EU FP 6 Project
GOCE 037063 with DG Environment

GEO-BENE

Global Earth Observation
-Benefit Estimation: Now, Next and Emerging

STREP
PRIORITY [1.1.6.3]
[Global Change and Ecosystems]

DELIVERABLE D6 (T14)
GEO-BENE DATA BASE REPORT

Due date of deliverable: 30 September 2007

Start date of project: 1 July 2006 Duration: 36 Months

Organization name of lead contractor for this deliverable:
International Institute for Applied Systems Analysis (IIASA)

Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Abstract

The GEO-BENE Data Base Report (D6) is due by the end of September 2007 (T14). The main objective of this task is the analysis of current standards and geodata infrastructures worldwide as well as the definition of Geobene's system architecture and high level data models for the work packages. Therefore, this deliverable covers the following topics:

- **OGC and ISO standards:**
  Compilation of standards and recommendations in the field of geomatics, especially the ISO/TC211 family and the recommendations of the OpenGeospatial Consortium.
- **Woldwide geodata infrastructures:**
  short review of global and regional initiatives like UN-SDI and INSPIRE
- **The Geobene database strategy** and the global database for epic modelling

Document history

Version	Status	Date	Author
0	Input for OGC standards	2007-09-14	FELIS
1	Detailed description of data strategy for high data-demanding application - global database for the EPIC modeling	2007-09-24	SSCRI, BOKU
2	Input for OGC standards	2007-09-24	FELIS
3	OGC Reference Model	2007-09-24	FELIS
4	Dublin Core and input for OGC included.	2007-09-27	FELIS
5	Detailed description of data strategy for high data-demanding application - global database for the EPIC modelling, final version	2007-09-27	SSCRI, BOKU
6	Metadata, Global SDI Initiatives.	2007-09-27	FELIS
7	Final Version	2007-09-28	FELIS

Table of Contents

Executive summary

The main objective of workpackage WP4000 has been the analysis of current geomatics standards, the definition of the geodata infrastructure within the GEOBENE project as well as the definition of the interfaces between different system architecture layers. The results of this workpackage will influence the data storage and will give recommendations and definitions to other work packages.

The outcome of this task / recommendations …..

- **OGC and ISO TC211 standards**, a general review of geomatioc standards that are used world wide
- **A review of Metadata standards (like Dcand ISO 19115)** which are obviously vitally important for this project
- **A review of global spatial data infrastructures**, like UN-SDI, INSPIRE initiative etc.
- since GEO's data infrastrucutre is still developing and in discussion we will review this SDI in deliverable D8
- **GEOBENE's system architecture** that we are going to use for our spatial data and model results within our project.

1 INTRODUCTION

This introduction will give an overview about the envisaged GEOBENE system architecture and geodata infrastructure used to serve both the modellers and interested researchers who can search the Geobene data pool via a standardized meta data interface based on ISO 19115 (and FAO's open source geonetwork server). Furthermore it will give summaries on geomatic standards from OGC and ISO, and describe the data models used within our simulation software (like EPIC,G-EPIC, FASOM, MESSAGE etc)

Geomatic standards:
The Open Geospatial Consortium is a well known organisation of the global geo market players worldwide. They have set up an extensive set of standards, cookbooks and recommendations that are available via http://www.opengeospatial.org (formerly known as OpenGIS).
They are defining a spatial infrastructure based on a Service Oriented Architecture (SOA). Nearly all global initiatives are adopting the ideas from OGC to form an interoperable worldwide community of geodata and services.
Now OGC and ISO, the International Standards Organisation that set the de-jure-standards are working closely together and the ISO technical committee 211 has adopted OGCs standards to form the ISO 19xxx standards family which will also be considered in this deliverable.
OGC makes use of standard web technologies like web services.
OGC defines services for the transmission of raster data like Web Mapping Service, Web Coverage service as well as for vector data (like Web Feature Service, Catalogue Services). Furthermore it defines metadata data that were adopted by ISO and culminate in the definition of metadata according to ISO 19115, which describes geospatial datasets.

The geobene system architecture:
We also describe the GEOBENE high level system architecture that we have developed and that we are implementing at the moment. The client frontend is based on the aforementioned OGC- and ISO standards to guarantee broad acceptance and,. more important, best interoperability. The backend is based on open source software like geonetwork and postgis as well as on propriatory data sources were the used models demand it and where it is not feasible to write connectors (like EPIC).
Parts of this deliverable are based on earlier works from the authors in different projects like in the EC funded projects INSEA, VEPS and NNR.

2 Existing Geomatics Standards

This chapter describes in brief the main international standardization bodies OGC and ISO as well as geomatics standards used in geodata infrastructures,. servicve oriented architectures and the viewpoints from OGC and ISO. See also [URL 2007,1], [URL 2007, 2]. From [Charvat et al, 2005] :

2.1 OGC Standard Specifications

2.1.1 Introduction

The OpenGeoSpatial consortium (OGC, formerly known as OpenGIS consortium) is well known to be a non profit standards organization, a union of (not only) the global players in the geospatial community. Together with ISO/TC211, the Technical Committee 211 it formulates Geomatic Standards. All information regarding their standards work can be found at http://www.geospatial.org [June 2005]

OGC does not only publish approved standards but also serves as a communication- and discussion platform within the geospatial community. OGC's "technical baseline document" holds links to > 140 documents describing Implementation specifications, Discussion papers, Abstract specifications, recommendation papers etc.

Therefore and due the highly dynamic process of publishing standards we do not only focus on approved well known standards like for instance the Web Map Service but also on Services currently in discussion and not yet approved as standards.

2.1.2 OGC Reference Model

The OGC Object Reference Model (ORM) describes the framework of all ongoing work; it is the technical baseline of OGC. It is a "must-read" for the understanding of OGC standards but however is not a standard itself but a document due to change.

The OpenGIS Reference Model (ORM) provides an architecture framework for the ongoing work of the OGC. Further, the ORM provides a framework for the OGC Technical Baseline. The OGC Technical Baseline consists of the currently approved OpenGIS Specifications as well as for a number of candidate specifications that are currently in progress.

The ORM has the following purposes:

- Provides a foundation for coordination and understanding (both internal and external to OGC) of ongoing OGC activities and the Technical Baseline;
- Update/Replacement of parts of the 1998 OpenGIS Guide;
- Describes the OGC requirements baseline for geospatial interoperability;
- Describes the OGC architecture framework through a series of non-overlapping viewpoints: including existing and future elements;
- Regularize the development of domain-specific interoperability architectures by providing examples.

The ORM is a living document that will be revised on a regular basis to continually and accurately reflect the ongoing work of the Consortium.

Background information for this chapter can be found at
http://portal.opengeospatial.org/files/?artifact_id=3836 [september 2007]

The ORM adopts the so called Reference Model for Open Distributed Processing (RM-ODP), a standard for architecting open distributed systems. The RM-ODP addresses five different *viewpoints* on systems:

Table 1 - RM-ODP viewpoints

Viewpoint Name	Definition of RM-ODP Viewpoint
Enterprise	Focuses on the purpose, scope and policies for that system.
Information	Focuses on the semantics of information and information processing.
Computational	Captures component and interface details without regard to distribution
Engineering	Focuses on the mechanisms and functions required to support distributed interaction between objects in the system.
Technology	Focuses on the choice of technology.

(source: 03-040_OpenGIS_Reference_Model_ORM_version_0.1.2.pdf)

The *Enterprise* viewpoint focussed on business perspective, and policies, the *Informational* viewpoint on semantics of information and information processing like Geographic Features, general feature models etc. The Computational viewpoint deals with a systems being treated as sets of (web) services that interact at interfaces. The *Engineering* viewpoint in turn is concerned with communication, computing systems, software processing etc. (keywords: mapping logical rchitectures to ohysical architectures, thick vs. thin clients etc). Last not least the *Technology* viewpoint deals with underlying infrastructures in distributed systems (keywords Feature Model, Coverage Model etc).

