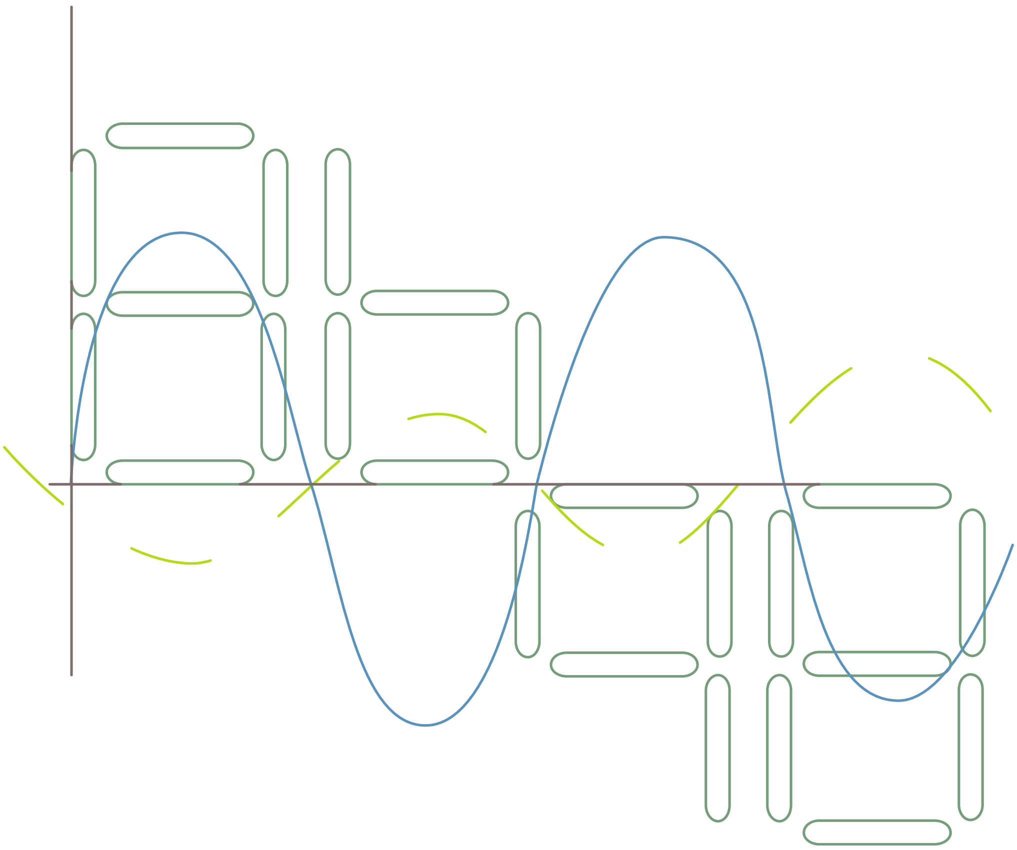
Report 31

GRDC Report Series

Globally agreed standards for metadata and data on variables describing geophysical processes

A fundamental prerequisite for an integrated global data and information infrastructure and thus improved management of the Earth System for our all future



Report 31

GRDC Report Series

Globally agreed standards for metadata and data on variables describing geophysical processes

A fundamental prerequisite for an integrated global data and information infrastructure and thus improved management of the Earth System for our all future

Thomas Maurer
(thomas.maurer@bafg.de)

Global Runoff Data Centre

GRDC operates under the auspices of the World Meteorological Organization (WMO) with the support of the Federal Republic of Germany within the Federal Institute of Hydrology (BfG)

Global Runoff Date Centre

in the
Federal Institute of Hydrology (BfG)

Am Mainzer Tor 1
56068 Koblenz, Germany

P.O.Box 20 02 53
56002 Koblenz, Germany

Phone: +49 261 1306-5224
Fax: +49 261 1306-5280
E-Mail: grdc@bafg.de
Internet: http://grdc.bafg.de

October 2004

Contents

1 Summary 1

2 Introduction 2

3 I have a(nother) dream... 3

4 Metadata, data and standards 5

4.1 General 5

4.2 Cost and benefits of metadata 6

4.3 Metadata standards 8

4.4 Requirements for a metadata standard 8

4.5 ISO 19115 - an international metadata standard for geographic information 9

4.6 WMO Core Metadata Standard 15

5 Reasoning about required metadata elements- a GRDC view centred on river discharge data 17

6 Approaches to a global data infrastructure 20

6.1 General 20

6.2 Example SIRIUS 22

6.3 Example MERCURY 23

6.4 Oceans Information Technology Pilot Project (OIT) 24

6.5 Future WMO Information System (FWIS) 25

6.6 INfrastructure for SPatial InfoRmation in Europe (INSPIRE) 27

6.7 Implementation Plan for the Global Observing Systems for Climate (GCOS-IP) 29

6.8 Global Earth Observation System of Systems (GEOSS) 29

7 Conclusion and Outlook 32

7.1 “Data” versus “model” integration 32

7.2 Further Development - Required long term vision 33

7.3 Perspectives for GRDC operation 36

8 References 38

8.1 Literature 38

8.2 Acronyms and Links 42

Annex 1

Example of application of the Draft WMO Core Metadata Standard to the GRDC database

Annex 2

Proposed Version 0-2 of the Draft WMO Core Profile (Sep 2004)

About the author:

Thomas Maurer has a background of a civil engineer with a specialisation in hydraulic engineering, hydrology and numerical modelling. In 1997 he earned a PhD from the University of Karlsruhe, Germany, in the field of hydrology, dealing with transport of water and matter in small rural watersheds. From 1998 he gained practical experiences in an engineering consultant company in the field of hydroinformatics before he changed to his current position as Head of the Global Runoff Data Centre (GRDC) in 2000.

About the Global Runoff Data Centre (GRDC):

The GRDC is acting under the auspices of the World Meteorological Organization (WMO) and is supported by WMO Resolutions 21 (Cg XII, 1995) and 25 (Cg XIII, 1999). Its primary task is to maintain, extend and promote a global database on river discharge aimed at supporting international organisations and programmes by serving essential data and products to the international hydrological and climatological research and assessment community in their endeavour to better understand the earth system. The GRDC was established at the Federal Institute of Hydrology in 1988. The National Hydrological and Meteorological Services of the 187 member states of WMO are the principal data providers for GRDC.

Globally agreed standards for metadata and data on variables describing geophysical processes

A fundamental prerequisite to improve the management of the Earth System for our all future

1 Summary

While we strive for increasing integration of our understanding of the Earth System across disciplines and scales our need for reliable, available and accessible data is increasing. At the same time the amount of data collected, both in situ and from satellites, is growing dramatically. These two tendencies perfectly coincide at the first glance, and the seemingly ever accelerating development of information technology also seems to provide the tools required to succeed in the fusion of models of higher and higher complexity and larger and larger amounts of data to advance the understanding of the Earth System. However, we are lacking a fundamental scheme to organise all the information available, and the danger is real that we will miss out and over time even loose much of what has been collected spending enormous amounts of tax payers money. The terms "information infarct" and "information tsunami" have been used to describe this unfavourable situation. A crucially important, though not sufficient prerequisite for the advancement of science is to ensure sustainable order and overview of all these valuable resources, similar as it is done with printed material in professional managed, well-organised library systems. The key is the development and rigorous application of a standardised way of thoroughly documenting data and information objects by means of general accessible, comprehensive metadata. The ISO 19100 family of standards provide a fundament to develop guidelines and tools to arrive at a global geosphere information system. No funding of observations should be granted by any funding organisation unless it is ensured, that all measurements made are thus documented and preserved.

Today, many and an increasing number of data integration efforts mushroom all over the globe and disciplines, however from a future perspective they still look scattered and fragmented. This report is a plea for the initiation of an international coordinating mechanism or body that guides in consultation with all major players involved (international

organisations, governmental authorities, leading companies, etc.) a process of defining the fundaments for an global geosphere information system by developing standards and a data infrastructure for geophysical, biogeochemical and socioeconomic variables observed and predicted in the geosphere.

2 Introduction

An inevitable prerequisite for the sustainable management of the complex earth system respectively parts or sub-systems of it is unrestricted access to sound and comprehensive data and information on the state variables and fluxes of the governing processes which we try to mimic in computer models of ever growing complexity and refinement.

Besides the extension of operational monitoring and observation networks itself, there is the urgent need for the development of a more general, globally standardised data infrastructure ensuring time saving, highly automated access to the huge variety of observational data. However, authority over data and information, especially in the terrestrial domain is often scattered regionally and sectorally, resulting in highly fragmented approaches to their management. Consequently, researchers and managers striving for integrated approaches including the development of indicators are on the horns of the dilemma of either spending too much of their valuable time on searching, retrieving and organising fundamental data (which, at a large scale, is a non-trivial task for which they typically are not optimally trained) or alternatively omitting relevant information, both being unprofessional approaches that ultimately lead to stagnation in the development of suitable solutions. Though the question of how to cope with the challenges of the earth system's future stands high on the agenda of international organisations and consequently related meetings are mushrooming all around the world, yet an overarching rigorous approach aimed at tackling the fundamental data organisation issue is pending. Without doubt there are many more prominent or more exciting and eye-catching (in one word: sexy) problems of scientific, technological, political and financial nature, however their solutions all heavily rely on the improvement of the organisational aspect of data (further reading e.g. Harmancioglu et al., 1997, Maurer, 2003a, 2003b; GRDC, 2003; JCOMM, 2002; WMO, 2004a).

Here it is argued that a more rigorous "bookkeeping" of information is required, i.e. the development and application of international standards for integrated documentation of measurements taken of geophysical processes, much like those librarians have already developed for their information objects, i.e. printed material. Standards and integrating technologies suitable for creating machine readable (meta-)databases and structures are emerging and need to be extended and generalised. The implicit higher level of organisation certainly will imply the acceptance of societies to spend increasing overheads for the integration of data and information. However, considering the enormous resources that are expended to make earth observations, both in situ and from satellites, it should be obligatory to make all necessary efforts to guarantee persistent accessibility of the valuable output of these investments. A global data infrastructure is key to preserving humankind's memory.

This report aims at contributing some thoughts towards the goal of standardised interfaces as a fundamental part of a global data infrastructure. The starting point of reasoning are experiences of the Global Runoff Data Centre (GRDC) and its database of river discharge data. Deliberations and examples will thus frequently relate to this specific field. However, in view of the fact that all kinds of earth system data share a large common denominator of descriptive metadata, an attempt is made here to develop a generalised scheme, suitable for super-ordinated classes of data along the following chain of generalisations (the previous class is always a subset of the following one):

- river discharge data
  - hydrological data
    - terrestrial earth science data
      - data on geophysical processes in all spheres in general

3 I have a(nother) dream...

In appropriation of the title of the famous speech given by Martin Luther King from the steps at the Lincoln Memorial in Washington D.C. on 28 August 1963 on an admittedly completely different topic I would like to outline a vision which I regard - in its domain - as important as King's plea. The bottom line is, that in essence its realisation will converge towards the same goal: Peaceful coexistence of humans on the planet. In the long term this will only be achieved if we manage to sustain our all environment, which in turn requires to understand the Earth System, which in turn requires to have access to fundamental data on the state variables and fluxes of the processes governing its complex interplay.

I thus have a dream of researchers and decision makers sitting in front of terminals of expert systems, formulating their questions and scenarios with ease and being served with the best possible projection and answer in due time. This inevitably will require the operation of complex models in the background, integrating all our understanding of earth system processes in the various spheres which have been analysed separately in past but which we nowadays know are dynamically linked across scales and disciplines. Achieving this will not be possible without coupling models of these processes in nested hierarchies and feeding them time and again with the most recent and comprehensive versions of observations and measurements taken both, in situ and by satellites (as well as with historic data). One fundamental (but not sufficient) prerequisite for such a system is to always know with minimal effort where which information is located, who owns it, how it can be accessed under which conditions, what are its limitations and what is its generation history. This is a plea for preparing the fundaments for an universal library and retrieval system on earth science data, that can both, be consulted in an interactive mode but also be queried automatically by machines. What does this require?

Required are three components, that build on each other. Only getting them right in the following order will help to ensure interoperability:

1. A globally standardised archival scheme (metadata) (-> common *vocabulary*)
2. Globally standardised interfaces to the archives based on standardised metadata (-> common *grammar*)
3. Globally agreed data and information infrastructure, i.e. a virtual global system of interlinked systems, based on a network of databases and models that exchanges all required information on request. (-> common *language*).

   Such an infrastructure had to consider practical constraints such as:

   - storage capacity
   - processing capacity (bandwidth, FLOPS of CPU, storage access time, etc.)
   - costs
   - right on intellectual property

Systems like this have been developed already to a high degree in many areas, e.g. for printed material. Entries on library filing cards have been standardised and are stored electronically (archival scheme). Interchange formats for transporting this information have been developed (interfaces) and the inventory of numerous libraries can nowadays be explored from all over the world via web browser based client software with only a single query (infrastructure). Often an order for a book can be initiated by mouse click - thus, even exotic books are on our fingertips.

Systems like this have already been developed and are being developed all the time by communities dealing with subsystems of the Earth System. However, the problem is that they lag behind the requirements of today's demands. While we aim at integrating more and more spheres, aspects and scales of the Earth System the integration and organisation of the associated basis information is not keeping up. Though initiatives of largely varying rigour are dynamically mushrooming all around the place yet a common denominator is lacking. A global system as required will not automatically evolve in an uncoordinated bottom-up approach unless at least some definitions are given in a top-down manner. To be practical, an intelligent compromise of both approaches has to be found, to not throw overboard the many solutions already achieved to date.

4 Metadata, data and standards

4.1 General

"Metadata"[1] is a term used within the computer science community to denote characteristics or quality of data. There are many definitions of what is classified as metadata, mostly dependent on the expected uses of the metadata. In an extreme view, metadata may mean absolutely everything ancillary to the "datum" or measurement (meaning there is little real data and lots of metadata). On the other extreme, metadata may be just a few identified properties that support a certain application.

[1] The following paragraph is adapted from a tutorial of the FGDC, the US Federal Geographic Data Committee (available at http://www.fgdc.gov/clearinghouse/clearinghouse.html) with slight changes and extensions.

Metadata[2], commonly defined as "data about data", is a structured summary of information that describes data. The term, however, is not restricted to descriptions of data. More broadly defined, metadata is descriptive information about any object or resource, as diverse as geospatial and non-geospatial datasets, data analysis tools, computer models, websites, graphics and textual information. Metadata may thus more up-to-date be defined as "supplementary information at a higher level of abstraction of information on a lower level of abstraction". At a minimum, metadata consists of the standard bibliographic information that supports resource discovery (discovery level metadata). However, it generally contains information that supports a wider range of operations, such as management, evaluation, access and use. Thus, a comprehensive metadata standard would contain sufficient descriptors to allow for automatic processing of the data by ensuring machine readability of a self explanatory format. Data that is not documented in accordance with a standard cannot be found by queries to search engines. Such data does not exist or at least is known only to a comparatively small community of insiders. Those who nowadays do not manage their data clearly and informatively and do not boost them own data that is not existent!

4.2 Cost and benefits of metadata[3]

Creating and managing metadata does involve a significant effort and hence introduces additional costs to managing an information resource. These costs can be reduced by ensuring metadata is produced at the same time and by the same people as the data is. The information needed to create metadata is often readily available when the data is collected.

