of the institution supervising data set	C/ data set not owned by individual	1	Character	20	
	Function of supervising institution	403	Proportion of copyright of sharing data set the supervising institution possesses or the function the institution plays	O	1	Character	30	1. All copyrights, 2. Partial copyright, 3. Organizing and implementation, 4. Management, 5. Production, 6. Supply of data
Management information	Authors or persons to manage data set	404	Name of authors or persons to manage data set	O	1	Character	50	
	Progress of data set	405	Progress of data set	O	1	Character	20	1. Finished, 2. Will be finished, 3. In production, 4. Planned, 5. In archive, 6. Cancelled
	Maintenance of data set	406	After the data has been finished, its revision, supplementation and maintenance	O	1	Character	80	
	Storage media of data set	407	Name of media used for storage of data set	O	N	Character	60	CD-ROM, floppy disk, tape, network, publication

Table 2. Continued.

Entity	Element	No. of element	Definition	Character-istics/ condition[a]	Largest occurrence	Data type	Field length	Remark
	Computer hardware	408	Hardware requirement for the management of data set	O	1	Character	50	
	Computer software	409	Software for the management of data set	O	1	Character	100	
	Format of storage	410	Name of convertible data format, for example, TXT, DBF, DLG, ARC/INFO, etc.	O	1	Character	50	
	Volume of data	411	The data volume of data set stored in the format of storage	O	1	Character	20	Unit: MB, amount of records, ten thousand words
Access information	Access means	501	Means to access to information, such as online, publication, on spot, etc.	O	1	Character	40	
	Access address	502	Address of Uniform Resource Locator (URL)	C/ online	1	Character	100	
	Limitation to access	503	While accessing data, regulation on confidentiality and protecting intellectual property and other special limitations	O	1	Character	80	
	Limitation to use	504	While using data, regulation on confidentiality and protecting intellectual property and other special limitations	O	1	Character	170	
Metadata management information	Institution responsible for metadata	601	Name of institution responsible for the content of the metadata	M	1	Character	50	

Table 2. Continued.

Entity	Element	No. of element	Definition	Characteristics/ condition[a]	Largest occurrence	Data type	Field length	Remark
	Author of metadata	602	Name of individual responsible for the content of the metadata	C/ individual responsible for metadata	1	Character	50	
	District	603	Province, prefecture, county in which the institution or individual responsible for the content of the metadata are located	O	1	Character	20	
	City	604	City in which the institution or individual responsible for the content of the metadata are located	O	1	Character	10	
	Address	605	Street No. or P.O. Box	O	1	Character	50	
	Postcode	606	Postcode	O	1	Character	6	
	Telephone	607	Telephone of the institution or individual responsible for the content of the metadata	O	N	Character	100	

[a]M: mandatory; C: conditional; O: optional.

The entity of information on the content of data set consists of elements including the purpose, main content, key words, coordinates, coverage, time range, scale, resolution, type of data representation, data item, and spatial reference system of the data set.

The entity of management information consists of elements including the supervising institution, author, progress, maintenance, storage media, format, and data volume of the data set.

The entity of access information consists of elements including the access means, access address, limitation to access, and limitation to use of the data set.

The entity of metadata management information consists of elements including the institution responsible for metadata, author of metadata, and address of metadata author of the data set.

The name, definition, condition, data type, and field length of metadata elements in these six categories (entities) are described in Table 2.

Discussion

There is an awareness of the barriers to location and access of information among thousands of data sources. In particular, uncertainty of the quality of data is seen to be a barrier to data use (Reynolds et al. 1997). The integrity and utility of a data set are dependent on the quality of its documentation. Data should be submitted with metadata, and flexible metadata standards need to be established to encourage the provision of metadata records, without providing a barrier to submission.

Many countries and international organizations have been working on metadata standards. The US Federal Geographic Data Committee (FGDC) has begun to work on the Standard for Digital Geospatial Metadata, and released its revised version in April 1997 (FGDC 1994, 1997). The content of metadata was composed of 7 categories in this standard, i.e. identification, quality, geospatial organization, spatial reference system, entity and attribute, release, and metadata reference information. Each category is composed of a number of entities and elements. This standard includes more than 300 metadata elements.

The Third Ad Hoc Group of the International Organization for Standardization developed the ISO/TC 211 Geographic Information Metadata Standard (ISO 15046-15), based upon the standard of FGDC (ISO/TC 211 WG3 1998). In this standard, the content of metadata is divided into three types: Mandatory (M), which means the content must be provided; Conditional (C), which means the content should be provided under some circumstances; Optional (O), which means the content is optional. This standard includes nearly 500 metadata elements that belong to 7 categories of metadata.