OGC's computational viewpoint in the Object Reference Model defines the OpenGIS Service Framework and within this categorizes the different services into 6 different top level categories in order to better classify, catalogue and find relevant services:

2.1.3 The OGC Web Services Service Framework OSF

The OWS Service Framework (OSF) identifies services, interfaces and exchange protocols that can be utilized by any application. OpenGIS Services are implementations of services that conform to OpenGIS Implementation Specifications. Compliant applications, called OpenGIS Applications, can then "plug into" the framework to join the operational environment.

By building applications to common interfaces, each application can be built without a-priori or run-time dependencies on other applications or services. Applications and services can be added, modified, or replaced without impacting other applications. In addition, operational workflows can be changed on-the-fly, allowing rapid response to time-critical situations. This loosely coupled, standards-based approach to development results in very agile systems—systems that can be flexibly adapted to changing requirements and technologies

The next figure, the OWS Service Framework, shows the collaboration and dependance of the categories and services:
The OSF is designed to meet the following purposes:

- Provide a framework for coordinated development of new and extended services
- Enable interoperable services through standard interfaces and encodings
- Support publishing, discovery and binding of services through service metadata
- Allow separation of data instances from service instances
- Enable use of a provider's service on another provider's data
- Define a framework that can be implemented in multiple ways

The OSF is a profile of the OGC services taxonomy. The OSF categorizes services into five categories that correspond to the OGC services taxonomy top-level domains as shown in next Table.

Table 3 - OSF and ISO 19119 Service Categories

OSF Service Categories	OGC Service Taxonomy Categories
Application Services	Geographic Human Interaction
Registry Services	Geographic Information Management
Data Services	Geographic Information Management
Portrayal Services	Geographic Human Interaction
Processing Services	Geographic Processing Interaction

fig. 1 OSF and ISO 19119 Service Categories

2.1.4 Geographic services taxonomy

The following sub-clauses provide examples of geographic services within the geographic services taxonomy. It is not required that a system provide any service listed in these sub-clauses. It is required that if a system provides a service named in these sub-clauses that the service shall be categorized as defined in these sub-clauses. A service catalogue compliant with this International Standard shall categorize service metadata instances in the categories of the geographic service taxonomy.
If a service uses the name of an example service, the service shall provide the functionality that is defined in these sub-clauses. For example, if a service titled catalogue viewer is provided, it shall perform the services defined for the catalogue viewer in the geographic human interaction services category. Systems providing services should name services as found in the service examples.

2.1.4.1 Geographic human interaction services

Geographic human interaction services shall be a category in the geographic service taxonomy. Examples of human interaction services for working with geographic data and services:

- Catalogue viewer. Client service that allows a user to interact with a catalogue to locate, browse, and manager metadata about geographic data or geographic services.

- Geographic viewer. Client service that allows a user to view one or more feature collections or coverages.This viewer allows a user to interact with map data, e.g., displaying, overlaying and querying. An example is the viewer client generator defined in ISO 19128.

2.1.4.1.1 Portrayal and human interface

Portrayal is the presentation of information to humans, e.g., a map. A map is a two-dimensional visual portrayal of geospatial data; a map is not the data itself. Two or more maps with the same geographic extent and coordinate reference system can be accurately layered to produce a composite map. Information types associated with geospatial data visualization are shown in the context of the portrayal process

1. Image or picture of the data, e.g., a map to be displayed.
2. Display elements, e.g., lexical description of graphics to be drawn onto the target display.

2.1.4.1.2 Thin and thick clients

The engineering viewpoint helps to articulate a key distinction among distributed systems:

• Thin clients rely on invoking the services of other components (servers, middleware) for most of the computation they need to function in the system; they also rely on other components to manage most of the data and metadata they need.

• Thick clients handle much of the necessary computation and data/metadata management themselves; and rather than invoking the processing services of other components, they obtain their inputs through low-level data-access requests.

A thick client requires less functionality on the part of the server and other components; but a thin client is easier to build or to embed into general-purpose software components. The distinction often has quite tangible implications: thin clients are typically simple software with limited functions and flexibility, and smaller RAM and CPU requirements, often suitable for handheld or mobile devices. Thick clients usually require a significant portion of (at least) a microcomputer’s resources, but provide greater flexibility and capacity to decode, transform, render, and interact with retrieved data.

2.1.5 Complete List of OGC specifications:

Catalogue Service	**http://www.opengeospatial.org/standards/cat**
Coordinate Transformation	**http://www.opengeospatial.org/standards/ct**
Filter Encoding	**http://www.opengeospatial.org/standards/filter**
Geographic Objects	**http://www.opengeospatial.org/standards/go**
Geography Markup Language	**http://www.opengeospatial.org/standards/gml**
GML in JPEG 2000	**http://www.opengeospatial.org/standards/gmljp2**
Grid Coverage Service	**http://www.opengeospatial.org/standards/gc**
Location Services (OpenLS)	**http://www.opengeospatial.org/standards/olscore**
Sensor Model Language (SensorML)">Sensor Model Language	**http://www.opengeospatial.org/standards/sensorml**
Sensor Planning Service	**http://www.opengeospatial.org/standards/sps**
Simple Feature Access 1	**http://www.opengeospatial.org/standards/sfa**
Simple Feature Access 2	**http://www.opengeospatial.org/standards/sfs**
Simple Features CORBA	**http://www.opengeospatial.org/standards/sfc**
Simple Features OLE/COM	**http://www.opengeospatial.org/standards/sfo**
Styled Layer Descriptor	**http://www.opengeospatial.org/standards/sld**
OpenGIS Symbology	**http://www.opengeospatial.org/standards/symbol**

Encoding Implementaion Specification">Symbology Encoding	
Transducer Markup Language (TML) Encoding Specification">Transducer Markup Language	**http://www.opengeospatial.org/standards/tml**
Web Coverage Service	**http://www.opengeospatial.org/standards/wcs**
Web Feature Service	**http://www.opengeospatial.org/standards/wfs**
Web Map Context	**http://www.opengeospatial.org/standards/wmc**
Web Map Service	**http://www.opengeospatial.org/standards/wms**
Web Service Common	**http://www.opengeospatial.org/standards/common**

Many more discussion papers, recommendations and best practices documents exist, but the above list shows the document that are approved as official OGC standard.

2.2 ISO TC211

The ISO is the International Standards Organisation located in Geeneva, Switzerland. It is one of the major standardisation bodies worldwide.

The Technical Committee TC2211 ist the committee that formulates standards in the field of geomatics. It was founded on behalf of the NATO geomatics working group DGIWG and national standards efforts in Caanada and the U.S: after a formal proposal from Canada in 1994

The scope of ISO TC211 covers the following [Kresse, Fadaiae, 2004]:
“The objective of the work of ISO/TC211 is to establish a set of standards for ‘Geographic Information/Geomatics’. The standards would specif an infrastructure and the required services for the handling of geographic data including management, acquisition,processing, analysis, access, presentation, and transfer.Where pssoible, the standards wouöld link to other appropriate standards for information technology and provide a framework for the development of sector specific applications (ISO04 2003)”

The Technical Committee TC211 is currently developing ISO’s standards in the geomatics field. They are workling jointly together with the OpenGeospatial Consortium OGC, and ISO standards are close to OGC’s standards.