An analysis which weighs the initial expense of documenting data against the potential costs of duplicated or redundant data generation will determine whether the documentation of metadata is justified. In general, investing the time and resources at the beginning of a new project will be found to pay dividends

---

[2] The following two paragraphs are a slightly changed and extended adaption from an article of SEDAC, the Socioeconomic Data and Applications Center (available at http://sedac.ciesin.org/metadata/overview.html). SEDAC is one of the Distributed Active Archive Centers (DAACs) in the Earth Observing System Data and Information System (EOSDIS) of the U.S. National Aeronautics and Space Administration (NASA). SEDAC focuses on human interactions in the environment. Its mission is to develop and operate applications that support the integration of socioeconomic and Earth science data and to serve as an "Information Gateway" between the Earth and social sciences.

Metadata produces benefits for both producers and users of data. These are outlined below:

Information Investment Management: Metadata helps organise and maintain an organisation's investment in data and provides information about an organisation's data holdings in catalogue form.

Greater Information Efficiency: Coordinated metadata development avoids duplication of effort by ensuring the organisation is aware of the existence of data sets.

Provide Complete Information: Users can locate all available data relevant to an area of interest.

Better Information Practice: Collection of metadata reinforces good data management practices (including fitness for purpose assessments) and ensures the long term value of the investment in data creation and collection.

Information Promotion: Data providers are able to advertise and promote the availability of their data and potentially link to on line services (eg. e-government) that relate to their specific datasets. Reporting of descriptive metadata also promotes the availability of environmental data beyond the environmental community.

Knowledge Management: Metadata is an important knowledge management tool preserving understanding, and preventing data from losing its value due to personnel change in an organisation.

Greater Information Longevity: Metadata maintains the value of data for the creator by assuring its continued use and update over time.

Users understand dataset: Metadata enables users to understand the purpose and intention of the dataset, and so be better able to know how to use the data and also determine its fitness for a particular use.

---

[3] This section is taken from chapter 3 of the very readable and informative New Zealand Government Geospatial Metadata Standard DRAFT v.1.2, Part 2 – Profile Guidelines, prepared by New Zealands Geospatial Metadata Project Team in June 2004, current versions are available at http://www.linz.govt.nz/rcs/linz/pub/web/root/core/Topography/ProjectsAndProgrammes/geospatialmetadata

4.3 Metadata standards

Because metadata serves a diversity of uses, there exist a number of standards (refer e.g. to http://sedac.ciesin.org/metadata/overview.html). These standards differ greatly in the level of information they support. Essentially, one can look at both the uses of metadata, and the various standards along a continuum of complexity. The most basic record enables data and resource discovery, much like records in a library catalogue, whereas the most complex provides essential information for processing and interpreting data, much like a user manual. Metadata facilitates comparisons between datasets from different sources and when placed in a searchable index, enables searching of domain specific information, such as geographic location, title or data type. Metadata may also serve as a tool for organising and maintaining an organisation's investment in its data, by providing a systematic way of recording information about the data it produces. Metadata may even provide protection for the producing organisation if a conflict arises over the misuse of data. In essence, metadata is documentation that can answer the who, what, when, where, why and how questions, describing every facet of the data or resource being documented - its content, quality, accessibility, collection methods, processing and availability.

4.4 Requirements for a metadata standard

According to KOGIS (2001) the following requirements are to be meet by an ideal standardisation of metadata in order to generate benefit:

Generality

The structure of a metadata model has general, i.e. independent from requirements of a specific area of business. It thus should be possible to accommodate a structured description of all types of themes.

Completeness

It should be possible to define all relevant attributes describing a large variety of special subjects. Depending on the application also a flexible abstraction should be permitted.

Flexibility, extensibility

As the predefined descriptors of a standard will never meet all requirements for any specific area, a standard should be sufficiently flexible to allow for the definition of additional descriptors. A standards structure should also allow for extensions as user needs and prerequisites change over time. Flexibility can be understood as the capability of a standard to be adjusted to the needs of a existing data scheme.

Interchangeability

Producers, administrators, and users of metadata should be able to exchange metadata. This requires standardised access tools to metadata.

Usability, comprehensibility

In order to guarantee a wide spread and acceptance of a standard, it has to be comprehensible and easy to use. This requires the ability to accommodate the standard in user friendly IT-environments.

Independence

A metadata standard should be arrange such that a metadatabase can be operated independently from a database. Besides data stored in databases it also should allow for inclusion of data (still) only available in non-digital form.

Ease of implementation

A standard should be defined such that it can be easily, non-ambiguously and uniformly implemented in a software such as input clients, search engines, clearinghouses, models etc. in order to ease the computer-based processing of metadata and data.

4.5 ISO 19115 - an international metadata standard for geographic information

There have been considerable recent achievements in the field of standardisation of data and metadata representation and transfer, especially in the field of geomatics. The International Organization for Standardisation (ISO[4]) is a non-governmental organisation founded 1947 as the successor of the International Federation of the National Standardising Associations (ISA), founded in 1926, that establishes standards to facilitate the international exchange goods and services. In 1994 the ISO Technical Committee 211: Geographic

Information/Geomatics (ISO/TC 211) was established to develop a set of standards (ISO 19100 series) in the field of digital geographic information, aiming to establish a structured set of standards for information concerning objects or phenomena that are directly or indirectly associated with a location relative to the Earth (as are also all measurements taken of state variables of geophysical processes, including water–related ones!). ISO/TC 211 built on achievements of the Technical Committee 287 of the Comité Européen de Normalisation (CEN/TC287) which ceased its activities (that started earlier in 1991!) to avoid duplication of work. A good overview of the subject of standardisation of geographic information, including its historical development is given in Kresse & Fadaie (2004).

In May 2003 ISO/TC 211 released the metadata standard ISO 19115 as a result of intensive consultations of organisations pioneering this field including the NATO Geomatics Working Group called Digital Geographic Information Working Group (DGIWG), the International Hydrographic Organization (IHO), the US Federal Geographic Data Committee (FGDC) and the OpenGIS Consortium. Its development was influenced by several standards, including the FGDC and ANZLIC, but is more comprehensive than any of them. The FGDC and ANZLIC are establishing "profiles" that will consist of metadata elements unique to the ISO 19115 standard. All the major standards will be interoperable with this standard.

ISO 19115 defines more than 300 metadata elements (86 classes, 282 attributes, 56 relations), most of which can be applied optionally. At the topmost level the classes (or entities) are grouped in 14 packages (the "root" "Metadata entity set information" plus 13 dependent packages) which are thus also available for use in other TC 211 standards as appropriate. The obligation of the various entities or classes is quite variable and flexible. The complex, hierarchical nested structure and relationships between the components are shown using 16 UML diagrams (ISO 19115, Annex A). Additionally, the definitions are listed in a tabulated dictionary (ISO 19115, Annex B). Both the UML diagrams and the dictionary are normative.

Below, the 14 top level packages along with their entity names are listed. Only the first two packages are mandatory indicated by (M) while the remaining packages are optional, indicated by (O). However, once a package is chosen, again within each package mandatory (M), optional (O) and "mandatory under certain conditions" (C) elements exist. Note that an

---

[4] According to Kresse & Fadaie (2004), ISO is not an acronym, but rather a word derived from the Greek *isos*, meaning

element in general may be a class or an attribute, i.e. there may be nested hierarchies of (sub)classes and also multiple use of classes in different (super)classes (e.g. to accommodate contact information). Elements from underlined packages are also used by the WMO Core Metadata Standard V0.1 as described further below.

Metadata entity set information [MD_Metadata] (M): is the top level package and contains general information on the metadata set as e.g. date, contact, language, character set used etc.

Identification information [MD_Identification] (M): contains information to uniquely identify the data. Identification information includes information about the citation for the resource, an abstract, the purpose, credit, the status and points of contact.

Constraint information [MD_Constraints] (O): contains information concerning the restrictions placed on data.

Data quality information [DQ_DataQuality] (O): contains a general assessment of the quality of the dataset.

Maintenance information [MD_MaintenanceInformation] (O): contains information about the scope and frequency of updating data.

Spatial representation information [MD_SpatialRepresentation] (O): contains information concerning the mechanisms used to represent spatial information in a dataset.

Reference system information [MD_ReferenceSystem] (O): contains the description of the spatial and temporal reference system(s) used in a dataset.

Content information [MD_ContentInformation] (O): contains information identifying the feature catalogue used and/or information describing the content of a coverage dataset.

Portrayal catalogue information [MD_PortrayalCatalogueReference] (O): contains information identifying the portrayal catalogue used.

Distribution information [MD_Distribution] (O): contains information about the distributor of, and options for obtaining, a resource.

Metadata extension information [MD_MetadataExtensionInformation] (O): contains information about user specified extensions.

Application schema information [MD_ApplicationSchemaInformation] (O): contains information about the application schema used to build a dataset.

Extent information [EX_Extent] (O): contains information that describe the spatial and temporal extent of the referring entity.

Citation and responsible party information [CI_Citation + CI_ResponsibleParty] (O): contains information needed for citing a resource (dataset, feature, source, publication, etc.), as well as information about the party responsible for a resource.

"equal", which points to one of the goals of international standardisation.

ISO 19115 defines an extensive set of metadata elements; typically only a subset of the full number of elements is used. However, the ISO 19115 specification also summarises the generic **core metadata** comprising the minimum elements that satisfy the requirements of an ISO conformant metadata record, again including "mandatory" (M), "optional" (O) and "mandatory under certain conditions" (C) elements.. Listed are the core metadata elements required to identify a data set, typically for catalogue purposes. This list contains metadata elements answering the following questions:

- Does a data set on a specific topic exist (‘**what**’) ?
- For a specific place (‘**where**’)?
- For a specific date or period (‘**when**’)?
- A point of contact to learn more about or order the dataset (‘**who**’)?

Using the recommended optional elements in addition to the mandatory elements will increase interoperability, allowing users to understand without ambiguity the geographic data and the related metadata provided by either the producer or the distributor. All metadata profiles based on ISO 19115 shall include this core. Examples of core metadata records end up containing a comparable amount of information as a typical record by other standards (i.e. 20-40 entries). The following table 1 lists all core elements (after table 3 of ISO 19115):

Table 1: ISO 19115 Core metadata for geographic datasets (M: mandatory, C: mandatory under certain conditions, O: optional)

Core metadata for geographic datasets	obligation	UML hierarchy
Dataset title	(M)	(MD_Metadata > MD_DataIdentification.citation > CI_Citation.title)
Dataset topic category	(M)	(MD_Metadata > MD_DataIdentification.topicCategory)
Abstract describing the dataset	(M)	(MD_Metadata > MD_DataIdentification.abstract)
Dataset reference date	(M)	(MD_Metadata > MD_DataIdentification.citation > CI_Citation.date)
Dataset language	(M)	(MD_Metadata > MD_DataIdentification.language)
Metadata point of contact	(M)	(MD_Metadata.contact > CI_ResponsibleParty)
Metadata date stamp	(M)	(MD_Metadata.dateStamp)
Dataset character set	(C)	(MD_Metadata > MD_DataIdentification.characterSet)
Geographic location of the dataset (by four coordinates or by geographic identifier)	(C)	(MD_Metadata > MD_DataIdentification.extent > EX_Extent > EX_GeographicExtent > EX_GeographicBoundingBox or
Metadata language	(C)	(MD_Metadata.language)
Metadata character set	(C)	(MD_Metadata.characterSet)
Dataset responsible party	(O)	(MD_Metadata > MD_DataIdentification.pointOfContact > CI_ResponsibleParty)
Additional extent information for the dataset (vertical and temporal)	(O)	(MD_Metadata > MD_DataIdentification.extent > EX_Extent> EX_TemporalExtent or EX_VerticalExtent)
Spatial resolution of the dataset	(O)	(MD_Metadata > MD_DataIdentification.spatialResolution > MD_Resolution.equivalentScale or MD_Resolution.distance)
Spatial representation type	(O)	(MD_Metadata > MD_DataIdentification.spatialRepresentationType)
Reference system	(O)	(MD_Metadata > MD_ReferenceSystem)
Lineage	(O)	(MD_Metadata > DQ_DataQuality.lineage > LI_Lineage)
Distribution format	(O)	(MD_Metadata > MD_Distribution > MD_Format.name and MD_Format.version)
On-line resource	(O)	(MD_Metadata > MD_Distribution > MD_DigitalTransferOption.onLine > CI_OnlineResource)
Metadata file identifier	(O)	(MD_Metadata.fileIdentifier)
Metadata standard name	(O)	(MD_Metadata.metadataStandardName)
Metadata standard version	(O)	(MD_Metadata.metadataStandardVersion)

The scope of the information covered by ISO 19115 is broadly that needed for a user to identify, evaluate, select, obtain, and possibly use the data sets described. The level of detail is much greater than other standards and the manner in which it is provided is much more structured. The content models are elaborate and hierarchical, and for those elements that still contain text (such as e.g. addresses), the information is generally disaggregated more finely (e.g. street, city, zip-code, country etc.). This supports the construction of more elaborate interfaces and more finely controlled queries, but places a much greater burden on the metadata provider and tool developer. ISO 19115 attempt to cover the needs of a wide range of potential applications, but at the cost of a rather daunting structure.

However, rules are given on how to define a **community or domain profile** (in ISO 19115 Annex C, see also box 1) which limits the elements used or the values or obligations of components of the standard, and also on how to add specialised extensions where it is found that the requirements are not satisfied by the components already defined in ISO 19115 (though the need for the latter is considered minimal and clearly discouraged!).

A very good example of documentation of the process of developing a community or domain profile is the New Zealand Government Geospatial Metadata Standard DRAFT v.1.2 (New Zealand Government, 2004).

Overall, ISO 19115 aims to define a comprehensive range of metadata elements that may be needed, so that any single application domain will normally select only a subset of the components available. The standard does not specify how metadata should be archived or presented to users and does not specify any particular implementation. It could be implemented as a database, a flat file, or any other suitable mechanism.

The OpenGIS Consortium aims to provide a comprehensive suite of open interface specifications to enable transparent access to heterogeneous geo-data and geo-processing resources in a networked environment. They work on the implementation of these standards applying the Extensible Markup Language (XML), which is an internet standard approved by the World Wide Web Consortium (W3C) allowing the separate definition of the logical and physical structure of a documentation object.

BOX 1: Metadata Community Profile according to ISO 19115, Annex C.5

ISO 19115 provides a mechanism to extend the metadata definitions. If the information to be added is extensive, involving the creation of many metadata elements within a metadata entity, specific to a discipline or application, coordination of the proposed extension via user groups and creation of a community profile is recommended. ISO 19115 defines almost 300 metadata elements, with most of these being listed as "optional". They are explicitly defined in order to help users understand exactly what is being described. Individual communities, nations, or organisations may develop a "community profile" of this International Standard. They will make a select set of metadata elements mandatory. A given metadata element (e.g., the "price" of a dataset) may be established as "mandatory" for a certain community that will always want that metadata element reported. A community of users may want to establish additional metadata elements that are not in this International Standard. For example, a community may want to develop metadata elements for the status of datasets within their system to help manage production. However, these added elements will not be known outside the community unless they are published. A community profile should establish field sizes and domains for all metadata elements to insure interoperability. See ISO 19106 for more information on community profiles. Figure B-1 illustrates the relationship between the Core Metadata components, the comprehensive metadata application profile and national, regional, domain specific or organisational profiles.