The Directory Interchange Format of National Aeronautics and Space Administration (NASA), the Metadata Standard of the Center for International Earth Science Information Network (CIESIN), USA, the Dublin Core Element Standard of UK, and the Core Metadata Standard of Australia and New Zealand are simpler

than the standards of FGDC and ISO/TC 211. Generally, they include no more than 40 metadata elements.

The Metadata Standard of US National Biological Information Infrastructure (NBII) is based on and revised and supplemented according to the standard of FGDC. The Core Metadata Standard on Australia Terrestrial Creature Survey includes categories on data set information, site information, site visitation information, creature observation information, etc. It includes over 90 metadata elements.

The Center for Biodiversity Information, Chinese Academy of Sciences, has laid down a metadata standard. The standard includes identification information, collection information, management information, description of data set, visitation information, information of metadata management, etc., with up to 58 metadata elements. But it does not have spatial description and data quality information.

Different metadata contents of institutions, data sets, databases, books and maps related to biodiversity conservation and sustainable use have been established in China (Xu et al. 1999). However, no coordinated effort was focused on establishing a comprehensive biodiversity metadata standard. These different metadata contents were modified and integrated into a single standard, and the core biodiversity metadata standard was developed in this paper. This standard includes such information as inventory information, collection information, information on content of data set, management information, access information, and metadata management information, with 51 metadata elements. This standard was kept as flexible as possible to facilitate data submission, and allow interoperable searching. It is more convenient than the earlier China metadatabase for users to locate data sources and to submit and update metadata through the Internet. This standard is also consistent with the framework and main content of ISO/TC 211, and is more comprehensive than the Directory Interchange Format of NASA, the Metadata Standard of CIESIN, the Dublin Core Element Standard of UK, and the Core Metadata Standard of Australia and New Zealand.

There are many institutions that are engaged in the collection and management of biodiversity data in China. Insufficient attention has been paid to metadata. Metadata differs from actual data, in that it is distributed in, and related to, many institutions. Metadata must be supplied by data sources which maintain, manage and document the actual data. The provision of metadata is very complicated, especially in the field of research, because regulations on obligations and available funds are lacking. Incentive mechanisms and quality control are needed to ensure the success of data collection. Firstly, the National Biodiversity Information Committee should be established to organize and coordinate the submission of biodiversity metadata so that a partnership can be shaped. Secondly, data sources know the importance and role of metadata, and have a strong interest in the submission of metadata. Thirdly, a technical mechanism should be designed to ensure the quality of metadata.

China has set up a CHM website on the Internet (www.biodiv.gov.cn). However, a search engine driven by a metadatabase with linkages to data sources is lacking. After the formulation of the core biodiversity metadata standard, the construction

of a biodiversity metadatabase should be strengthened, which will give service to online query, management and use of biodiversity data.

References

Chrisman N.R. 1994. Metadata required to determine the fitness of spatial data for use in environmental analysis. In: Michener W.K., Brunt J.W. and Stafford S.G. (eds), Environmental Information Management and Analysis: Ecosystem to Global Scales. Taylor & Francis, London.

FGDC 1994. Content standards for digital geo-spatial metadata. Federal Geographic Data Committee, Reston, Virginia.

FGDC 1997. Content standards for digital geo-spatial metadata. Federal Geographic Data Committee, Reston, Virginia.

ISO/TC 211 WG3 1998. ISO Standard 15046-15 Geographic Information-Metadata, V.4.4.

Reynolds J., Bray S., Stone S. and Fisher T. 1997. Pre-feasibility study on the publishing of data sets and metadata. European Commission Contract No. 12055-96-07F1EI ISP GB.

Slagle R.L. 1994. Standards for integration of multisource and cross-media environmental data. In: Michener W.K., Brunt J.W. and Stafford S.G. (eds), Environmental Information Management and Analysis: Ecosystem to Global Scales. Taylor & Francis, London.

State Environmental Protection Administration of China 1998. China's Biodiversity: A Country Study. China Environmental Science Press, Beijing, China.

Xu H.G., Gao Z.N., Xue D.Y. and Wu X.M. 1999. China national biodiversity information query system. Journal of Environmental Management 56: 45–59.

Xu H.G. and Wu X.M. 1999. The study and development of China biodiversity homepage. Chinese Environmental Sciences 19: 313–317.

Xu H.G., Wang D.H. and Sun X. 2000. Biodiversity Clearing-House Mechanism in China: Present Status and Future Needs. Biodiversity and Conservation 9: 361–378.