The following table gives an overview about ISO/TC211s available standards in geomatics [URL 2007, 13]

Standard and/or project	**Description**
ISO 6709:1983	Standard representation of latitude, longitude and altitude for geographic point locations
ISO/DIS 6709	Standard representation of geographic point location by coordinates
ISO 19101:2002	Geographic information -- Reference model
ISO/CD TS 19101-2	Geographic information -- Reference model -- Part 2: Imagery
ISO/TS 19103:2005	Geographic information -- Conceptual schema language
ISO 19105:2000	Geographic information -- Conformance and testing
ISO 19106:2004	Geographic information -- Profiles
ISO 19107:2003	Geographic information -- Spatial schema
ISO 19108:2002	Geographic information -- Temporal schema
ISO 19108:2002/Cor 1:2006	
ISO 19109:2005	Geographic information -- Rules for application schema

ISO 19110:2005	Geographic information -- Methodology for feature cataloguing
ISO 19110:2005/CD Amd 1	
ISO 19111:2007	Geographic information -- Spatial referencing by coordinates
ISO/CD 19111-2	Geographic information -- Spatial referencing by coordinates
ISO 19112:2003	Geographic information -- Spatial referencing by geographic identifiers
ISO 19113:2002	Geographic information -- Quality principles
ISO 19114:2003	Geographic information -- Quality evaluation procedures
ISO 19114:2003/Cor 1:2005	
ISO 19115:2003	Geographic information -- Metadata
ISO 19115:2003/Cor 1:2006	
ISO/CD 19115-2	Geographic information -- Metadata -- Part 2: Extensions for imagery and gridded data
ISO 19116:2004	Geographic information -- Positioning services
ISO/NP 19117	Geographic information -- Portrayal
ISO 19117:2005	Geographic information -- Portrayal
ISO 19118:2005	Geographic information -- Encoding
ISO/CD 19118	Geographic information -- Encoding
ISO 19119:2005	Geographic information -- Services
ISO 19119:2005/DAmd 1	Extensions of the service metadata model
ISO/TR 19120:2001	Geographic information -- Functional standards
ISO/TR 19121:2000	Geographic information -- Imagery and gridded data
ISO/TR 19122:2004	Geographic information / Geomatics -- Qualification and certification of personnel
ISO 19123:2005	Geographic information -- Schema for coverage geometry and functions
ISO 19125-1:2004	Geographic information -- Simple feature access -- Part 1: Common architecture
ISO 19125-2:2004	Geographic information -- Simple feature access -- Part 2: SQL option
ISO/CD 19126	Geographic information -- Feature concept dictionaries and registers
ISO/TS 19127:2005	Geographic information -- Geodetic codes and parameters
ISO 19128:2005	Geographic information -- Web map server interface
ISO/WD TS 19129	Geographic information -- Imagery, gridded and coverage data framework
ISO 19131:2007	Geographic information -- Data product specifications
ISO 19132	Geographic information -- Location-based services -- Reference model
ISO 19133:2005	Geographic information -- Location-based services -- Tracking and navigation
ISO 19134:2007	Geographic information -- Location-based services -- Multimodal routing and navigation
ISO 19135:2005	Geographic information -- Procedures for item registration
ISO 19136:2007	Geographic information -- Geography Markup Language (GML)
ISO 19137:2007	Geographic information -- Core profile of the spatial schema
ISO/TS 19138:2006	Geographic information -- Data quality measures
ISO/TS 19139:2007	Geographic information -- Metadata -- XML schema implementation
ISO/DIS 19141	Geographic information -- Schema for moving features
ISO/CD 19142	Geographic information -- Web Feature Service
ISO/CD 19143	Geographic information -- Filter encoding
ISO/CD 19144-1	Geographic information - Classification Systems -- Part 1: Classification system structure
ISO/CD 19144-2	Geographic information - Classification Systems -- Part 2: Land Cover Classification System LCCS

ISO/WD 19146	Geographic information -- Cross-domain vocabularies
ISO/NP 19149	Geographic information -- Rights expression language for geographic information -- GeoREL
ISO/NP 19151	Dynamic position identification scheme for Ubiquitous space (u-position)

2.3 Relevant standards for the GEOBENE project

The most important services for the GEOBENE project are the standard services for the transmission of raster data, vector dat and metadata, respectively.

The transmission of raster data is guaranteed by the Web Map Service (WMS) and Web Coverage Service (WCS), whereas the Web Feature service is suitable for vector data. In the field of metadata we use the ISO19115 standard (Dublin Core also possible) and a Catalogue Service.

2.3.1 The Web Map Service (WMS)

[URL 2007 ,15]
"A Web Map Service (WMS) produces maps of spatially referenced data dynamically from geographic information. This International Standard defines a "map" to be a portrayal of geographic information as a digital image file suitable for display on a computer screen. A map is not the data itself. WMS-produced maps are generally rendered in a pictorial format such as PNG, GIF or JPEG, or occasionally as vector-based graphical elements in Scalable Vector Graphics (SVG) or Web Computer Graphics Metafile (WebCGM) formats. This International Standard defines three operations: one returns service-level metadata; another returns a map whose geographic and dimensional parameters are well-defined; and an optional third operation returns information about particular features shown on a map. Web Map Service operations can be invoked using a standard web browser by submitting requests in the form of Uniform Resource Locators (URLs). The content of such URLs depends on which operation is requested. In particular, when requesting a map the URL indicates what information is to be shown on the map, what portion of the Earth is to be mapped, the desired coordinate reference system, and the output image width and height. When two or more maps are produced with the same geographic parameters and output size, the results can be accurately overlaid to produce a composite map. The use of image formats that support transparent backgrounds (e.g., GIF or PNG) allows underlying maps to be visible. Furthermore, individual maps can be requested from different servers. The Web Map Service thus enables the creation of a network of distributed map servers from which clients can build customized maps."

2.3.2 Web Feature Service (WFS)

From [URL 2007, 15] OGC WFS spcification:
"The OGC Web Map Service allows a client to overlay map images for display served from multiple Web Map Services on the Internet. In a similar fashion, the OGC Web Feature Service allows a client to retrieve and update geospatial data encoded in Geography Markup Language (GML) from multiple Web Feature Services.
The requirements for a Web Feature Service are:

1. The interfaces must be defined in XML.
2. GML must be used to express features within the interface.
3. At a minimum a WFS must be able to present features using GML.

4. The predicate or filter language will be defined in XML and be derived from CQL as defined in the OpenGIS Catalogue Interface Implementation Specification.
5. The datastore used to store geographic features should be opaque to client applications and their only view of the data should be through the WFS interface.
6. The use of a subset of XPath expressions for referencing properties. "

The exchange format between WFS client and server is based on the Geographic Markup Language GML.
An example for a WFS request and the returned document can be found in the annex
For details on the WFS specification please see [URL 2007,. 15].