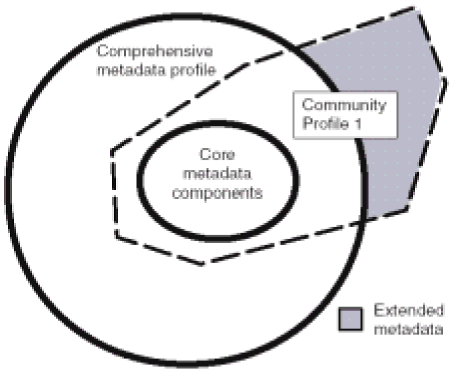


Figure B-1: *Metadata community profile as defined in ISO 19115, annex C.5. The inner circle contains the core metadata components. The comprehensive metadata includes the core metadata components. A community profile shall contain the core metadata components, but not necessarily all the other metadata components. Additionally it may contain metadata extensions (shaded area) which shall be defined following the metadata extension rules in ISO 19115, annex C.*

Currently TC 211 plans an extension for imagery and gridded data in ISO 19115-2 (i.e. part 2) by 2006. Furthermore, TC 211 is planning to finalise a metadata dataset implementation specification (ISO 19139) by November 2004. Given the abstract nature of the ISO 19115 specification, the actual execution of geographic information metadata could vary based on the interpretation of the metadata producers. In an attempt to facilitate the standardisation of implementations, the comprehensive dataset metadata implementation specification (ISO 19139) provides a single UML interpretation and a XML schema document (XSD) as an implementation of the ISO 19115 metadata standard. This specification is meant to enhance interoperability by providing a common specification for describing, validating and exchanging metadata about geographic datasets.

4.6 WMO Core Metadata Standard

In a series of workshops (ET-IDM 2001, 2002, 2003, 2004), an Expert Team on Integrated Data Management (ET-IDM) of WMO's Commission For Basic Systems (CBS) has developed a Draft WMO Core Metadata Standard (v0-2) based on ISO 19115 (WCMS, http://www.wmo.int/web/www/WDM/Metadata/documents.html). This core provides a general definition for directory searches and exchange that should be applicable to a wide variety of WMO datasets at the discovery level. Like the fundamental standard ISO 19115, the WMO core metadata standard does neither specify how these metadata should be archived or presented to users nor does specify any particular implementation (this is currently underway with the development of ISO 19139, see above).

As already mentioned above, the ISO 19115 specifies a process (in ISO 19115, annex C) where a community can adopt parts of the standard which it feels relevant (including the "Core Elements") and also extend the elements, keywords and code table instances to suite that community (see also box 1). The ET-IDM (2001) noted that the WMO might need to accept more than one Community Core Profile (CCP) according to ISO 19115 Annex C.5 for the different WMO Programmes, but that there should be a CCP which could be adopted by all of WMO, with the potential for further extensions under ISO 19115 Annex C where necessary.

The core elements define a minimum set of information required to exchange data for WMO purposes and are not exhaustive. To fully meet the requirements of all WMO Programmes for metadata, application of far more comprehensive standards would be required. The ET-IDM (2002) felt that the development of a comprehensive WMO metadata standard would be a difficult, lengthy and expensive undertaking (here the author of this report agrees) and the potential benefits of a such a standard would be very limited and would not justify the large commitment of resources that would be required (here the author of this report completely disagrees, for reasons that are implicit throughout the report!). ET-IDM (2002) suggested that each WMO Programme use the WMO Core Metadata as a starting point to develop more detailed metadata standards (CCPs) in response to its own requirements. These more-detailed programme-specific standards should be based on the ISO standard with any necessary extensions. Reliance on the ISO standard as a common starting point would reduce the effort required by the Programmes and would greatly enhance the compatibility between the various Programme-specific standards and with the WMO Core Metadata standard.

There are many possible ways of representing WMO metadata and ET-IDM (2002) recommended that XML be adopted as the common language (or format) for exchange. To ensure interoperability, the experts developed a framework, as an XML Schema, for mapping the proposed metadata standard into XML. It will have to be harmonised with the currently developing comprehensive data set metadata implementation specification (ISO 19139, see previous section). Annex 1 shows the attempt to document the entire GRDC database using the current XML implementation of the WMO Core Metadata Standard.

However, if one tries to figure out which elements the ISO 19115 standard are and especially which are NOT applied in the WMO Core Metadata Standard V0.2, one stumbles over some incongruities. E.g., for identification information ISO 19115 requires the following hierarchy of metadata information objects (cf. ISO 19115, Annex B.2.2):

MD_Metadata (1) > MD_Identification (23) > MD_DataIdentification (36)

while the WCMS seems to follow a different structure, i.e. leaving out one level of hierarchy:

MD_Metadata (1) > MD_DataIdentification (36)

Correspondingly, some of the attributes that are associated with 2 different classes in ISO are merged together in one class in WCMS. This could be due to the application of a draft version of ISO 19115 during development of WCMS in 2002 (there also seem to still exist some contrariness in the ISO 19115 standard, comparing e.g. some of the hierarchies given in table 3 on core metadata, as opposed to those listed in the normative data dictionary for geographic metadata in annex B.2.2.)

The Draft WMO Core Metadata Standard as currently provided as a XML representation at http://www.wmo.int/web/www/WDM/Metadata/documents.html thus probably will require further development regarding:

- conformance with ISO 19115
- conformance with ISO 19139
- possibly definition of additional items
- provision of a more comfortable environment to apply it as an incentive to data producers to apply it.

Furthermore, the above cited web page on the WCMS provides the XML-scheme only (i.e. the XSD-file, however no example XML metadata set (for use as a template) and no example XSL stylesheet to look at a XML metadata set in an easy to read way is supplied.

In addition, a metadata input client would be desirable, i.e. a software that reads the XML-scheme (XSD-file) and that offers a form to input the requested fields by a human-friendly interface and finally dumps a XML representation (which normally only should be dealt with by machines).

Finally, a graphical interactive stylesheet (XSL) generator software would be desirable, that reads the XML-scheme (XSD-file) and allows to design an human-friendly, easy to read output format (in HTML).

5 Reasoning about required metadata elements- a GRDC view centred on river discharge data

As it has been stated before data can be described on different levels, ranging between the discovery level and sophisticated data description level. Though progress regarding the standardisation of the first level will already yield significant improvements as it will allow humans to *interactively* browse through catalogues of metadata in a semi-automated way, using comfortable filters to chase down their target, similar to what we know from library search clients.

On the long run however, the degree of automation will have to increase to allow fully unattended access to data by computer programs. This consequently will require an even higher degree of abstraction and formalisation of metadata to make it fully *machine readable*, likely also requiring to break down information units in even smaller standardised units. Essentially, it is required to avoid every kind of input of descriptive text, but rather code the same statement by using a number of (possibly hierarchically nested) codes from a predefined set (as e.g. supported by the code lists structure within ISO 19115).

An example for this is breaking down an address (of, say, a data owner), which in currently used schemes often is a lumped text block, into its atoms, i.e. prefix, title, first name, name, street, house number, town, zip-code, state etc. Only this way an address can be exchanged across different systems or thoroughly be analysed for e.g. regions, as no procedure can in general dissemble an address given as a text block unambiguously into its smallest parts. In fact this is one of the problems encountered in current migration activities from older and simpler standards to ISO 19115.

An address is also a good example for the aspect of the need to use code lists, as free text inputs, though they intend to express exactly the same thing, will in general differ from each other, be it due to transcription, the existence of multiple names, languages, abbreviation or the like. This is also exactly the reason why organisations come up with standardised keyword lists to describe their information. On a global scale this extends to the issue of developing all-embracing multilingual thesauri, as e.g. the GEneral Multilingual Environmental Thesaurus (GEMET, http://www.eionet.eu.int/GEMET) of the European Union, which is nothing else than a already quite generalised code list. As long as several code lists are used, it will be necessary to map them into each other (which is not a trivial task!) .

So what do we need to know to thoroughly describe a data set? And is it possible to describe a data set in an entirely generalised form, that is independent from the actual physical variable? What does this require? And is it already laid out in the ISO 19115 standard or does it require extensions?

It will not be possible to think of all necessary items in one step. But it will be important to communicate to all involved players in the standardisation endeavour the need of adjusting their developments from time to time, which of course will require resources for maintenance.

In the following an attempt is made to match information objects required to describe river discharge time series with the corresponding objects available in the ISO 19115 as well as in the WMO Core Metadata Standard. Due to the complexity of the standard this overview is still incomplete, but indicated that both the ISO standard and the WMO profile allow to document the most obvious information objects. Amore detailed analysis has to follow, as the standard develops and user interfaces are generated.

Table 2: Matching information objects required to describe a river discharge data set with definitions available in ISO 19115 and the Draft WMO Core Metadata Standard (v0-2)

Information objects required to describe a river discharge data set	**ISO 19115** (in parenthesis: line number used in annex B of the standard; in blue: elements not (fully) used in WMO Satndard)	**Draft WMO Core Metadata Standard (v0-2)** (in parenthesis: line number used in annex B of the standard)
Information content Name and dimension of measured quantity/ geophysical parameter	MD_DataIdentification (36) MD_ContentInformation (232)	MD_DataIdentification (36) MD_ContentInformation (232)
Position in space and time Including three measures, the so-called scale triplet of a measurement or modelled quantity (in space x, y, z and time t) • extend • spacing • (control-)volume i.e. in general: specification of three space and one time window. Special cases can be derived from that, e.g. • time series (space window -> point), • digital elevation models (time window practical infinite)	referenceSystemInfo (13) MD_ReferenceSystem (186) EX_Extent (334) MD_Resolution (59) MD_SpatialRepresentation (156)	referenceSystemInfo (13) MD_ReferenceSystem (186) EX_Extent (334)
Accuracy	dataQualityInfo(18) LI_Lineage (82) DQ_Element (99)	dataQualityInfo(18) LI_Lineage (82)

Origin and history of generation
Which methods have already applied to the data? (could be realised by a chain of links to other metadata sets)
• author
• investigators
• processors
• measured data
  - applied measurement methods, technical devices
  - framework conditions of a measurement campaign
• model output data
  - applied models, model versions
  - parameter sets used
	LI_Lineage (82)
	LI_ProcessStep (86)
		LI_Lineage (82)

Data formats (very secondary as compared to "information content"!)	MD_Distribution (270)	MD_Distribution (270)

• interfaces • software/ operation system • hardware		
Data types • analog • digital	MD_Distribution (270)	MD_Distribution (270)
Place of storage (e.g. library, server)	MD_Distribution (270) CI_OnlineResource (396)	MD_Distribution (270)
Usage of data • owner • conditions of use • restrictions • costs	MD_Metadata (1) contact (8) pointOfContact (29) MD_Constraints (67) MD_Mainteance Information (142)	MD_Metadata (1) contact (8) pointOfContact (29) MD_Constraints (67)
Documentation	descriptiveKeywords (33) MD_ContentInformation (232) MD_BrouwseGraphic (48)	descriptiveKeywords (33) MD_ContentInformation (232)
User experiences, reports		
Applicable methods to data (=available models) • interpolation • scaling (of measurement and model scale) - downscaling, disaggregation - upscaling, aggregation • model applications	MD_Usage (62)	

6 Approaches to a global data infrastructure

6.1 General

All what is provided by a standard for metadata is the model of an interface. This is only a small, though indispensable, but not sufficient part of a larger system required. The standard has to be implemented and to be used in a framework of a global data infrastructure, often also referred to as a clearinghouse, to allow to search and retrieve data. Such an infrastructure has to define the flow of information and the location of storage, including the assignment of responsibilities of organisations to provide data, to run databases and networks.

Standards usually cover a sector (e.g the - admittedly broad - sector of geomatics, as described above) and still are not accepted or implemented everywhere on a global scale. The reasons for this are manifold and are related to the heterogeneity discussed in introductory sections.

Denzer et al. (1993, 1995b) have early pointed out clearly in the context of developing Environmental Information Systems (EIS) that no matter how much effort will be invested to integrate distributed and heterogeneously spread information of different meaning, syntax and structure, from a practical point of view there is no realistic way to combine them in a single unified or monolithic system in a reasonable time frame. According to Denzer et al. (1993, 1995b) this is due to mainly three reasons, namely heterogeneity, autonomy and dynamics. Their elaboration on these characteristics are summarised below:

Heterogeneity: In practice, different systems to be integrated are heterogeneous in different respects:

1. **Syntactical heterogeneity** means that systems differ with respect to hardware, operating system, storage technology etc. Syntactical heterogeneity is a pure computer problem and should be hidden from the user.
2. **Semantical heterogeneity** means that there are different notions about the semantics of a single piece of information. This includes the development of different terminologies in parallel projects in different regions.
3. **Structural heterogeneity** evolves due to the fact that different parties combine different sets of simple information to different structures (or objects) denoting the same type of information, but in a different way, resulting in aggregates of different syntax and different meaning (although some part of these objects may have the same meaning).

Autonomy: Many EIS which have been built are information systems for public authorities, supporting public services in their every day work. Due to the legal authority of these institutions, they are completely autonomous in their decisions concerning information technology. Due to scattered sectoral and regional competencies with regard to environmental management one is confronted with a fragmented situation of approaches in different regions and sub-regions. This holds especially true for water as a traditionally locally managed resource in many regions of the world..

The task of building a data network in such a situation means, that it is not possible to apply a unique data model for such a network because one can never force anyone to use this data model or stick to it and its enhancements.
Software developers have to accept the fact that fragmented autonomy is something that will not vanish quickly. This makes integration more difficult but not impossible, i.e. this boundary condition has severe implications on the types of software architectures which can apply to autonomous systems.

Dynamics: It is usually impossible to thoroughly describe the tasks ahead in environmental management in a single step from scratch, and this consequently holds true for the definition of a final data model too. Even if a perfect data model could be defined beforehand, it is unlikely that this model remains valid for more than one year, given the rapidly changing demands. Thinking of the integration of hundreds of environmental data sources and linking them with hundreds of thousands of potential clients, it becomes clear what it means to keep such a system up-to-date when data sources change their features all the time.

6.2 Example SIRIUS

To overcome these problems Denzer et al. (1993, 1995a, 1995b, 1998, 2000) developed an Environmental Information System (EIS) that serves as a meta-database system and as data retrieval system capable to integrate distributed existing database systems of different structure and level of abstraction without touching their grown internal structure nor necessarily their ownership, i.e. control of local administrators. This system features a flexible internet based client server architecture that ensures applicability across heterogeneous environments. The system is designed in a completely generic way by means of a communication server (termed SIRIUS, Denzer et al. 1995a) between local service programmes and distributed clients, which thus features two interfaces. The system is furthermore prepared to automatically translate all features client-dependently.

Different local systems feature different levels of abstraction; few of them give access to their catalogues, almost none of them is able to describe itself, e.g. by object-classes and structures they provide. To enable the outer world nevertheless to see what local systems have to offer in a unified way, each local system participating has to be equipped with a slim local interfacing database and a number of service applications running on the local system communicating with the communication server. These are the only parts of the system which have to be adjusted and have to restrict to some standards of the system. The local interfacing database can be regarded as the "table of contents" of the local database designated for integration and remains completely under the control and responsibility of the local administration. It defines who is permitted to view or retrieve what information and also contains the methods how to access the local system for retrieval. Once a local candidate system is set up as described its data is readily available to the outer world.

On the client side of the system the only prerequisite is a WWW-browser. A JAVA-application collects all meta-information which a user is authorised to view from the communication server and jointly displays it as a multi-hierarchical tree from which single data sets can be selected.. Alternatively the interface also allows to selectively query the metadata including location, which can be both, selected and displayed, in an integrated internet map server window.

A system like the one described by Denzer et al. (1993, 1995a, 1995b, 1998, 2000) provides the slimmest possible approach to integration by ideally combining centralised and decentralised features, thus being flexible enough to be adjusted with minimal effort to the ever changing boundary conditions as discussed above. Recapitulating, the introduction of a new data-source in the system is thus achieved

- without making changes or enhancements to the communication server;
- without making changes or enhancements to the clients, i.e. all end-user applications;
- without having to write too much new code for each new data source.

Though such an architecture may be the slimmest possible solution it still will require a minimum of standardisation, that is to say at the interfaces.

6.3 Example MERCURY

Even though Denzer et al. (1993, 1995b) was among the first to promote the ideas outlined above, there are other initiatives around showing developments along similar lines, e.g. the MERCURY approach for scientific data management launched at the Oak Ridge National Laboratory (ORNL DAAC Mercury, http://mercury.ornl.gov), a federal research facility operated by Lockheed Martin Energy Research Corporation for the US Department of Energy.

This metadata management scheme builds on existing WWW technology and commercial-of-the-shelf (COTS) products as well as on agreed metadata standards. The basic idea here is to keep metadata sets in XML format on the servers of providers and their maintenance in the hands of the providers who also maintain a "locator file" (the table of contents!) at their system. This "locator file" has to be registered with the Mercury staff. Based on this a specialised Mercury web-crawler extracts the latest versions of metadata sets in nightly "harvesting"-runs and stores them in a central database, which again is made available to the public by a web-browser application. Several US organisations already joined the system.
See an example of a GRDC metadata set compatible with MERCURY at

- http://mercury.ornl.gov/servlet/ornldaac/retrieve?pn=1&el=MercuryFull&db=current&rp=2&mr=1&ac=current&cid=67 (formatted)
- http://www-eosdis.ornl.gov/data/bluangel_harvest/RGED/curtis/metadata/hydrology/grdc.xml (xml-source)

6.4 Oceans Information Technology Pilot Project (OIT)

The International Oceanographic Data and Information Exchange (IODE, http://ioc.unesco.org/iode) network was established by the Intergovernmental Oceanographic Commission of UNESCO (IOC, http://ioc.unesco.org/iocweb) in 1960 to:

- facilitate and promote the exchange of oceanographic data and information.
- develop standards, formats, and methods for the global exchange of oceanographic data and information.

assist member states to acquire the necessary capacity to manage oceanographic data and information and become partners in the IODE network.

IODE's OceanTeacher (http://www.oceanteacher.org), which is a comprehensive self-training and resource tool for data management, is but one of the many services IODE provides to foster its aims.

The Joint WMO-IOC Technical Commission for Oceanography and Marine Meteorology (JCOMM, http://ioc.unesco.org/goos/jcomm.htm) was established in 1999 by WMO Resolution 14 (Cg-XIII) and IOC Resolution XX-12 to improve the development, maintenance, coordination and guidance of the operation of the global marine meteorological and oceanographic observing systems and supporting communications facilities of these organizations to meet the needs of the IOC and WMO Programmes and in particular of the Global Ocean Observing System (GOOS), the Global Climate Observing System (GCOS) and the World Weather Watch (WWW).

Inspite of all efforts made so far, JCOMM (2002) published a prospectus for a new comprehensive initiative in data and information management for the ocean and marine environment. The JCOMM Data Management Coordination Group (DMCG) was requested to consider the development of an Oceans Information Technology Project (OIT) in general and the implementation of an Oceans Information Technology Pilot Project (OIT, http://www.oceans-it.net) in particular. Smith (2002) emphasises the need to align ocean-data management more closely with evolving information technology. There is a need to improve the telemetry, uptake of technology, mode of data transport, and links with the scientific community, in order to create an efficient and effective data and information management system for the ocean and marine environment.

Hankin et al. (2002) argue along similar lines from the perspective of the U.S. Integrated Ocean Observing System (IOOS, http://www.ocean.us) and its Data Management And Communications (DMAC, http://dmac.ocean.us) network and subsystem.

6.5 Future WMO Information System (FWIS)

WMO's Commission For Basic Systems (CBS, http://www.wmo.int/web/www/BAS/CBS-info.html) debates since 1998 the Future WMO Information System (FWIS, 1999, 2000, 2001, 2002, 2003). As it is summarised in WMO (2004a), it became apparent that the various WMO Programmes either had already or were in the process of developing their own information systems independently of each other. Since the multiplicity of systems resulted in incompatibilities, inefficiencies, duplication of effort and higher overall costs for Members, the continued development of the systems in this uncoordinated manner would have exacerbated these problems and would have further isolated the WMO Programmes from each other and from the wider environmental community. It would have increased the difficulty in sharing information between programmes, which was essential for them to fulfil their requirements. During its fourteenth session in 2003, 5 to 24 May 2003, the World Meteorological Congress reviewed a preliminary concept proposed by CBS and confirmed that an overarching approach was required: a single coordinated global infrastructure, the Future WMO Information System (FWIS):

- FWIS would be used for the collection and sharing of information for all WMO and related
- FWIS would provide a flexible and extensible structure that would allow the participating centres to enhance their capabilities as their national and international responsibilities grew.
- Implementation of FWIS should build upon the most successful components of existing WMO information systems in an evolutionary process.
- FWIS development should pay special attention to a smooth and coordinated transition.
- The basis for the core communication network would be the present communication links used within the World Weather Watch (WWW) for the high priority real-time data.
- FWIS should utilise international industry standards for protocols, hardware and software.

FWIS is intended to serve all relevant WMO programmes. It would bring savings to the meteorological/hydrological community as a whole and increase the efficiency of their operations.

Reviewing the requirements of the different WMO programmes, the following needs were highlighted:

- A widely available and electronic (on-line) catalogue of all meteorological and related data for exchange to support WMO Programmes is required.
- It should be possible to rapidly integrate real-time and non-real-time (archive) data sets to better interpret weather events in a climatological context.
- There is a need to identify and utilise the potential of data from observation sites established by one Programme to meet the requirements of other Programmes.

- There is a need to harmonise data formats, transmission standards, archiving and distribution mechanisms to better support inter-disciplinary use of data and products.
- Standard practices for the collection, electronic archival and exchange of metadata, both discovery level and detailed, especially for stations and instruments, are needed.

FWIS will consist of three major components (which are functions, that can be assigned to different organisation or combined in one):

- National Centres (NC)
- Data Collection or Product Centres (DCPC)
- Global Information System Centres (GISC)

together with a data communication network connecting the components.

As FWIS in meant to use off-the-shelf hardware and software systems, it should be affordable and highly flexible Actual development and implementation of FWIS should be pursued through a gradual introduction and evaluation of enabling technologies through pilots and prototypes. The major innovation is needed in the development of metadata directories for which all programmes should contribute. Successful prototypes could then be expanded to serve additional communities and/or distributed to other Members and centres for wider implementation.

6.6 INfrastructure for SPatial InfoRmation in Europe (INSPIRE)

An impressive example for the increasing awareness of the importance of state-of-the-art data infrastructures by governments is the recent proposal for a directive of the European Parliament and of the Council establishing an INfrastructure for SPatial InfoRmation in Europe (INSPIRE, http://inspire.jrc.it, EC, 2004). The rational behind this proposal is given in the introduction of the in initial explanatory memorandum of EC (2004):

"Good policy depends on high-quality information and informed public participation. Policymakers have recognised the growing interconnection and complexity of the issues affecting the quality of life today, and this recognition is influencing the way new policies are now being formulated. For instance, the Sixth Environment Action Programme (6th EAP) emphasises that environment policy needs to be based on sound knowledge and informed participation, and this new approach is transforming the way EU environment policy decisions are being taken.

A new approach is therefore needed to deal with monitoring and reporting and with data management and delivery across the different levels of government. Policies need to be employed to reduce duplicated data collection and to assist and promote the harmonisation, broad dissemination and use of data. Such policies should result in increased efficiency, the benefits of which can be reinvested in improving the availability and quality of information. In turn, the increased availability of information will stimulate innovation among information providers in the commercial sector.

Spatial information can play a special role in this new approach because it allows information to be integrated from a variety of disciplines for a variety of uses. A coherent and widely accessible spatial description of the Community territory would deliver the requisite framework for coordinating information delivery and monitoring across the Community. Spatial information may also be used to produce maps, which are a good way of communicating with the public. Unfortunately, the technical and socio-economic characteristics of spatial information make the problems of coordination, information gaps, undefined quality and barriers to accessing and using the information particularly acute.

The Commission has therefore decided to submit to the European Parliament and the Council of the European Union the present proposal to make interoperable spatial information readily available in support of both national and Community policy and to enable the public to access to this information. This initiative derives from the commitment of several Commission services in particular DG Environment, Eurostat and the Joint Research Centre, who have already and will continue to play an important role in the adoption and implementation of this Directive."

"..."

"The proposed Directive creates a legal framework for the establishment and operation of an Infrastructure for Spatial Information in Europe, for the purpose of formulating, implementing, monitoring and evaluating Community policies at all levels and providing public information.

A key objective of INSPIRE is to make more and better spatial data available for Community policy-making and implementation of Community policies in the Member States at all levels. INSPIRE focuses on environmental policy but is open for use by and future extension to other sectors such as agriculture, transport and energy."

"..."

"INSPIRE will not set off an extensive programme of new spatial data collection in the Member States. Instead, it is designed to optimise the scope for exploiting the data that are already available, by requiring the documentation of existing spatial data, the implementation of services aimed at rendering the spatial data more accessible and interoperable and by dealing with obstacles to the use of the spatial data. INSPIRE will pave the road for a progressive harmonisation of spatial data in the Member States."

6.7 Implementation Plan for the Global Observing Systems for Climate (GCOS-IP)

Second Report on the Adequacy of the Global Climate Observing Systems for Climate (2AR, see http://www.wmo.int/web/gcos/gcoshome.html) published by the Global Climate Observing System (GCOS) in 2003 on request of the 187 parties who signed the United Nations Framework Convention for Climatic Change (UNFCCC).

Implementation Plan for the Global Observing Systems for Climate (GCOS IP, see http://www.wmo.int/web/gcos/gcoshome.html) following up the 2AR

6.8 Global Earth Observation System of Systems (GEOSS)

The Global Earth Observation System of Systems (GEOSS, http://earthobservations.org) is an international governmental initiative

Origin and Purpose of this Plan

The World Summit on Sustainable Development, Johannesburg 2002, highlighted the urgent need for coordinated observations relating to the state of the Earth. The First Earth Observation Summit was convened in Washington, DC in July 2003, attended by high- level officials of 33 countries and the European Commission and 21 international organisations involved in Earth observations 1. Governments adopted a Declaration signifying a political commitment to move toward development of a comprehensive, coordinated, and sustained Earth observation system. The Summit established the ad hoc intergovernmental Group on Earth observation (GEO), co- chaired by the European Commission, Japan, South Africa and the United States of America, and tasked it with the development of an initial 10-Year Implementation Plan by February 2005. GEO established five technical subgroups and a small secretariat. A series of subgroup meetings and a plenary meeting led to a Framework Document 2, negotiated at GEO-3 in Cape Town and adopted at the Second Earth Observation Summit in Tokyo in April 2004 by 47 nations and the European Commission, joined by 25 international organisations. The Framework defines the scope and intent of a Global Earth Observation System of Systems (GEOSS). A small task team was charged by the GEO with the drafting of an Implementation Plan, building on inputs from the subgroups and other sources.

The Implementation Plan establishes the operating principles, institutions and commitments relating to GEOSS. It is supported by a longer Reference Document (this document), which is consistent with the Implementation Plan, and provides the substantive detail necessary for implementation. The Implementation Plan was negotiated by the GEO in Ottawa in November 2004, and adopted at the Third Earth Observation Summit in Brussels, February 2005. The Reference Document was extensively reviewed by technical experts, nations and international organisations.

Scope of the GEOSS Implementation Plan

The Washington Summit Declaration establishes the objective "to monitor continuously the state of the Earth, to increase understanding of dynamic Earth processes, to enhance prediction of the Earth system, and to further implement our international environmental treaty obligations", and thus the need for "timely, quality, long-term, global information as a basis for sound decision making". The Framework Document adds that to move from principles to action, a "10-Year Implementation Plan for establishing the Global Earth Observation System of Systems (GEOSS)", which should be "comprehensive", "coordinated", and "sustained" is needed. The first 10-Year Implementation Plan of GEOSS defines a sequence of actions and responsibilities, commencing from the Third Earth Observation Summit in February 2005. GEOSS has an indefinite lifetime, subject to periodic review of its continued effectiveness.

A global…

In the GEOSS context, the word 'global' has two meanings. In the first sense, GEOSS aspires to be as inclusive as possible, embracing all nations and parts of the world and the organisations with Earth observation mandates. In the second sense, its priority focus is Earth system processes that operate at scales greater than the individual nation, for instance the global climate system. Phenomena that operate at lesser scales are the primary responsibility of local and national observing systems, but may be included in GEOSS if any of the following three conditions are met:

- They have global consequences in aggregate (e.g. desertification),
- They have significant global-scale causes (e.g. biodiversity loss);
- Their observation is enhanced by global systems (e.g. natural hazards)

…system of systems…

The components of GEOSS consist of existing and future Earth observation systems across the processing cycle from data collection to information production. Contributors maintain their respective responsibilities, ownership and mandates, but commit to making all or a portion of their observations available and easily accessible for collective use. GEOSS thus makes it possible to combine information from currently unconnected sources, in order to obtain a view that is sufficiently comprehensive to meet user needs.

…for Earth Observation

GEOSS will facilitate access to direct observations as well as products based on the collation, interpolation and processing of observations, and the services necessary for such a coordinated system, such as the maintenance of data description and exchange standards. The observations provided by GEOSS will originate entirely from contributing national, intergovernmental and non-governmental systems. They will include observations made outside the territory of any nation, for example of open oceans, Antarctica and from space. GEOSS will give priority to the development of observation-based products that are not currently available.

The content of GEOSS will be defined, from time to time, by its governance structures. Initially it covers the nine topic areas agreed by the second Earth observation Summit to be beneficial to many nations, and included in the Framework Document. GEOSS shall be built step-by-step through cooperation among existing observing and processing systems, while encouraging and accommodating new components as needs and capabilities develop. The plan includes the actions needed to build capacity, particularly in developing countries, that will permit the system to be useful to all participants.

The Case for a Global Earth Observation System of Systems

- Agreements to make systems interoperable and to share data
- Collective optimisation of the observation strategy
- Cooperative gap filling
- Commitments to observational adequacy and continuity

Societal Benefits, Requirements, and Earth Observation Systems

1. Reducing loss of life and property from natural and human induced disasters
2. Understanding environmental factors affecting human health and well being
3. Improving management of energy resources
4. Understanding, assessing, predicting, mitigating and adapting to climate variability and change
5. Improving water resource management through better understanding of the water cycle
6. Improving Weather Information, Forecasting and Warning
7. Improving the management and protection of terrestrial, coastal and marine ecosystems
8. Supporting sustainable agriculture and combating desertification
9. Understanding, monitoring and conserving biodiversity

7 Conclusion and Outlook

7.1 "Data" versus "model" integration

It should be mentioned that there are also endeavours to not only integrate environmental data but also to integrate model components developed by various specialists in a homogeneous environment, a prominent example being the Modular Modeling System (MMS) developed by Leavesley et al. (1996, 1998).

MMS is an integrated system of computer software that has been developed to provide the research and operational framework needed to support the development, testing and

evaluation of physical-process algorithms and to facilitate the integration of user-selected sets of algorithms into an operational model. MMS provides a common framework intended to focus multidisciplinary research and operational efforts. Scientists in a variety of disciplines can develop and test model components to investigate questions in their own areas of expertise as well as work cooperatively on multidisciplinary problems without each scientist having to develop the complete system model.
Continued advances in physical and biological sciences, GIS technology, computer technology, and data resources will expand the need for a dynamic set of tools to incorporate these advances in a wide range of interdisciplinary research and operational applications. MMS is being developed as a flexible framework in which to integrate these activities.

7.2 Further Development - Required long term vision

A user friendly intuitive environment is required hiding the considerable overhead generated by complex standards and XML definitions from the user and, on the other hand, guide (and force) the data set producer through the process of metadata input at data generation time. This also will involve the development of multilingual unified code lists and thesauri, which will ensure the information to be brought together and be comparable across all systems.