2.3.3 Web Coverage Service (WCS)

From [URL 2007, 17] OGC WCS specification:
"The Web Coverage Service (WCS) supports electronic retrieval of geospatial data as "coverages" – that is, digital geospatial information representing space-varying phenomena.
A WCS provides access to potentially detailed and rich sets of geospatial information, in forms that are useful for client-side rendering, multi-valued coverages, and input into scientific models and other clients. The WCS may be compared to the OGC Web Map Service (WMS) and the Web Feature Service (WFS); like them it allows clients to choose portions of a server's information holdings based on spatial constraints and other criteria.
Unlike the WMS [OGC 04-024], which portrays spatial data to return static maps (rendered as pictures by the server), the Web Coverage Service provides available data together with their detailed descriptions; defines a rich syntax for requests against these data; and returns data with its original semantics (instead of pictures) which may be interpreted, extrapolated, etc. – and not just portrayed.
Unlike WFS [OGC 02-058], which returns discrete geospatial features, the Web Coverage Service returns coverages representing space-varying phenomena that relate a spatio-temporal domain to a (possibly multidimensional) range of properties.
The Web Coverage Service provides three operations: *GetCapabilities*, *DescribeCoverage,* and *GetCoverage*. The *GetCapabilities* operation returns an XML document describing the service and brief descriptions of the coverages that clients may request. Clients would generally run the *GetCapabilities* operation and cache its result for use throughout a session, or reuse it for multiple sessions. When the *GetCapabilities* operation does not return such descriptions, then equivalent information must be available from a separate source, such as an image catalog.
The *DescribeCoverage* operation lets clients request a full description of one or more coverages served by a particular WCS server. The server responds with an XML document that fully describes the identified coverages.
The *GetCoverage* operation is normally run after *GetCapabilities* and *DescribeCoverage* operation responses have shown what requests are allowed and what data are available. The *GetCoverage* operation returns a coverage (that is, values or properties of a set of geographic locations), encoded in a well-known coverage format. Its syntax and semantics bear some resemblance to the WMS *GetMap* and WFS *GetFeature* requests, but several extensions support the retrieval of coverages rather than static maps or discrete features."

2.3.4 Web Catalog Services (OGC CS)

From [URL 2007, 18]:
„Catalogue services support the ability to publish and search collections of descriptive information (metadata) for data, services, and related information objects. Metadata in

catalogues represent resource characteristics that can be queried and presented for evaluation and further processing by both humans and software. Catalogue services are required to support the discovery and binding to registered information resources within an information community. “

2.3.5 Metadata standard ISO 19115

For a thorough discussion of ISO 19115 please see chapter 3 “Metadata Standards”.

3 Metadata Standards

Although Metadata standards are part of the aforementioned standards, specification information on metadata has been compiled into an extra chapter because metadata are most important for geodata retrieval.

In its practical meaning metadata are "data about data" – that is, data describing resources or data, so that they can be retrieved from a large collection of data

From [http://www.library.uiuc.edu/systems/metadata_doc/what_is_metadata.htm]:

"Metadata has been with us since the first librarian made a list of the items on a shelf of handwritten scrolls. The term "meta" comes from a Greek word that denotes "alongside, with, after, next." More recent Latin and English usage would employ "meta" to denote something transcendental, or beyond nature. Metadata, then, can be thought of as data about other data. It is the Internet-age term for information that librarians traditionally have put into catalogs, and it most commonly refers to descriptive information about Web resources.

A metadata record consists of a set of attributes, or elements, necessary to describe the resource in question. For example, a metadata system common in libraries -- the library catalog -- contains a set of metadata records with elements that describe a book or other library item: author, title, date of creation or publication, subject coverage, and the call number specifying location of the item on the shelf."

In the field of geomatics the most widely used international standards are covered by Dublin Core, FGDC and ISO19115 with its xml schema ISO19139.

3.1 Dublin Core

Dublin Core is not only used in Geomatics, but also in all kind of digital information and in particular in library management. Its core set of elements only covers 15 elements and is therefore easy to use.

[Charvat et al, 2005]"The Dublin Core metadata element set is a standard for cross-domain information resource description. Here an information resource is defined to be "anything that has identity". This is the definition used in Internet RFC 2396, "Uniform Resource Identifiers (URI): Generic Syntax", by Tim Berners-Lee et al. There are no fundamental restrictions to the types of resources to which Dublin Core metadata can be assigned.

Three formally endorsed versions exist of the Dublin Core Metadata Element Set, version 1.1:

- ISO Standard 15836-2003 (February 2003): http://www.niso.org/international/SC4/n515.pdf
- NISO Standard Z39.85-2001 (September 2001): http://www.niso.org/standards/resources/Z39-85.pdf
- CEN Workshop Agreement CWA 13874 (March 2000, no longer available)

The current document has been brought into line with the ISO and NISO standards. The more comprehensive document "DCMI Metadata Terms" (http://dublincore.org/documents/dcmi-terms/) includes the latest and authoritative term declarations for the Dublin Core Metadata Element Set, Version 1.1.

For an overview and links to full specifications of all metadata terms maintained by the Dublin Core Metadata Initiative please see: http://dublincore.org/usage/documents/overview/.

The Elements

Element Name: Title	
Label:	Title
Definition:	A name given to the resource.
Comment:	Typically, Title will be a name by which the resource is formally known.
Element Name: Creator	
Label:	Creator
Definition:	An entity primarily responsible for making the content of the resource.
Comment:	Examples of Creator include a person, an organization, or a service. Typically, the name of a Creator should be used to indicate the entity.
Element Name: Subject	
Label:	Subject and Keywords
Definition:	A topic of the content of the resource.
Comment:	Typically, Subject will be expressed as keywords, key phrases or classification codes that describe a topic of the resource. Recommended best practice is to select a value from a controlled vocabulary or formal classification scheme.
Element Name: Description	
Label:	Description
Definition:	An account of the content of the resource.
Comment:	Examples of Description include, but are not limited to: an abstract, table of contents, reference to a graphical representation of content or a free-text account of the content.
Element Name: Publisher	
Label:	Publisher
Definition:	An entity responsible for making the resource available
Comment:	Examples of Publisher include a person, an organization, or a service. Typically, the name of a Publisher should be used to indicate the entity.
Element Name: Contributor	
Label:	Contributor
Definition:	An entity responsible for making contributions to the content of the resource.

Comment:	Examples of Contributor include a person, an organization, or a service. Typically, the name of a Contributor should be used to indicate the entity.
Element Name: Date	
Label:	Date
Definition:	A date of an event in the lifecycle of the resource.
Comment:	Typically, Date will be associated with the creation or availability of the resource. Recommended best practice for encoding the date value is defined in a profile of ISO 8601 [W3CDTF] and includes (among others) dates of the form YYYY-MM-DD.
Element Name: Type	
Label:	Resource Type
Definition:	The nature or genre of the content of the resource.
Comment:	Type includes terms describing general categories, functions, genres, or aggregation levels for content. Recommended best practice is to select a value from a controlled vocabulary (for example, the DCMI Type Vocabulary [DCT1]). To describe the physical or digital manifestation of the resource, use the FORMAT element.
Element Name: Format	
Label:	Format
Definition:	The physical or digital manifestation of the resource.
Comment:	Typically, Format may include the media-type or dimensions of the resource. Format may be used to identify the software, hardware, or other equipment needed to display or operate the resource. Examples of dimensions include size and duration. Recommended best practice is to select a value from a controlled vocabulary (for example, the list of Internet Media Types [MIME] defining computer media formats).
Element Name: Identifier	
Label:	Resource Identifier
Definition:	An unambiguous reference to the resource within a given context.
Comment:	Recommended best practice is to identify the resource by means of a string or number conforming to a formal identification system. Formal identification systems include but are not limited to the Uniform Resource Identifier (URI) (including the Uniform Resource Locator (URL)), the Digital Object Identifier (DOI) and the International Standard Book Number (ISBN).