All this will certainly not happen in one single step but rather in an evolutionary process. However, it has to be the goal to minimise the number of iterations involved in this process, especially as the iterations become more and more complex as the system evolves and as resources are spent by an increasing number of organisations on harmonising their data holdings according to the latest developments. Thus, great care has to be taken to detect "dead ends" in the evolutionary process as early as possible and prevent dissipation of energy and subsequent frustration of participants. It will require a smart strategy to cope with the task of finding an efficient way ahead. The key to success will be in the interplay of a strongly focused supervising structure (top-down) and a number of organisations developing prototypes and putting them in test beds (bottom-up), however, without falling in love with their approaches, as to remain able to abolish their development, if something else proves to better serve the purpose.

In essence it will be of crucial importance to prioritise the development of a common denominator (not the least one!) for the archival and retrieval of data and metadata for general geophysical process measurements in the geosphere (i.e. applicable to data management in the atmospheric, oceanic and terrestrial domain). Whatever data current initiatives such as GEOSS and GCOS-IP succeed to collect, this data will not be sufficiently accessible if the data management issue is not improved at the same time (a crucially important, though not sufficient prerequisite). It will be the only way to control and consequently efficiently use and preserve the ever increasing amount of data.

We urgently need to carve out over-arching standards for data representation. Only a top-down approach will save in the mid to long term the global community a number of (costly and time consuming) subsequent iterations, that will be the consequence if domains continue to develop their individual standards in a bottom-up fashion.

It seems to be advisable to advocate for an international technical commission for this purpose that is not specialised to any specific domain (as e.g. WMO) but rather cares for the geosphere as a whole and is associated to a more neutral organisation such as a Technical Committee of ISO (recruiting of course specialists from all domains and specialised agencies), which as a side effect would also help the acceptance of the result by a wider community. Moreover, ISO has already published a metadata standard for geographic information, i.e. ISO 19115 in 2003 and almost finished a standardised XML-implementation of it with ISO 19139. This is the ideal starting point for the search for the common denominator, as *all* data on *any* geophysical variable share geographic information associated with it. *Everything* defined there (in a years-long process!) could be used right away as a start for the yet to find common denominator of geophysical variables, but needs to be extended. In fact, that is what is happening all over the world right now, especially in the geographic information community. Also the Future WMO Information System (FWIS), the Draft WMO Metadata Standard as well as JCOMM with its Oceans Information Technology Project (OIT) will build on it, however, care has to be taken that no unnecessary divergence takes place at a too early stage!

Summarising the preceding paragraphs, let's (at least in parallel to other activities) trigger the further development of all-embracing, generalised standards to the extend possible and only *afterwards* add domain and subdomain specific extensions if necessary. The geographic information community with its ISO 19100 series of standards has luckily already put a highly integrated "through-ball" which just waits to be picked up. It should not be recommended to pursuit individual approaches in a fragmented way which just will mean that we will have to come back to this point as time goes by.

Currently the path to the future seems not yet to be completely clear, that is why there are as many parallel approaches which are similar but will remain incompatible! That's exactly why it is recommend to initiate a (ISO TC-)process as argued for above. Such a process will probably not come to a final result within a typical project frame of 5-10 years (and will be an ongoing process anyway), but if it is not initiated now, we will be on a very similar point in 5-10 years. And of course, striving for the ambitious goal of a very general international standard is only one thing among many, because business has to go on and other actions will be required to support goals that can be achieved in shorter term, especially on the data collection side.

But what could be done to ensure shorter-term progress on some particular variables or data sets? Well, it is doing anything between the many approaches found today and the very ambitious and general process recommend above! Wherever one looks, all over the world and disciplines people and organisations can be observed struggling hard to improve access to their data, integrating some of the resources they are aware of and which they need from their current perspective for the domain(s) they currently consider. This certainly helps local/regional respectively sectoral communities to some extend for some time. And it is the only way to proceed instantaneously and it thus will not stop. Though this certainly can be called integration, it will be only until a group's perspective widens that they need to revise their scheme and harmonise it again with those developments that took place elsewhere in the area/by the communities that they now want to include. Probably this kind of comparatively uncoordinated bottom-up approach ("muddling around") will eventually also arrive at a universal standard, but in the authors view certainly at higher costs and with longer development time (alternatively it also can be imagined that all limited energy available is lost in friction). And in any case it will require at some point somebody who will substantiate, streamline and synchronise the process by setting a very general standard. So why not starting/initiating it now, taking the rare chance of the current awareness related to GEOSS and GCOS-IP?

7.3 Perspectives for GRDC operation

To be more concrete for a particular variable, River Discharge (and what GRDC does for shorter term progress):

The GRDC is one of such mechanisms currently in place, solely helping to integrate one of the many varíables of interest. It's progress is unsatisfying, as usually a terrible amount of manual/personal communication is required per data acquisition . Besides contacting National Hydrological Services (NHS) directly GRDC also scans their Web Sites and find more and more online presentations of country's individual data holdings, but guess what is the problem: They all follow quite different philosophies and provide different level of detail and change from time to time without prior notice. It is still very tedious to retrieve data, as it involves individual treatment of each source. GRDC is now close to finalising a near real time monitoring software for river discharge, however this will be only the basic infrastructure as the bulk of work will follow, i.e. writing interface routines for each online resource GRDC will trace down or alternatively convince the data providers to provide the data in our proposed format and with an information depth that GRDC defines (GRDC STANDARD!). This has not to be done only once, but GRDC will have to follow up all changes that will be made remotely without notifying us.

Moreover, other river discharge compilations exist statically or are maintained, but these groups of course prefer other interfaces. GRDC has encountered the situation that NHS reacted annoyed when GRDC contacted them, undertone: "Every other day someone contacts us to provide data in the format he requires, sorry we do not have the capacity". Also GRDC itself does not have the capacity to follow manually all possible channels, and this is why - besides problems on other levels, e.g. political - a data collection effort like that of GRDC will not be optimally up-to-date unless it can be substantially improved.

There are two ways to improve the situation:

- raising capacity by increasing manpower (the brute force approach) thus multiplying the manual/personal acquisition activity or
- introducing automation/machine-readability, which inevitably requires standardisation.

So should we have a ISO process starting for each single essential variable (EV) (because there is no way to convince all involved parties related to one EV to stick to a standard which is defined only by one party)? Only to eventually find that all the standards developed for individual EVs are not compatible with each other anyway and a climate researcher will still have to search for each of the EVs he needs to run his complex model in heterogeneous, though individual standardised systems? This is not recommendable, as it will multiply the efforts required.

On the other hand for the moment all what GRDC alone can do (and does) within it's limited capacity is trying to integrate in a bottom up fashion step by step. As long as there is no development of an overarching approach driven by something that is much more influential than GRDC, this is the only option. The current international trends, manifesting in activities as GEOSS and GCOS-IP provide a great if not unique chance for a huge step forward, rather than continuing in infinitesimal steps, always prone to the danger of stagnation due to high friction loss. This is about the question to either scratch with the chickens or fly with the eagles...

An impressive example for the increasing awareness of the importance of state-of-the-art data infrastructures by governments is the afore mentioned very recent proposal for a directive of the European Parliament and of the Council establishing an Infrastructure for SPatial InfoRmation in Europe (INSPIRE, http://inspire.jrc.it).

The rational behind this proposal is given in the introduction of the *Explanatory Memorandum* of EC (2004). Essentially for pretty much the same reasons that were stressed in this section, EC is starting by this directive (in the field of *Environmental Spatial Information*) exactly such a top-down approach that needs to be initiated by a yet to establish international coordinative body on standards for geophysical variables. This EC proposal could in fact become a template/starting point for discussing and developing a constituting document of the (yet to establish JCOMM-like) coordinative international body that was mentioned in the 2AR (GCOS, 2003c) as well as in the GCOS-IP. Just imagine to replace some terms used in the early sections of the *Explanatory Memorandum*, e.g.

- Good policy => Good science and consequently policy
- Spatial Information => Information on geophysical variables including its associated spatial information and metadata

- The Commission => "The (JCOMM-like) coordinative international body" (whatever will be its name and legal structure)

Let's not be too pusillanimous! This is a plea for the initiation of an international coordinating mechanism or body that guides in consultation with all major players involved (international organisations, governmental authorities, leading companies, etc.) a process of defining the fundaments for an global geosphere information system by developing standards and a data infrastructure for geophysical, biogeochemical and socioeconomic variables observed and predicted in the geosphere.

8 References

8.1 Literature

Denzer, R. , Schimak, G. and Humer, H. (1993): Integration in Environmental Information Systems, in D. Russell (ed.), International Symposium on Engineered Software Systems, Malvern, USA, May 1993, World Scientific, Singapore.

Denzer, R. and Güttler, R. (1995a): SIRIUS - Saarbrücken Information Retrieval and Interchange Utility Set, in International Symposium on Environmental Software Systems 1995 (ISESS 1995), Malvern, PA, USA, June 1995, Chapman & Hall.

Denzer, R. and Güttler, R. (1995b): An overview of integration problems in environmental information systems, J. Computing and Information 1(2), 1112-1120. (available at http://eig.htw-saarland.de/publications/english_pub/jci_95.pdf)

Denzer, R., Güttler, R., Houy, P. and TEMSIS-Consortium (2000): TEMSIS - a transnational system for public information and environmental decision support, J. Environmental Modelling and Software 15(3), 235-243.

Denzer, R., Schimak, G., Güttler, R., Houy, P. and TEMSIS-Consortium (1998): TEMSIS - a Transnational System for Public Information and Environmental Decision Support, HICSS-31, January 1998.

EC (2004): Proposal for a directive of the European Parliament and of the Council establishing an infrastructure for spatial information in the Community (INSPIRE), Brussels, 23 July 2004, COM(2004) 516 final, 2004/0175 (COD) (online available at http://inspire.jrc.it/proposal/COM_2004_0516_F_EN_ACTE.pdf)

ET-IDM (2001): Meeting of the Expert Team on Integrated Data Management, Geneva, 5 - 8 November 2001, Commission for Basic Systems, World Meteorological Organization, Geneva, Switzerland. (online available at http://www.wmo.int/web/www/WDM/reports/ET-IDM-2001.html; see also http://www.wmo.int/web/www/WDM/ET-IDM/documents.html)

ET-IDM (2002): Second Meeting of the Expert Team on Integrated Data Management, Shinfield Park, Reading, 13-16 May 2002, Commission for Basic Systems, World

Meteorological Organization, Geneva, Switzerland. (online available at http://www.wmo.int/web/www/WDM/reports/ET-IDM-2002.html; see also http://www.wmo.int/web/www/WDM/ET-IDM-2/documents.html)

ET-IDM (2003): Third Meeting of the Expert Team on Integrated Data Management, Geneva,15 - 18 December 2003, Commission for Basic Systems, World Meteorological Organization, Geneva, Switzerland. (online available at http://www.wmo.int/web/www/WDM/ET-IDM-3/documents.html)

ET-IDM (2004): Fourth Meeting of the Expert Team on Integrated Data Management, Geneva, 1-3 September 2004, Commission for Basic Systems, World Meteorological Organization, Geneva, Switzerland. (online available at http://www.wmo.int/web/www/WDM/ET-IDM-4/documents.html)

FWIS (1999): First Meeting of the Inter-Programme Task Team on the Future WMO Information System, Melbourne, 1-5 November 1999, Final Report. Commission For Basic Systems, World Weather Watch, World Meteorological Organization, Geneva, Switzerland. (online available at http://www.wmo.int/web/www/CBS-Reports/ISS-index.html#WDM or directly at http://www.wmo.int/web/www/reports/TT-FWIS-99.doc)

FWIS (2000): Second Meeting of the Inter-Programme Task Team on the Future WMO Information System, Monterey, California, 28 August - 1 September 2000, Final Report. Commission For Basic Systems, World Weather Watch, World Meteorological Organization, Geneva, Switzerland. (online available at http://www.wmo.int/web/www/CBS-Reports/ISS-index.html#WDM or directly at http://www.wmo.int/web/www/WDM/reports/FWIS-2000.doc and http://www.wmo.int/web/www/WDM/reports/FWIS-2000.html)

FWIS (2001): Third Meeting of the Inter-Programme Task Team on the Future WMO Information System, Langen, Germany, 25 - 29 June 2001, Final Report. Commission For Basic Systems, World Weather Watch, World Meteorological Organization, Geneva, Switzerland. (online available at http://www.wmo.int/web/www/CBS-Reports/ISS-index.html#WDM or directly at http://www.wmo.int/web/www/WDM/reports/FWIS-2001.doc and http://www.wmo.int/web/www/WDM/reports/FWIS-2001.html)

FWIS (2002): Fourth Meeting of the Inter-Programme Task Team on the Future WMO Information System, Johannesburg, South Africa, 23-27 September 2002, Final Report. Commission For Basic Systems, World Weather Watch, World Meteorological Organization, Geneva, Switzerland. (online available directly at http://www.wmo.int/web/www/WDM/reports/FWIS-2002.doc and http://www.wmo.int/web/www/WDM/reports/FWIS-2002.html)

FWIS (2003): Fifth Meeting of the Inter-Programme Task Team on the Future WMO Information System, Kuala Lumpur, 20 - 24 October 2003, Final Report. Commission For Basic Systems, World Weather Watch, World Meteorological Organization, Geneva, Switzerland. (online available at http://www.wmo.int/web/www/FWIS/FWIS-2003-final.pdf ; see also http://www.wmo.int/web/www/FWIS/documents.htm)

GCOS (2000) Establishment of a Global Hydrological Observation Network for Climate. Report of the GCOS/GTOS/HWRP Expert Meeting, Geisenheim, Germany, June 26-30, 2000. J. Cihlar, W. Grabs, and J. Landwehr (Editors), (GCOS-63; GTOS-26)

(WMO/TD-No. 1047). (Report online available at http://www.fao.org/gtos/doc/pub26.pdf)

GCOS (2002): Report of the GCOS/GTOS/HWRP Expert Meeting on the Implementation of a Global Terrestrial Network - Hydrology (GTN-H), Koblenz, Germany, June 21-22, 2001. W. Grabs and A.R. Thomas (Editors), (GCOS-71; GTOS-29) (WMO/TD-No. 1099). (Report online available at http://www.wmo.int/web/gcos/Publications/gcos-71.pdf)

GCOS (2003a): Global Terrestrial Network-Hydrology (GTN-H). Report of the GTN-H Coordination Panel Meeting, Toronto, Canada, 21-22 November 2002. K.D. Harvey, W. Grabs and A.R. Thomas (Editors), (GCOS-83; GTOS-33) (WMO/TD-No. 1155). (Report online available at http://www.wmo.int/web/gcos/Publications/gcos-83.pdf)

GCOS (2003b): Report of the GCOS/GTOS/HWRP Expert Meeting on Hydrological Data for Global Studies, Toronto, Canada, 18-20 November 2002. K.D. Harvey and W. Grabs (Editors), (GCOS-84; GTOS-32) (WMO/TD-No. 1156). (Report online available at http://www.wmo.int/web/gcos/Publications/gcos-84.pdf)

GCOS (2003c): Second Report on the Adequacy of the Global Climate Observing Systems. Developed on request of the United Nations Framework Convention for Climatic Change/ Subsidiary Body for Scientific and Technological Advice (UNFCCC/SBSTA). Available from the GCOS-homepage at http://www.wmo.int/web/gcos/gcoshome.html)

GRDC (2003): Development of an operational internet-based near real time monitoring tool for global river discharge data. A contribution to the Global Terrestrial Network for Hydrology (GTN-H). T. Maurer, GRDC Report No. 30, Koblenz, Germany. (Report online available at ftp://ftp.bafg.de/pub/REFERATE/GRDC/report30.pdf)

Hankin, S., Bahner, L., Bernard, L., Bogden, P., Cohen, R., Cornillon, P., Dantzler, L., Glenn, S., Grassle, F., Legler, D., Nowlin, W., Orsi, T., Sherman, B., Spaulding, M., Starke, S. (2002): A Data and Communications Infrastructure for the U.S. Integrated Sustained Ocean Observing System. Ocean.US 2002 working paper. 2300 Clarendon Blvd. Suite 1350 Arlington, VA 22201-3667 (online available at http://www.ocean.us/documents/docs/BAKDOC8_Data-Information-ManagementPaper.doc)

Harmancioglu, N. B., Alpaslan, M. N., Ozkul, S. D. (1997): Conclusions and recommendations. In: Harmancioglu, N. B., Alpaslan, M. N., Ozkul, S. D., Singh, V.P. (eds.): "Integrated Approach to Environmental Data Management Systems", Proc. NATO Advanced Research Workshop, 16 - 20 September 1996, Nato ASI Series, 2. Environment, Volume 31, Part 9, pp. 423-436, Kluwer Academic Publishers

ISO (2003): ISO 19115: Geographic information – Metadata

ISO/TC211/WG7 (2004): Draft Technical Specification 19139 Geographic information – Metadata - XML schema implementation, Due date: 2004-09-30 http://www.isotc211.org/protdoc/211n1663

ISO/IEC (1995-2004) ISO 11179: Information Technology – Metadata (MDR) -- Part 1: Framework, Part 2: Classification, Part 3: Registry metamodel and basic attributes; Part 4: Formulation of data definitions, Part 5: Naming and identification principles, Part 6: Registration (see e.g. http://www.cio-dpi.gc.ca/im-gi/meta/mdregistry/mdregistr_e.asp)

ISO (1998): ISO 23950: Information and documentation – Information retrieval (Z39.50) – Application service definition and protocol specification http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=27446 (Protocol for Information Search and Retrieval)

JCOMM (2002): An Ocean Information Technology Project. Document DMCG-I/28 (20 May 2002) to the First Session of the Data Management Coordination Group (DMCG) of the Joint WMO-IOC Technical Commission for Oceanography and Marine Meteorology (JCOMM) of the Intergovernmental Oceanographic Commission (of UNESCO) (IOC) and the World Meteorological Organization (WMO), Paris, 22-25 May 2002 (online available at http://ioc.unesco.org/oit/files/DMCG1_doc28_oit.pdf)

Jeffery, Keith G. (2001): Metadata: The Future of Information Systems. Paper presented by the Director, Information Technology, CLRC-Rutherford Appleton Laboratory, Chilton, Didcot, Oxfordshire, UK to the Meeting of the Expert Team on Integrated Data Management, Geneva, 5 - 8 November 2001, Commission for Basic Systems, World Meteorological Organization, Geneva, Switzerland. (online available via http://www.wmo.int/web/www/WDM/ET-IDM/documents.html)

KOGIS (2001): Entwicklung eines ISO/DIS 19115 kompatiblen Metadatenmodells für die Schweiz, Version 1.2d. Koordination der Geoinformation und geografischen Informationssysteme. (Development of a ISO/DIS 19115 compatible metadatamodel for Switzerland, Version 1.2d. Coordination of geographic information and geographic information systems) KOGIS, c/o Bundesamt für Landestopographie, Seftigenstrasse 264, Postfach, CH-3084 Wabern Tel. +41 31 963 21 11 / Fax +41 31 963 24 59, Web http://www.kogis.ch .

Kresse, W., Fadaie, Kian (2004): ISO Standards for Geographic Information, Springer, ISBN 3-540-20130-0

New Zealand Government (2004): New Zealand Government Geospatial Metadata Standard DRAFT v.1.2, Part 1 - Profile Definition and Part 2 – Profile Guidelines, prepared by New Zealands Geospatial Metadata Project Team in June 2004, current versions are available at http://www.linz.govt.nz/rcs/linz/pub/web/root/core/Topography/ProjectsAndProgrammes/geospatialmetadata

Maurer, Th. (2003a): Challenges in transboundary and transdisciplinary environmental data integration in a highly heterogeneous and rapidly changing world - A view from the perspective of the Global Runoff Data Centre. In: Harmancioglu, Nilgun B., Ozkul, Sevinc D., Fistikoglu, Okan, Geerders, Paul (eds.): "Integrated Technologies for Environmental Monitoring and Information Production" , Proc. NATO Advanced Research Workshop, 10 - 14 September 2001, Marmaris, Turkey. Nato Science Series IV Volume 23, pp. 25-38, Kluwer Academic Publishers (ISBN HB: 1-4020-1398-1, PB: 1-4020-1399-X). (online available at http://grdc.bafg.de/?2535)

Maurer, Th. (2003b): Intergovernmental arrangements and problems of data sharing. Monitoring Tailor-Made IV Conference, Information to support sustainable water management: From local to global levels, St. Michielsgestel, The Netherlands 15-18 September 2003. (online available a). (online available at http://grdc.bafg.de/?3997)

Maurer, Th. (2004): Transboundary and transdisciplinary environmental data and information integration - an essential prerequisite to sustainably manage the Earth System.

INDUSTRY IDS-Water Europe 2004 Online Conference (http://www.idswater.com), 10 May - 28 May 2004, 14pp (online available also at http://grdc.bafg.de/?6413)

Smith, N. (2002): Ocean Information Technology - New opportunities for marine data management. Presentation at IOGOOS-I in October 2002 (online available at http://ioc.unesco.org/oit/files/Smith_IOGOOS_OIT_files/frame.htm; see also JCOMM, 2002 and Hankin et al., 2002)

ST-OIT (2002): Report of the First Session of the Steering Team of the Oceans Information Technology Pilot Project, Brussels, Belgium, 29 November 2002. Hosted by the Research and Development Department, Scientific Technical and Cultural Affairs (OSTC). IOC, UNESCO (online available at http://ioc.unesco.org/oit/files/INF1178.pdf)

WWDR (2003): World Water Development Report of the Secretariat of the World Water Assessment Programme (WWAP) at UNESCO, Paris

WMO (2001a): Commission for Hydrology, Eleventh Session, Abuja, Nigeria, 6-16 November 2000, abridged final report with resolutions and recommendations, Secretariat of the World Meteorological Organization, Geneva, Switzerland. (WMO-No. 921).

WMO (2001b): Exchanging hydrological data and information, WMO policy and practice, World Meteorological Organization, Geneva, Switzerland. (WMO brochure No. 925).

WMO (2001c): Exchange of Hydrological Data and Products, WMO Technical Reports in Hydrology and Water Resources No.74, P. Mosley, World Meteorological Organization, Geneva, Switzerland. (WMO/TD-No. 1097). (Report online available at http://www.wmo.int/web/homs/documents/TD74.pdf)

WMO (2004a): The Future WMO Information System, WMO-FWIS-DOC-SSA-5239-03.pdf by Prof. Geerd-R. **Hoffmann**, Deutscher Wetterdienst (German Weather Service), Offenbach, Germany, Email: geerd-ruediger.hoffmann@dwd.de, February 2004. See also: FWIS (1999, 2000, 2001, 2002, 2003).

WMO (2004b): WMO Core Metadata Standard (v0-2): http://www.wmo.int/web/www/WDM/Metadata/documents.html

8.2 Acronyms and Links

DCMI, Dublin Core Metadata Initiative: Homepage, http://dublincore.org

FGDC, Federal Geographic Data Committee: Homepage, http://www.fgdc.gov

GCOS-IP, Implementation Plan of the Global Climate Observing System http://www.wmo.int/web/gcos/gcoshome.html

GEOSS, Global Earth Observation System of Systems of GEO, http://earthobservations.org

HWRP, Hydrology and Water Resources Programme of the World Meteorological Organisation: Homepage, http://www.wmo.int/web/homs/hwrphome.html

ISO/JTC1 http://www.iso.org/iso/en/stdsdevelopment/tc/tclist/TechnicalCommitteeDetailPage.TechnicalCommitteeDetail?COMMID=1

ISO/JTC1/SC32
http://www.iso.org/iso/en/stdsdevelopment/tc/tclist/TechnicalCommitteeDetailPage.TechnicalCommitteeDetail?COMMID=160

ISO/TC 211 Geomatics: Homepage, http://www.isotc211.org

OGC, Open GIS Consortium: Homepage, http://www.opengis.org

WMO, World Meteorological Organization , http://www.wmo.int

WWAP, World Water Assessment Programme: Homepage, http://www.unesco.org/water/wwap

XML, Extensible Markup Language: Homepage, http://www.xml.org

Annex 1

Example of application of the Draft WMO Core Metadata Standard to the GRDC database

```
  <?xml version="1.0" encoding="UTF-8" ?>
  <!--  XML file generated by Thomas Maurer GRDC Koblenz Germany   -->
  <?xml-stylesheet href="XML.xsl" type="text/xsl"?>
- <metaData xmlns="http://www.wmo.ch/www/metadata"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://www.wmo.ch/www/metadata
    http://www.wmo.ch/www/metadata/WMO19115_metadata_v0_1.xsd">
    <metadataLanguage>en</metadataLanguage>
    <metadataCharacterSet>utf8</metadataCharacterSet>
  - <metadataContact>
      <organisationName>Global Runoff Data Centre (GRDC) in the Federal Institute of
        Hydrology (BfG)</organisationName>
      <role>The digital world-wide depository of discharge data and associated metadata in
        support of Global Change Research and Integrated Water Resources
        Management</role>
      <individualName>Thomas Maurer</individualName>
      <positionName>Head</positionName>
    - <contactInfo>
      - <phone>
          <voice>+49 261 1306 5224</voice>
          <facsimile>+49 261 1306 5280</facsimile>
        </phone>
      - <address>
          <deliveryPoint>Am Mainzer Tor 1</deliveryPoint>
          <city>Koblenz</city>
          <administrativeArea />
          <postalCode>56068</postalCode>
          <country>Germany</country>
          <electronicMailAddress>grdc@bafg.de</electronicMailAddress>
        </address>
        <onlineAddress>http://grdc.bafg.de</onlineAddress>
      </contactInfo>
    </metadataContact>
    <metadataDateStamp>2004-04-29</metadataDateStamp>
    <metadataStandardName>WMO19115_metadata_v0_1</metadataStandardName>
    <metadataStandardVersion>0.1</metadataStandardVersion>
  - <referenceSystemInfo>
      <referenceSystemIdentifier />
    - <referenceAuthority>
        <organisationName />
        <role />
        <individualName />
        <positionName />
      - <contactInfo>
        - <phone>
            <voice />
            <facsimile />
          </phone>
        - <address>
            <deliveryPoint />
            <city />
            <administrativeArea />
            <postalCode />
            <country />
            <electronicMailAddress />
          </address>
          <onlineAddress />
        </contactInfo>
      </referenceAuthority>
      <referenceDescription>Global set of discharge station locations given by latitude
        longitude pairs</referenceDescription>
    </referenceSystemInfo>
```

```
- <identificationInfo>
  - <citation>
      <resourceTitle>Global Runoff Data Base</resourceTitle>
      <resourceIdentifier>GRDB</resourceIdentifier>
      <resourceIdentifierType />
    </citation>
    <abstract>Global database of time series from 6700 discharge stations (5000 featuring
      daily data, 6700 featuring monthly data)</abstract>
  - <pointOfContact>
      <organisationName>Global Runoff Data Centre (GRDC) in the Federal Institute of
        Hydrology (BfG)</organisationName>
      <role>The digital world-wide depository of discharge data and associated metadata
        in support of Global Change Research and Integrated Water Resources
        Management</role>
      <individualName>Thomas Maurer</individualName>
      <positionName>Head</positionName>
    - <contactInfo>
      - <phone>
          <voice>+49 261 1306 5224</voice>
          <facsimile>+49 261 1306 5280</facsimile>
        </phone>
      - <address>
          <deliveryPoint>Am Mainzer Tor 1</deliveryPoint>
          <city>Koblenz</city>
          <administrativeArea />
          <postalCode>56068</postalCode>
          <country>Germany</country>
          <electronicMailAddress>grdc@bafg.de</electronicMailAddress>
        </address>
        <onlineAddress>http://grdc.bafg.de</onlineAddress>
      </contactInfo>
    </pointOfContact>
  - <resourceMaintenance>
      <maintenanceAndUpdateFrequency>Irregular, depending on data provider between 1
        and 20 years</maintenanceAndUpdateFrequency>
      <maintenanceNote>Information will be disseminated 1 week past data
        receipt</maintenanceNote>
    </resourceMaintenance>
  - <resourceConstraints>
      <useLimitation>A user declaration has to be signed. Restrictions as determined by
        the GRDC policy guidelines for the dissemination of data and costing of services.
        In brief: (1) Free and unrestricted (but identified) access to all hydrological data
        and products (2) Data are free of charge (only a fee for services and reproduction
        may be charged) (3) No commercial use of the data (4) Ownership of the data
        and responsibility for errors lie with the data providers (5) No redistribution of
        the data by the user (6) No distribution of the entire database (or of substantial
        parts)</useLimitation>
    - <legalConstraints>
        <accessConstraints>Only by formal data request to GRDC</accessConstraints>
        <useConstraints>As signed in the User Declaration</useConstraints>
        <otherConstraints />
      </legalConstraints>
    </resourceConstraints>
  - <resourceConstraints>
      <useLimitation />
    - <securityConstraints>
        <classification />
        <userNote />
        <classificationSystem />
        <handlingDescription />
      </securityConstraints>
    </resourceConstraints>
```

```
  <spatialRepresentationType>Irregular distributed points and associated time
    series</spatialRepresentationType>
  <spatialResolution />
  <dataLanguage>Data held in Oracle database, export in ASCII files</dataLanguage>
  <dataCharacterSet>utf8</dataCharacterSet>
  <topicCategory>Global, Hydrology, Water, Climatology</topicCategory>
  <descriptiveKeywords>River discharge, runoff, streamflow, global coverage, daily and
    monthly time series</descriptiveKeywords>
- <referenceDate>
    <date />
    <dateType>referenceDate</dateType>
  </referenceDate>
- <referenceDate>
    <period>1806-01-01T00:00:00 until 2003-12-31T00:00:00</period>
    <dateType>referenceDate</dateType>
  </referenceDate>
- <dataExtent>
    <description>Globally irregular distributed points</description>
  - <geographicElement>
    - <geographicBoundingBox>
        <westBoundLongitude>-180</westBoundLongitude>
        <eastBoundLongitude>180</eastBoundLongitude>
        <southBoundLatitude>-90</southBoundLatitude>
        <northBoundLatitude>90</northBoundLatitude>
      </geographicBoundingBox>
    </geographicElement>
  - <geographicElement>
    - <polygon>
      - <point>
          <latitude>180</latitude>
          <longitude>90</longitude>
        </point>
      - <point>
          <latitude>-180</latitude>
          <longitude>90</longitude>
        </point>
      - <point>
          <latitude>-180</latitude>
          <longitude>-90</longitude>
        </point>
      - <point>
          <latitude>180</latitude>
          <longitude>-90</longitude>
        </point>
      - <point>
          <latitude>180</latitude>
          <longitude>90</longitude>
        </point>
      </polygon>
    </geographicElement>
  - <geographicElement>
      <geographicIdentifier>World</geographicIdentifier>
    </geographicElement>
  </dataExtent>
- <dataExtent>
    <description>Validity Time Range</description>
  - <temporalElement>
      <beginDateTime>earliest record 1806</beginDateTime>
      <endDateTime>Latest record 2003</endDateTime>
      <dataFrequency>daily, monthly</dataFrequency>
    </temporalElement>
```

```
    </dataExtent>
  - <dataExtent>
      <description>Corresponding to the altitude of the individual data
        points</description>
    - <verticalElement>
        <minimumValue>1</minimumValue>
        <maximumValue>4000</maximumValue>
        <unitOfMeasure>m.a.s.l.</unitOfMeasure>
      </verticalElement>
    </dataExtent>
  </identificationInfo>
- <distributionInfo>
  - <distributionFormat>
      <formatName>GRDC Export Format</formatName>
      <formatVersion>Version 2</formatVersion>
    </distributionFormat>
  - <transferOptions>
      <transferSize>Variable, few kB to several MB</transferSize>
      <onLineSource>Data itself not online, however some products (statistics,
        integrations)</onLineSource>
      <WMO_Source>GTN-H, GTN-R</WMO_Source>
      <offlineMedium>File by email, on CD or diskette</offlineMedium>
    </transferOptions>
  </distributionInfo>
- <dataQualityInfo>
    <lineageStatement>Coordinates are of varying accuracy, typically 1 km, depending on
      data provider, in general NHS of WMO member states</lineageStatement>
    <dataProcessInfo>Data is distributed as received from data provider (in general NHS of
      WMO member states). Only crude plausibility checks are performed by
      GRDC</dataProcessInfo>
    <dataSourceInfo>GRDC data providers are in general the NHS of WMO member
      states</dataSourceInfo>
  </dataQualityInfo>
</metaData>
```

Annex 2

Proposed Version 0-2 of the Draft WMO Core Profile (Sep 2004)

Draft WMO Core Profile of the ISO Metadata Standard
Proposal for Version 0.2, August 2004

Notes: The following table provides an overview of the WMO Community Core Metadata Profile suitable for use by decision makers and users - NOT implementers. To implement this standard the ISO DIS 19115 document, which describes the complete ISO standard, must be consulted.