Element Name: Source	
Label:	Source
Definition:	A Reference to a resource from which the present resource is derived.
Comment:	The present resource may be derived from the Source resource in whole or in part. Recommended best practice is to identify the referenced resource by means of a string or number conforming to a formal identification system.
Element Name: Language	
Label:	Language
Definition:	A language of the intellectual content of the resource.
Comment:	Recommended best practice is to use RFC 3066 [RFC3066] which, in conjunction with ISO639 [ISO639]), defines two- and three-letter primary language tags with optional subtags. Examples include "en" or "eng" for English, "akk" for Akkadian", and "en-GB" for English used in the United Kingdom.
Element Name: Relation	
Label:	Relation
Definition:	A reference to a related resource.
Comment:	Recommended best practice is to identify the referenced resource by means of a string or number conforming to a formal identification system.
Element Name: Coverage	
Label:	Coverage
Definition:	The extent or scope of the content of the resource.
Comment:	Typically, Coverage will include spatial location (a place name or geographic coordinates), temporal period (a period label, date, or date range) or jurisdiction (such as a named administrative entity). Recommended best practice is to select a value from a controlled vocabulary (for example, the Thesaurus of Geographic Names [TGN]) and to use, where appropriate, named places or time periods in preference to numeric identifiers such as sets of coordinates or date ranges.
Element Name: Rights	
Label:	Rights Management
Definition:	Information about rights held in and over the resource.
Comment:	Typically, Rights will contain a rights management statement for the resource, or reference a service providing such information. Rights information often encompasses Intellectual Property Rights (IPR), Copyright, and various Property Rights. If the Rights element is absent, no assumptions may be made about any rights held in or over the resource.

3.2 FGDC

The FGDC is a national standard from the U.S: (Federal Geographic Data Committee).

From [http://www.fgdc.gov/metadata/geospatial-metadata-standards]

"The Content Standard for Digital Geospatial Metadata (CSDGM), Vers. 2 (FGDC-STD-001-1998) is the US Federal Metadata standard. The Federal Geographic Data Committee originally adopted the CSDGM in 1994 and revised it in 1998. According to Executive Order 12096 all Federal agencies are ordered to use this standard to document geospatial data created as of January, 1995. The standard is often referred to as the FGDC Metadata Standard and has been implemented beyond the federal level with State and local governments adopting the metadata standard as well.

The international community, through the International Organization of Standards (ISO), has developed and approved an international metadata standard, ISO 19115. As a member of ISO, the US required to revise the CSDGM in accord with ISO 19115. Each nation can craft their own profile of ISO 19115 with the requirement that it include the 13 core element. The FGDC is currently leading the development of a US Profile of the (ISO) international metadata standard, ISO 19115."

ISO 19115 Core Metadata Elements

Mandatory Elements:	Conditional Elements:
Dataset title	Dataset responsible party
Dataset reference date	Geographic location by coordinates
Dataset language	Dataset character set
Dataset topic category	Spatial resolution
Abstract	Distribution format
Metadata point of contact	Spatial representation type
Metadata date stamp	Reference system
	Lineage statement
	On-line Resource
	Metadata file identifier
	Metadata standard name
	Metadata standard version
	Metadata language
	Metadata character set

"

The core set of FGDC consists of the following:

```xml
<?xml version="1.0" encoding="UTF-8"?>
<metadata>
	<idinfo>
		<citation>
			<citeinfo>
				<origin/>
				<pubdate/>
				<title>Template for FGDC</title>
				<pubinfo>
					<pubplace/>
					<publish/>
				</pubinfo>
			</citeinfo>
		</citation>
		<descript>
			<abstract/>
			<purpose/>
		</descript>
		<timeperd>
			<timeinfo>
				<rngdates>
					<begdate/>
					<enddate/>
				</rngdates>
			</timeinfo>
			<current/>
		</timeperd>
		<status>
			<progress/>
			<update/>
		</status>
		<spdom>
			<bounding>
				<westbc/>
				<eastbc/>
				<northbc/>
				<southbc/>
			</bounding>
		</spdom>
		<keywords>
			<theme>
				<themekt/>
				<themekey/>
			</theme>
		</keywords>
		<accconst/>
		<useconst/>
	</idinfo>
	<metainfo>
		<metd/>
		<metc>
			<cntinfo>
				<cntperp>
					<cntper/>
				</cntperp>
				<cntaddr>
					<addrtype/>
					<city/>
					<state/>
					<postal/>
				</cntaddr>
				<cntvoice/>
			</cntinfo>
		</metc>
		<metstdn/>
		<metstdv/>
	</metainfo>
</metadata>
```

3.3 ISO 19115

The ISO 19115 together with it's xml schema formulation ISO 19139 form the international standard for the description of metadata. Therefore it should be the standard of choice for all geodata and it is the standard that is used within in the GEOBENE project.

The metadata elements in ISO19115 are grouped into two levels: core obligatoty elements and the mandatory elements.

The following UML diagram shows the different packages of ISO19115 [Kresse, Fadaie, 2004], [OGC 2000]:

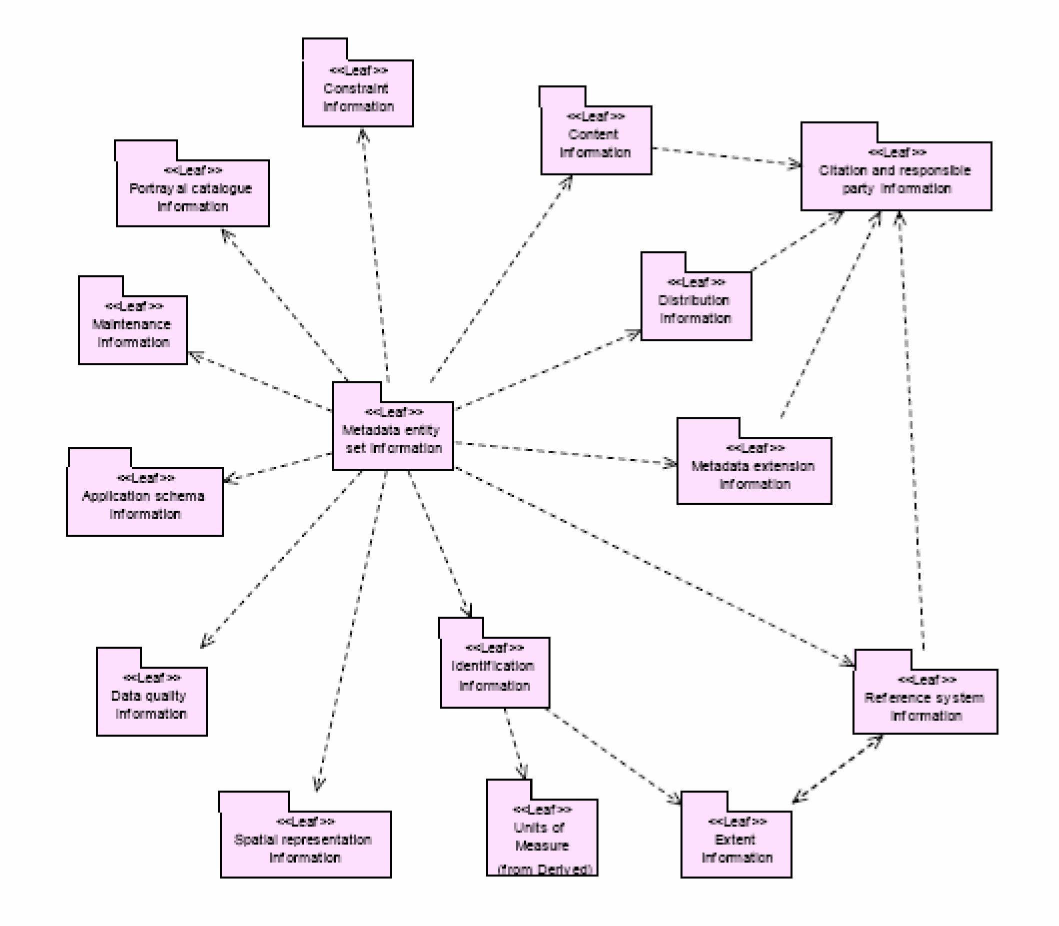


fig. 2 UML package diagram from ISO19115

For a complete example of ISO 19115 please see the annex I with a metadata description from one of our datasets as exported from metadata server *geonetwork.*

From [OGC 2000]:

"Core metadata for geographic datasets

Listed are the core metadata elements required to identify a dataset, typically for catalogue purposes. This list contains metadata elements answering the following questions: "Does a dataset on a specific topic exist ('what')?", "For a specific place ('where')?", "For a specific date or period ('when')?" and "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. Dataset metadata profiles of this International Standard shall include this core.
Listed below are the core metadata elements (mandatory and recommended optional) required for describing a dataset. An "M" indicates that the element is mandatory. An "O" indicates that the element is optional. A "C" indicates that the element is mandatory under certain conditions."