This standard provides a general definition for directory searches and exchange that should be applicable to a wide variety of WMO datasets. It does not specify how these metadata should be archived or presented to users. It also does not specify any particular implementation and could be implemented as a database, a flat file, or any other suitable mechanism.

Of the core elements listed, those in **bold** are required, with all others being optional.

It must be remembered that this list defines a minimum set of information to describe data for WMO exchange and is not exhaustive. To fully meet the requirements of WMO Programmes for metadata, application of far more comprehensive standards would be required. The development of these comprehensive standards should be pursued by the individual programmes.

Changes to the profile developed by ET-IDM 2 are marked in green for those proposed by ET-IDM-3 and in yellow for those proposed by ET-IDM-4
.

Generic Name	ISO Field/Class Name and Reference Lines	Definition
	MD_Metadata (1)	
Metadata ID	fileIdentifier (2)	Unique identifier for this metadata item
Metadata language	**language** (3)	Language of this metadata item
Metadata char. set	**characterSet** (4)	Character set used for this metadata item (Default of ISO 10646-1 but any standard character set can be used)
Metadata contact	**contact** (8) CI_ResponsibleParty (see 374 below)	Party responsible for this metadata item
Metadata date	**dateStamp** (9)	Date that this metadata item was created
Metadata name	metadataStandardName (10)	Name of the metadata standard (including profile name) used
Metadata Version	metadataStandardVersion (11)	Version (profile) of the metadata standard used
	referenceSystemInfo (13)	Description of the data temporal and spatial reference system
	MD_ReferenceSystem (186)	Information about the reference systems used (temporal, coordinate and geographic)
Data Reference System	ReferenceSystemIdentifier (187)	Name of reference system

	ReferenceAuthority (206)	Person or party responsible for maintenance of the reference system
	CI_ResponsibleParty (374) (see below)	
	ReferenceDescription (207)	Description of the Reference System
	IdentificationInfo(15)	
Data information	MD_DataIndentification (see 36 below)	Basic information about the data
	distributionInfo(17)	
Data distribution Information	MD_Distribution (see 270 below)	Information about the data distribution and availability
	dataQualityInfo(18)	
Data Lineage or Quality	LI_Lineage (see 82 below)	Information about the data lineage or quality
	MD_DataIdentification (36)	Basic information required to uniquely identify a dataset
	Citation (24)	
Title	**title** (360)	Name of the dataset
Reference Date	**referenceDate** (362)	Reference date for the dataset. This and other dates referring to the maintenance of the data set use the Julian Calendar. Dates that describe the data themselves may use other calendars, but if so this must be specified in MD_ReferenceSystem.
	CI_Date (see 393 below)	
Identifier	**identifier** (365)	Unique identifier for dataset
	identifierType (366)	Form of the unique identifier (if standardized)
Abstract	**abstract** (25)	Brief narrative summary of the contents of the dataset
Dataset Contact	**pointOfContact** (29)	Identification of, and means of communication with, person(s) and organizations(s) associated with the dataset
	CI_ResponsibleParty (see 374 below)	
	resourceMaintenance(30)	
Update frequency	maintenanceAnd UpdateFrequency(143)	Frequency with which changes are made to the dataset after the intial dataset is created. This maintenance update frequency should either be the WMO code or of the form {Date type, Time type or DateTime Type}.The time elements can be repeated as often as needed to describe the data.
	maintenanceNote(148)	Information regarding specific requirements for maintaining the dataset
Access Rights or Restrictions	resourceConstraints (35)	
	MD_Constraints (67) (see below)	Restrictions on the access and use of the resource or metadata
Spatial Resolution	spatialResolution (38)	Spatial density of the data in the dataset (e.g. grid spacing)

	spatialRepresentationType (37)	Method used to spatially represent data in the dataset [Code list: B.5.26]
Language	language(39)	Language(s) used in the dataset, if applicable
Character set	characterSet (40)	Character set used in the dataset, if applicable
Topic Category	**topicCategory** (41)	Discipline covered by this dataset [ISO code list B.5.27] - Note this field is of limited use for WMO purposes but is a required field within the ISO standard and is included to ensure conformity.
Keywords	**descriptiveKeywords** (33) keywordReferenceSource(new)	List of predefined and other keywords used to describe the dataset. Keywords should be taken from a standard thesaurus (the URI for this thesaurus should be given – this, for example, would facilitate searching in different languages), or other defined list but free form keywords are permitted as well.
Date or period	**CI_Date** (393)	
	date or period (394)	Reference date or period for the dataset
	dateType (395)	Type of date [code list: creation, publication or revision date]
Responsible Party	CI_ResponsibleParty (374)	
Organization	**organisationName** (376)	Name of the responsible organization
Org. role	**role**(379)	Function performed by the responsible party [code list: resourceProvider, custodian, owner, user, distributer, originator, etc¨]
Individual name	individualName (375)	Name of the responsible person
Position	positionName (377)	Position of the responsible person
	contactInfo (378)	
	CI_Contact (387)	NOTE: Either a phone number or address is required
	phone (388)	
Phone number	voice (408)	Telephone by which individuals can speak to the responsible party
Fax number	facsimile (409)	Telephone number of a fax machine for the responsible party
Address	address(389)	
	deliveryPoint (381)	Address line for the location
	city (382)	City of the location
	postalCode (384)	Postal code
	country (385)	Country
E-mail address	electronicMailAddress (386)	Electronic mail address of the responsible party
Web Site	onLineAddress (390)	URL of organization
	EX_Extent (334)	Information about spatial, vertical, and temporal extent of the dataset
	Description (335)	Spatial and temporal extent for the dataset (in text)
Vertical Extent	Ex_VerticalExtent (354)	Vertical domain of the dataset (Note: There is potential ambiguity about vertical

		extent, particularly in oceanography. This can be resolved by the unitOfMeasure.)
	minimumValue (355)	Lowest vertical extent contained in the dataset
	maximumValue (356)	Highest vertical extent contained in the dataset
	unitOfMeasure (357)	Vertical units used for vertical extent information (E.g.: metres, feet, hectopascals) This must include the sign convention for height (whether values increase upwards or downwards).
	verticalDatum (358)	Information about the origin from which the maximum and minimum elevation values are measured (see ISO 1911).
	signConvention (new)	Whether the vertical co-ordinate increases or decreases upwards.
Geographic Extent	EX_GeographicExtent (339)	WMO metadata must contain the “bounding box” where relevant – even if global. However, either or both of a geographical name and/or a bounding polygon and/or an irregular point set *should* be used as well.
	HorizontalCoordinateType (new)	If the horizontal co-ordinate type and datum are not specified then the standard is assumed to be WGS84. This refers to the metadata – MD_ReferenceSystem specifies these for the data.
Irregular point set	irregularPointSet (new)	This class is a list of co-ordinates defining positions of data in the datset
Geographic name	EX_GeographicDescription (348)	Description of the geographic area using identifiers (names)
	geographicIdentifier (349)	Identifier used to represent a geographic area or location. While it is preferable to use names from a well-known Gazetteer (this should be referred to in the identifier), it is acceptable to use names that are not in a Gazetteer.
Bounding box	EX_GeographicBoundingBox	NOTE This is only an approximate reference so specifying the co-ordinate system is unnecessary. Using latitude and longitude, for any box surrounding a Pole, the limits are +/-90 and the southern (northern) most latitude, and the longitude extent must be +/-180. Bounding box may not be effective when used to search for data that cross the international date line or a pole.
	westBoundLongitude (344)	Western-most limit of the dataset, longitude in decimal degrees (positive east)
	eastBoundLongitude (345)	Eastern-most limit of the dataset, longitude in decimal degrees (positive east)
	southBoundLatitude (346)	Southern-most limit of the dataset, latitude in decimal degrees (positive north)
	northBoundLatitude (347)	Northern-most, limit of the dataset, latitude in decimal degrees (positive north)
		Sets of points defining a bounding polygon.
Bounding polygon	EX_BoundingPolygon (341)	The polygon is defined as a set of co-ordinate pairs with the last pair the same as the first. When the points in the polygon are traversed, the interior is to the left of the direction of travel. If the region has “holes”, multiple polygons may be used. The points of the outer polygon will be traversed anti-clockwise, and those of inner polygons will be traversed clockwise.
	polygon (342)	
Temporal Extent	EX_TemporalExtent (350)	NOTE: Each of the Extent fields below is **required if applicable**
	Extent (351)	

	referenceDateTime (new)	Creation or issuing time of data.
	beginDateTime (new)	Beginning date of the data in the dataset
	endDateTime (new)	Ending date of data in the dataset. For datasets that are still being added to the end date should indicate the expected end date of the series (if known), but "continuing" is acceptable.
	dataFrequency (new)	Observing frequency of the data in the dataset [code: WMO_DataFrequencyCode] The data update frequency should either be the WMO code or of the form {Date type, Time type or DateTime Type}. The time elements can be repeated as often as needed to describe the data.
	CalendarType (new)	Type of calendar used by the data. Default is Julian calendar.
Access Rights or Restrictions	MD_Constraints (67)	Restrictions on the access and use of the dataset (Could specify WMO Additional Data as free text) Note: At present the WMO Core Metadata will not contain the ISO parameters that describe access constraints to the metadata, but implementers should be aware that the ISO parameters exist and might be required in later versions of the WMO Core. Any metadata "published" through a system developed for the WMO Core is therefore likely to be disclosed regardless of privacy markings on the metadata.
	useLimitation (68)	
	MD_LegalConstraints (69)	Restrictions and legal prerequisites for accessing and using the dataset
	accessConstraints (70)	Any special restrictions or limitations on obtaining the dataset
	useConstraints (71)	Any special restrictions or limitations or warnings on using the dataset
	otherConstraints (72)	Other restrictions and legal prerequisites for accessing and using the dataset
	MD_SecurityConstraints (73)	Handling restrictions imposed on the dataset for security reasons
	classification (74)	Name of the handling restrictions on the dataset
	userNote (75)	Explanation of the application of the legal constraints or other restrictions and legal prerequisites for obtaining and using the dataset
	classificationSystem (76)	Name of the classification system
	handlingDescription (78)	Additional information about the restrictions on handling the dataset
	MD_Distribution (270)	Information about the distributor of and options for obtaining the dataset
	distributionFormat (271)	Provides a description of the format of the data to be distributed
Format name	**name** (285)	Name of the data transfer format(s). This is an unrestricted string that allows multiple formats to be included. It could also include reference to documentation of compression methods.
Format version	version (286)	Version of the format (date, number, etc.)
	TransferOptions (273)	NOTE: At least either on-line source or off-line media is required
On-line source	OnLine (277)	Information about online sources from which the dataset can be obtained

Off-line media	linkage (397) WMO_Source (new) Offline (278) mediumName (292)	Location (address) for on-line access using a Uniform Resource Locator WMO centre identifier Information about offline media on which the dataset can be obtained Name of the medium on which the dataset can be received [code list: ISO B.5.20]
FeatureCatalogue	MD_ContentInformation (232) MD_FeatureCatalogueDescription (233) complianceCode (234) language (235) includeWithDataset (234) featureTypes (237) featureCoverage (new) featureCatalogueCitation (238)	Describes the content of the dataset in more detail than the keywords. The ISO standard provides both "feature catalogues" and "Coverage" to describe the attributes of the data held in the dataset. This WMO Core Metadata chooses to use "feature" to describe all aspects of these attributes, including those relating to grids of data. Value 1 if feature catalogue is compliant with ISO19110. Default is 0 (not compliant) Language(s) used in the Catalogue **Required if feature Catalogue is used.** Value 1 in feature catalogue is included in dataset, 0 if not. Subset of feature types from the cited feature catalogue occurring in the dataset. Note: the physical variables described by the data are attributes of a feature (which could be an observed profile or a field of data, for example). Information about grids and other qualifiers for features (such as which pressure level a temperature field refers to). **Required if featureCatalogue is used.** Bibliographic reference to the feature catalogue(s) used.
Processing Level	**LI_Lineage** (82) statement(83) processStep(84) source(85)	Information about the level of processing applied to the dataset. This field should be used to indicate whether the data are observations, analyses (re-analyses), forecast (based on initial states including observations), simulations or other sources of data. Could also be used to include the platform/mission in the source of data (eg Ship, aircraft, satellite, satellite id). May need to use pairs of [source, processing step] to provide additional information. May contain references (eg URI) to external information on the processing and source. Information about the events or source data used in constructing the dataset Information about an event in the creation process for the dataset Information about the source data used in creating the dataset
Reference System	MD_ReferenceSystem (186) referenceSystemIdentifier (187) authority (206)	Information about the reference systems used (temporal, coordinate and geographic) Name of reference system

	CI_ResponsibleParty (374) (see above)	Person or party responsible for maintenance of the reference system namespace
	code(207)	Alphanumeric value identifying an instance in the namespace

Extensions to ISO Code Lists

B.5.26 MD_SpatialRepresentationTypeCode <<CodeList>>

	Name	Domain code	Definition
1.	MD_SpatialRepresentationTypeCode	SpatRepTypCd	method used to represent geographic information in the dataset
2.	vector	001	vector data is used to represent geographic data
3.	grid	002	grid data is used to represent geographic data
4.	textTable	003	textual or tabular data is used to represent geographic data
5.	tin	004	triangulated irregular network
6.	stereoModel	005	three-dimensional view formed by the intersecting homologous rays of an overlapping pair of images
7.	video	006	scene from a video recording
Additional entries			
8.	**irregularPoints**	**007**	**Irregularly-spaced points, such as meteorological stations**

NEW:
WMO_DataFrequencyCode <<CodeList>>

	Name	Domain code	Definition
1.	WMO_DataFrequencyCode	DataFreqCd	Temporal sampling frequency of the data within the dataset
2.	Continuous	001	More than once per minute
3.	1minute	002	
4.	5minute	003	
5.	10minute	004	
6.	15minute	005	
7.	30minute	006	
8.	Hourly	007	
9.	3hourly	008	
10.	6hourly	009	
11.	8hourly	010	
12.	12hourly	011	
13.	Daily	012	
14.	Weekly	013	
15.	10day	014	
16.	Fortnightly	015	
17.	Monthly	016	
18.	3monthly	017	

}

	Name	Domain code	Definition
19.	6monthly	018	
20.	Annual	019	
21.	decade	020	Decade or longer

}

Keywords for Describing WMO Datasets

Note: The list is not, and cannot be, exhaustive but is included to allow metadata providers to include them in their data descriptions and for users to use them for searching. However, to avoid the situation where data cannot be described, data creators are able to define new keywords, and a mechanism will be put in place to assess proposals for new keywords for inclusion in the list (and hence for having multi-lingual equivalents defined).

The additions to the list proposed by ET-IDM-2 are marked in grey.

- Absolute
- Absorbing
- Absorption
- Acceleration
- Accumulated
- Accumulation
- Acid
- Acoustic
- Active
- Adiabatic
- Adjoint
- Advection
- Aeorological
- Aeorology
- Aeronomy
- Aerosol
- Age
- Ageostrophic
- Aggregated
- Agriculture
- Agrometeorological
- Agrometeorology
- Air
- Albedo
- Alkaline
- Alpine
- Altimeter
- Altitude
- Altocumulus
- Altostratus
- Amount
- Analysis
- Annual
- Anomaly
- Anomalous
- Anticyclone

- Anticlyclonic
- Anthropogenic
- Applied
- Arbitrary
- Ash
- Assimilation
- Asymmetry
- Atmosphere
- Atmospheric
- Automatic
- Avalanche
- Average
- Aviation
- Backscatter
- Balance
- Baroclinic
- Barometer
- Barometric
- Base
- Basic
- BGC (biogeochemical)
- Biennial
- Biogeochemistry
- Biogeochemical
- Biology
- Biomass
- Biometeorology
- Biosphere
- Boundary
- Brightness
- Budget
- BUFR
- Bulb
- Buoy
- Burst
- Cap

- Cave
- Carbon
- Carbon dioxide
- Ceiling
- CFC
- Change
- Chemical
- Chemistry
- Chill
- Chilly
- Circulation
- Cirrocumulus
- Cirrostratus
- Cirrus
- Climate
- Climatology
- Cloud
- CO2
- Coast
- Coastal
- Cold
- Colour
- Column
- Component
- Composite
- Composition
- Compound
- Condensation
- Conductivity
- Constant
- Contour
- Convection
- Convective
- Convergence
- Cooling
- Core

- Coriolis
- Correlation
- Coupled
- Cover
- CREX
- Crop
- Cross
- Cryosphere
- Cryospheric
- Crystal
- Cumulonimbus
- Cumulus
- Current
- Cycle
- Cyclone
- Cyclonic
- Daily
- Data
- Day
- Decadal
- Decay
- Deep
- Degree
- Density
- Depth
- Derivative
- Derived
- Detection
- Dew
- Diabatic
- Diagnostic
- Dielectric
- Differentiate
- Diffusion
- Dimension
- Dioxide

Direction
Discharge
Dispersion
Dissolved
Distance
Distribution
Disturbance
Diurnal
Divergence
Drifting
Drogue
Drop
Droplet
Drought
Dry
Duration
Dust
Dynamical
Dynamics
Earth
Ecology
Eddy
Effect
Electricity
Element
Elevation
Emission
Emissivity
Emittance
Energy
Environment
Episodic
Equatorial
Equilibrium
Equivalent
Ergodic
Erosion
Eulerian
Evaluate
Evaporation
Evapotranspiration
Events
Evolution
Exchange

Extent
Exterior
Extinction
Extra
Extremes
Factor
FAPAR
Feedback
Field
Finite
Fire
Floe
Flow
Fluid
Flux
Fog
Force
Forced
Forcing
Forecast
Forestry
Forward
Freeze
Freshwater
Friction
Front
Frozen
Frost
Future
Gale
Gas
Gaseous
Gauge
GCM
General
Geopotential
Geostationary
Geostrophic
GHG
Glacial
Glacier
Glaciology
Global
GPS

Gradient
Gravity
Greenhouse
GRIB
Grid
Gridded
Ground
Groundwater
Growing
Growth
Gust
Hail
Hair
Halide
Halocarbon
Halogen
Heat
Heating
Heavy
Height
Helium
Hemisphere
Hemispheric
High
Higher
Horizontal
Horizontally
Hour
Hourly
Human
Humidity
Hurricane
Hybrid
Hydrocarbon
Hydrographic
Hydrography
Hydrological
Hydrology
Hydrometeorology
Hydrosphere
Hydrostatic
Ice

Iceberg
Imagery
Imaging
Impact
Incoming
Index
Infrared
Initial
Initialization
Inorganic
Insolation
Instruments
Inter
Integral
Integrate
Intensity
Interaction
Interannual
Interface
Interior
Intermittent
International
Interpolation
Interseasonal
Intersection
Intra
Intraseasonal
Inversion
Invert
Ion
Ionic
Ionosphere
Irradiance
Isentropic
Isobar
Isolate
Isopicnal
Isopleth
Isotherm
Isotope
Isotropic
Kinetic
Lagrangian
Lake

}

Land
Lapse
Laser
Latent
Layer
Level
Lifted
Lightning
Linear
Liquid
Lithosphere
Long
Longwave
Low
Lower
Map
Margin
Marine
Mass
Maximum
Mean
Measurement
Mechanics
Median
Melt
Meridional
Mesoscale
Mesosphere
Mesospheric
Meteorological
Meteorology
Methan
Micro
Micrometeorology
Microwave
Middle
Military
Minimum
Missing
Mist
Mixed
Mixing
Mode
Model

Modeling
Modelling
Modification
Moisture
Molecule
Momentum
Monsoon
Month
Monthly
Moored
Mountain
Multi
Natural
Net
Nimbostratus
Nitrate
Nitrogen
Non
Normal
Nuclei
Nutrients
NWP
Observation
Ocean
Oceanic
Oceanographic
Oceanography
Optic
Optical
Ordinary
Organic
Oscillation
Oxygen
Ozone
Pack
Palaeoclimate
Palaeoclimatology
Paleoclimate
Paleoclimatic
Paleoclimatology
Paleoglacial
Parallel
Parameterization
Partial

Periglacial
Period
Periodic
Permafrost
Perturbation
pH
Phenomena
Phenomenon
Phosphate
Photochemical
Photochemistry
Photolysis
Photon
Photosynthesis
Physical
Physics
Planetary
Point
Polar
Polarization
Pollutant
Pollution
Polynia
Polynya
Potential
Precipitable
Precipitation
Prediction
Present
Pressure
Process
Processed
Profile
Properties
Qualitative
Quality
Quantitative
Quantity
Quasi
Radar
Radiance
Radiation
Radiative
Radiological

Radiosonde
Rain
Rainfall
Rate
Ratio
Re-analysis
Reanalysis
Record
Reflectance
Reflection
Reflectivity
Region
Relative
Remote
Reservoir
Respiration
Resolution
Resources
Response
Ridge
Rings
Rise
River
Rock
Runoff
Salinity
Salt
Sampling
Sand
Satellite
Saturation
Scale
Scaling
Scatter
Scattering
Sea
Season
Seasonal
Seasonally
Section
Sediment
Sedimentation
Semi
Sensed

}

Sensible
Sensing
Sensor
Shallow
Shape
Shear
Sheet
Shelf
Ship
Short
Shortwave
Silicate
Sigma
Significance
Significant
Size
Sky
Sleet
Smog
Smoke
Snow
Snowfall
Snowflake
Soil
Solar
Sounding
Space
Spectra
Spectral
Spectrum
Speed
Sporadic
Spout
Stability
Standard
Static
Station
Storage
Storm
Stratocumulus
Stratopause
Stratosphere
Stratospheric
Stratus

Stream
Streamflow
Streamline
Stress
Storage
Study
Sub
Sub-surface
Sulfur
Sulphide
Sulphure
Sun
Sunshine
Surface
Surge
Swell
SYNOP
Synoptic
System
Teleconnection
Temperature
Tendency
Terrestrial
Thermal
Thermocline
Thermohaline
Theta
Thickness
Thunder
Thunderstorm
Tidal
Tide
Top
Topographic
Topography
Tornado
Total
Trace
Tracer
Track
Tracked
Tracking
Transect
Transient

Transmission
Transmittance
Transpiration
Transport
Tree
Trend
Tritum
Tropical
Tropopause
Troposphere
Tropospheric
Trough
Tsunami
Turbidity
Turbulence
Type
Typhoon
Ultraviolet
Unforced
Uniform
Upper
Urban
Use
Vacillation
Vapor
Vapour
Variability
Variable
Variance
Variation
Vector
Velocity
Vertical
Vertically
Visibility
Voltage
Vortex
Volume
Virtual
Volcanic
Vorticity
Warm
Warming
Water

Wave
Wavelength
Weather
Wet
Wetness
Wind
World
Year
Yearly
Yield
Zonal

}

```
}
```

Reference list of GRDC Reports

Report No. 1 (May 1993)	Second Workshop on the Global Runoff Data Centre, Koblenz, Germany, 15 - 17 June, 1992. (17 pp, annex 73 pp)
Report No. 2 (May 1993)	Dokumentation bestehender Algorithmen zur Übertragung von Abflußwerten auf Gitternetze. (incl. an English abstract in English by the GRDC: Documentation of existing algorithms for transformation of runoff data to grid cells) / G.C. Wollenweber. (71 pp)
Report No. 3 (June 1993)	GRDC - Status Report 1992. (5 pp, annex 5 pp)
Report No. 4 (June 1994)	GRDC - Status Report 1993. (16 pp, annex 34 pp)
Report No. 5 (Nov 1994)	Hydrological Regimes of the Largest Rivers in the World - A Compilation of the GRDC Database. (275 pp)
Report No. 6 (Dec 1994)	Report of the First Meeting of the GRDC Steering Committee, Koblenz, Germany, June 20 - 21, 1994. (10 pp, annex 38 pp)
Report No. 7 (June 1995)	GRDC - Status Report 1994. (12 pp, annex 20 pp)
Report No. 8 (July 1995)	First Interim Report on the Arctic River Database for the Arctic Climate System Study (ACSYS). (34 pp)
Report No. 9 (Aug 1995)	Report of the Second Meeting of the GRDC Steering Committee, Koblenz, Germany, June 27 - 28. (17 pp, annex 34 pp)
Report No. 10 (March 1996)	Freshwater Fluxes from Continents into the World Oceans based on Data of the Global Runoff Data Base / W. Grabs, Th. de Couet, J. Pauler (49 pp, annex 179 pp)
Report No. 11 (April 1996)	GRDC - Status Report 1995. (16 pp, annex 45 pp)
Report No. 12 (June 1996)	Second Interim Report on the Arctic River Database for the Arctic Climate System Study (ACSYS). (39 pp, annex 8 pp)
Report No. 13 (Feb 1997)	GRDC Status Report 1996 (25 pp, annex 36 pp)
Report No. 14 (Feb 1997)	The use of GRDC - information. Review of data use 1993/1994. Status: January 1997 (18 pp, annex 34 pp)

PO-Box 20 02 53, 56002 Koblenz, Germany
Am Mainzer Tor 1, 56068 Koblenz, Germany
phone +49 261 1306-5224
fax +49 261 1306-5280
email grdc@bafg.de
web http://grdc.bafg.de

Reference list of GRDC Reports

Report No. 15 (June 1997)	Third Interim Report on the Arctic River Data Base (ARDB) for the Arctic Climate System Study (ACSYS): Plausibility Control and Data Corrections (Technical Report) (3 pp, annex 20 pp)
Report No. 16 (Aug 1997)	The GRDC Database. Concept and Implementation / J. Pauler, Th. de Couet (38 pp, annex 4 pp)
Report No. 17 (Sep 1997)	Report on the Third Meeting of the GRDC Steering Committee, Koblenz, Germany June 25-27, 1997 (30 pp, annex 137)
Report No. 18 (July 1998)	GRDC Status Report 1997 (13 pp, annex 37 pp)
Report No. 19 (Aug 1998)	Evaluation of Statistical Properties of Discharge Data of Stations Discharging Into the Oceans - Europe and Selected World-Wide Stations / F. Portmann (80 pp)
Report No. 20 (July 1998)	Water Resources Development and the Availability of Discharge Data in WMO Region II (Asia) and V (South-West Pacific) W. Grabs, J. Pauler, Th. de Couet (51 pp, annex 68 pp)
Report No. 21 (Sep 1998)	Analysis of long runoff series of selected rivers of the Asia-Pacific region in relation with climate change and El Niño effects / D. Cluis (23 pp, annex 58 pp)
Report No. 22 (April 1999)	Global, Composite Runoff Fields Based on Observed River Discharge and Simulated Water Balances / B. M. Fekete, C. Vörösmarty, W. Grabs (36 pp, annex 77 pp)
Report No. 23 (Oct 1999)	Report of the fourth Meeting of the GRDC Steering Committee, Koblenz, Germany, 23-25 June 1999 (29 pp, annex 140 pp)
Report No. 24 (Nov 1999)	Use of the GRDC Data 1993-1999: A Comprehensive Summary (48 pp)
Report No. 25 (June 2000)	GIS-related monthly Balance of Water Availability and Demand in Large River Basins - case study for the River Danube / I. Dornblut (27 pp, annex 46 pp)
Report No. 26 (Nov 2000)	Modelling raster-based monthly water balance components for Europe / Carmen Ulmen (133 pp)
Report No. 27 (July 2002)	Water Resources Management Country Profile Germany. A contribution to the Global Water Information Network WWW.GLOBWINET.ORG / R. Winnegge and T. Maurer (32 pp)
Report No. 28 (Nov 2002)	Report of the Fifth Meeting of the GRDC Steering Committee, Koblenz, Germany, 25-28 June 2001 (36 pp, annex 300 pp)

GRDC operates under the auspices of the World Meteorological Organization (WMO) with the support of the Federal Republic of Germany within the Federal Institute of Hydrology (BfG)

Koblenz, November 2004

PO-Box 20 02 53, 56002 Koblenz, Germany
Am Mainzer Tor 1, 56068 Koblenz, Germany
phone +49 261 1306-5224
fax +49 261 1306-5280
email grdc@bafg.de
web http://grdc.bafg.de

Reference list of GRDC Reports

Report No. 29 (Feb 2003)	GRDC Status Report 2002	(28 pp, annex 32 pp)
Report No. 30 (Dec 2003)	Development of an Operational Internet-based Near Real Time Monitoring Tool for Global River Discharge Data / T. Maurer	(23 pp, annex 5 pp)
Report No. 31 (Oct 2004)	Globally agreed standards for metadata and data on variables describing geophysical processes. A fundamental prerequisite to improve the management of the Earth System for our all future / T. Maurer	(43 pp, annex 28 pp)
Report No. 32 (Nov 2004)	Detection of change in world-wide hydrological time series of maximum annual flow / Z.W. Kundzewicz, D. Graczyk, T. Maurer, I. Przymusinska, M. Radziejewski, C. Svensson, M. Szwed	(36 pp, annex 52 pp)
Report No. 33 (Nov 2004)	Trends in flood and low flow series / C. Svensson, Z.W. Kundzewicz, T. Maurer	(26 pp, annex 18 pp)

also available from the GRDC-Homepage as PDF-file

GRDC operates under the auspices of the
World Meteorological Organization (WMO) with
the support of the Federal Republic of Germany
within the Federal Institute of Hydrology (BfG)

Koblenz, November 2004

PO-Box 20 02 53, 56002 Koblenz, Germany
Am Mainzer Tor 1, 56068 Koblenz, Germany
phone +49 261 1306-5224
fax +49 261 1306-5280
email grdc@bafg.de
web http://grdc.bafg.de