Core metadata for geographic datasets

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

4 Global and regional Geodata Infrastructure Initiatives

The following chapter will give a short summary of geodata initiatives. Among the global initiatives like GSDI and UN-SDI also regional initiatives like the European INSPIRE directive exist. We will not focus on local initiatives like for example the German Geodata infrastructure GDI-DE (http://www.gdi-de.org/de/f_start.html).

For a comprehensive list of global, regional and local initiaticves please see

http://www.gsdi.org/SDILinks.asp (Sep. 2007)

4.1 GSDI

"The GSDI Association is an inclusive organization of organizations, agencies, firms, and individuals from around the world. The purpose of the organization is to promote international cooperation and collaboration in support of local, national and international spatial data infrastructure developments that will allow nations to better address social, economic, and environmental issues of pressing importance." [http://www.gsdi.org]I

In it's "Spatial Data Infrastructure Cookbook " [URL 2007, 5] the GSDI promotes the use of standarditzes interfaces for the exchange of spatial data, especially the use of OGC- and ISO standards

4.2 INSPIRE and ESDI

INSPIRE (Infrastructure for Spatial Information in Europe) is an ambitious aim of the European Council to harmonise and make available spatial data in Europe, which will lead towards a European Spatial Data Infrastructure (ESDI). The Inspire principles comprise of the following:

INSPIRE: INfrastructure for SPatial InfoRmation in Europe

The INfrastructure for SPatial InfoRmation in Europe initiative (INSPIRE) aims at making available relevant, harmonised and quality geographic information for the purpose of formulation, implementation, monitoring and evaluation of Community policy-making.

INSPIRE Principles

- Data should be collected once and maintained at the level where this can be done most effectively
- It should be possible to combine seamlessly spatial data from different sources and share it between many users and applications
- Spatial data should be collected at one level of government and shared between all levels
- Spatial data needed for good governance should be available on conditions that are not restricting its extensive use
- It should be easy to discover which spatial data is available, to evaluate its fitness for purpose and to know which conditions apply for its use.

The initiative has now reached Directive status:

"Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)

was published in the official Journal on the 25th April 2007. It is known in short as the 'INSPIRE
Directive' and entered into force on the 15th May 2007.
The full text of the published Directive is available at the following link:
http://inspire.jrc.it/directive/l_10820070425en00010014.pdf" [URL 2007, 3]

The INSPIRE initiative has formulated requirements for the directive in their INSPIRE Work Programme, Transposition Phase 2007-2009, [URL, 2007, 19]:

"2 Requirements of the INSPIRE Directive

The aim of this chapter is to summarize the requirements for Implementing Rules based on the final text of the Directive. The following subsections address these requirements for metadata, network services, data specifications, data and service sharing, and monitoring and reporting. The roadmap at the end of the chapter summarises the deadlines associated to the development and implementation of the Implementing Rules.

2.1 Metadata

Member States shall create metadata and keep them up to date
• Metadata shall include:
– conformity with IR on interoperability,
– conditions for access and use of data sets and services,
– quality and validity,
– the public authorities responsible, and
– limitations on public access.
• An Implementing Rule for metadata shall be adopted within one year after the entry into force of the Directive.
• Once the Implementing Rule for metadata is adopted, metadata must be created:
– within 2 years for Annex I, II spatial data themes;
– within 5 years for Annex III spatial data themes.

2.2 Network services

Member States shall operate a network of the following services available to the public for data sets and services for which metadata has been created:
− Discovery services; No charge
− View services; No charge (exceptions)
− Download services;
− Transformation services,
− Services allowing spatial data services to be invoked;
Moreover,
− Member States shall ensure the technical possibility, for public authorities, to link their spatial data sets and services;
− Access to services may be restricted;
− Services shall be available on request to 3rd parties under conditions;
− An INSPIRE Geo-portal at Community level shall be established.
An Implementing Rule shall be adopted for the different types of service according to the INSPIRE Roadmap in section 2.6.

2.3 Data sharing

Member States shall adopt measures for the sharing of data and services between public authorities for public tasks relating to the environment without restrictions occurring at the point of use.
Public authorities may license and/or charge other public authorities and Community institutions provided that:
Infrastructure for Spatial Information in Europe Reference: INSPIRE IR WP2007-2009-v1 0.doc
CT Implementing Rules development 2007-05-15 Page 9 of 46
• It is compatible with the objective to facilitate sharing between public authorities.
• It is restricted to the minimum necessary to ensure sustained availability and quality of the data and services.
When spatial data or services are provided to Community institutions for reporting obligations under Community law relating to the environment then this will not be subject to charging.
Member States shall provide the institutions and bodies of the Community with access to spatial data sets and services in accordance with harmonised conditions. Implementing Rules governing those conditions shall be adopted according to the INSPIRE Roadmap presented in section 2.6.

2.4 Interoperability of spatial datasets and services

An Implementing Rule shall be adopted for interoperability and where practical for harmonisation of spatial data sets and services.
The Implementing Rule will be adopted within two years after entry into force for data sets corresponding to the data themes in the Annex I to the INSPIRE Directive and within 5 years for those covered in Annex II and III.
• Required for harmonised data specifications are:
– for Annex I, II, III:
• definition and classification of spatial objects,
• geo-referencing;
– and for Annex I, II:
• common system of unique identifiers for spatial objects;
• relationship between spatial objects;
• key attributes and corresponding multilingual thesauri;
• how to exchange the temporal dimension of the data;
• how to exchange updates of the data.
• 3rd parties shall have access to these specifications at conditions not restricting their use;
• Cross-border issues shall be agreed on.

2.5 Monitoring and reporting

Member States shall:
• Monitor the implementation and use of their infrastructures for spatial information, and make the results accessible to the Commission and to the public on a permanent basis.
• Send, no later than 3 years after the entry into force, to the Commission a report including summary descriptions of:

- Co-ordination between public sector providers and users;
- Relationship with third parties;
- Organisation of quality assurance;
- Contribution by public authorities or third parties;
- The use of the infrastructure;
- Data sharing agreements between public authorities;
- Costs and benefits of implementing the Directive.

An Implementing Rule governing the monitoring process and the reporting obligation shall be adopted within one year after the entry into force of the Directive.
Infrastructure for Spatial Information in Europe Reference: INSPIRE IR WP2007-2009-v1 0.doc
CT Implementing Rules development 2007-05-15 Page 10 of 46"

For a roadmap of the work of INSPIRE please see Annex I

4.3 UN-SDI

http://www.ungiwg.org/docs/unsdi/UNSDI_Compendium_13_02_2007.pdf

The United Nations are establishing a United Nations Spatial Data Infrastructure to support the aims and objectives of their work:

"Whether in the quest for peace, advancing the health, education or well being ofchildren, women and the afflicted, or reducing poverty, improving food security, responding to natural disasters, safeguarding the environment or advancing sustainable development, the United Nations' increasingly requires spatially representative information to realize its goals for the benefit of all. These 'geospatial' data contain embedded information about the location, shape and relationships among and between geographic features. Typically, they include topographic and cadastral surveys, satellite imagery and aerial photographs, censuses and household surveys, biological inventories and the like."

Their vision is:

"Spatial data infrastructures provide the institutional and technical foundation of policies, interoperable standards and procedures that enable organizations and technologies to interact in a way that facilitates spatial data discovery, evaluation and applications. Thus, the development of a UNSDI is considered essential for increasing system coherence for the use and exchange of geospatial data and information for UN activities."

The UN-SDI has mainly adopted and adjusted the service specifications and viewpoints from OGC. During the implementation phase they are also developing an open source ISO-based metadata server called "geonetwork". GEOBENE also uses this metadata server and will link into their existing metadata servers and vice versa.

5 The GEO data architecture

In it's ambitious 10year-implem,entation plan ,GEOSS has formulated the aims of data management within the GEOSS architecture:

> In the implementation of GEOSS, increased sharing of methods for modelling and analysis needed to transform data into useful products will be advocated. The implementation of GEOSS will facilitate, within 6 years, data-management approaches that encompass a broad perspective of the observation-data life cycle, from input through processing, archiving, and dissemination, including reprocessing, analysis and visualization of large volumes and diverse types of data. The implementation of GEOSS will establish, within 6 years, international information sharing and dissemination drawing on existing capabilities through appropriate technologies, including, but not limited to, Internet-based services.
> *GEOSS 10-Year Implementation Plan, Section 5.1&5.2*

From http://www.earthobservations.org/progress/transverse_areas/data_management.html:

The GEO Secretariat Expert responsible for Data Management is Michael Rast.

Data management tasks will focus on the following:

- Initiating steps for promoting the agreed GEO data sharing principles
- Developing GEO data quality assurance strategy
- Supporting the development and use of emerging assimilation and modelling techniques for new applications
- Identifying and improving the access to common data across GEOSS societal benefit areas
- Developing common data access tools, portals and best practices for users across societal benefit areas"

In the work plan for 2007 to 2009 [URL2007, 4] GEOSS has assigned a set of tasks to various member institutions to cover progress in the fields of data sharing principles, metadata, quality assurance, global land cover etc. For details see the work plan.

The Group on Earth observation is currently analysing their situation and will also adopt ISO TC/211 and OGC standards.

Geobene will focus on GEOs data management strategies and it's impact to data providers worldwide in the Deliverable D8.

6 The Geo-BENE data infrastructure

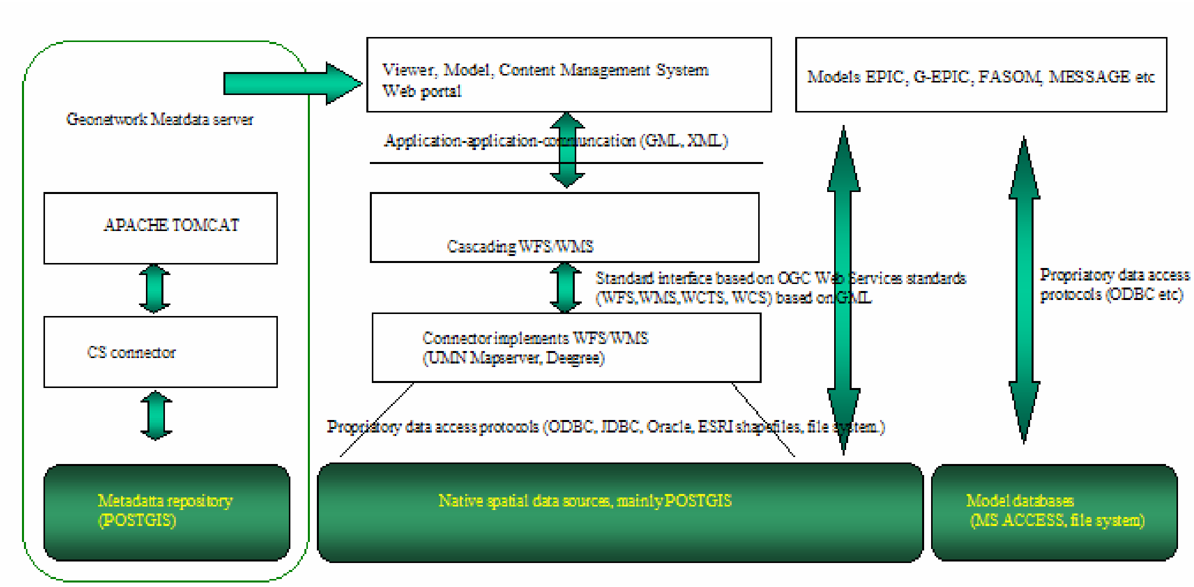
The GEOBENE data infrastructure serves both the needs of modellers as well as the needs of data clients. Whilst modellers are in need of specific data models for the use within their models (cf. EPIC, G-EPIC and the like) the data clients will be served via OGC interfaces, especially the WMS and WFS where approbriate.
Metadata based on ISO19115 are collected and published within an OGC Catalogue service that can be connected to remotely from other metadata servers.

6.1 Detailed description of data strategy for high data-demanding application - global database for the EPIC modelling

The generic models that are used for simulation runs within this project like EPIC, G-EPIC and FASOM mainly use their native data sources. For a thorough discussion of EPICs datamodel please see ANNEX II "Detailed description of data strategy for high data-demanding application - global database for the EPIC modelling". There the data structure as well as a common strategy of data retrieval, storage and meta catalogueing is described.

6.2 The Geobene System architecture

The Geobene System architecture comprises of a database layer and a web services layer together with a meta data server based on FAO's geonetwork java servlet:



:

fig. 3 The GEOBENE system architecture

Also included are the model databases that use their native protocol to connect to their data sources (odbc etc). The model results are then stored as either raster or vector data accessible via web services and also are registered in the meta data repository that also offers OGC catalogue services. We plan to publish our metadata server to the FAO and also to the INSPIRE metadata portal at http://geoportal.jrc.it/geoportal/

The central core of our database layer in this architecture is based on the open source database management system postgresql together with it's spatial language extension postgis, making it highly scalable and community driven. The OGC services are also built on open source software: The Web Map Service (WMS) is the well known and widely used Minnesota Mapserver, whereas the Web Feature Service for vector data is based an the "deegree"-OGC-infrastructure, a java servlet to be used within a java servlet container. In our case this is the open source Apache Tomcat webserver.

As central OGC/ISO compliant metadata server we use FAO's open source *geonetwork* service, which is also a java servlet running in the same Apache tomcat instance.

6.3 The Geobene metadata server "Geonetwork"

Geobene's metadata server is based on FAO's open source solution which is available at http://geonetwork-opensource.org/

It aims at making available "standards based, Free and Open Source catalog application to manage spatially referenced resources through the web. It provides powerful metadata editing and search functions as well as an embedded interactive web map viewer."

As already mentioned it consists of a java war file which is easily deployable in a java servlet capable web server, typically Apache Tomcat. However, for easy testing, it also ships with an integrated server *jetty*.
The latest version *geonetwork 2.1* now also supports postgresql. Since we are using postgresql for our web feature services it is straightforward to also use the postgresql database engine also for our metadata catalogue.

Geonetwork is a true Open Geospatial Consortium Catalog Service for the Web (OGC-CSW). It provides remote access via the Z.90 protocol (see also [URL 2007, 7] and [URL 2007, 8], the CSW specifications at opengeospatial's. This allows for cascading catalogue services. See Annex I for a detailed compilation of present geonetwork services.

The Geobene metadata server will consecutively be filled with metadata about our models and available geodata. An example can be found in the following figure:

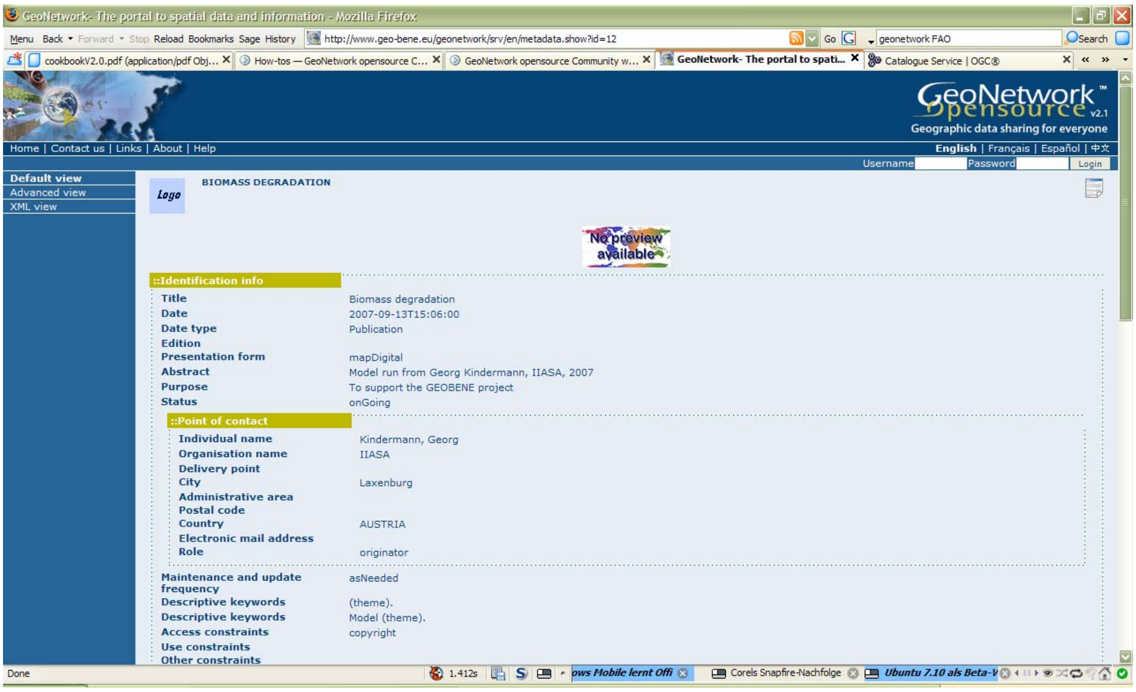


fig. 4 Example aof dataset in Geobene's geonetwork metadata server

6.4 The Geobene Web Map Server

Within our spatial data infrastructure we also employ an OGC compatible Web Map Service (WMS) which is based on the well known Minnesota Mapserver. It runs as a cgi-program within a "normal" apache2.0 installation. In order to keep things together we also use apache's mod_jk module to integrate the java servlet applications like geonetwork and deegree (the web feature service) within one single portal, http://www.geo-bene.eu:

The WMS hosts the different model run results and is capable of dealing with the all major geo-raster data formats by using the open source library "gdal". Dependent on compilation gdal supports a lot of different raster formats, see Annex I for details.

The data can be requested by a request according to the OGC WMS spec:

http://lyra2.felis.uni-freiburg.de/cgi-bin/feliswms?REQUEST=GETMAP&WMTVER=1.0&WIDTH=800&HEIGHT=600&LAYERS=biomassdegradation&BBOX=-180,-90,180,90&SRS=epsg:4326&

A request of this form renders an image of biomassdegradation according to a model that has been calculated by G. Kindermann, from IIASA. Via the WMS interface it is easily possible to cpnnect to this service with different clients – from COTS Geographic Information Systems to open source GIS like gvSIG, Quantum GIS etc. Furthermore NASA's Worldwind, a digital 3D globe, also supports the WMS interface, making it possible to overlay global 3D data with model runs.
The following images shows the aforementioned biomassdegradation model run as an overlay over NASA's "blue marble" data set in WorldWind:

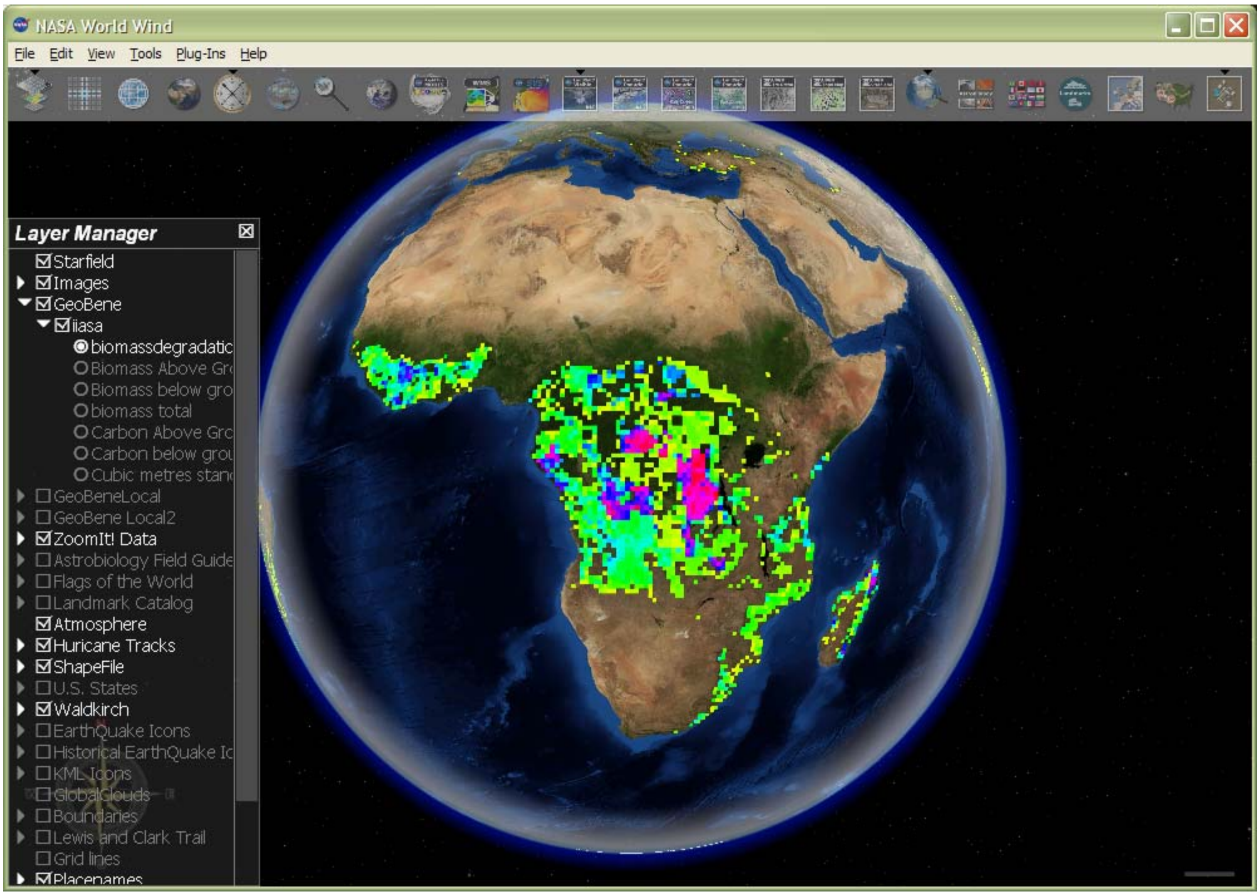


fig. 5 Example of integration of Geobene's WMS layers into a 3D globe

6.5 The Geobene Content Management System

The Geobene Web portal at

http://www.geo-bene.eu

serves as the CPI, the Central Point of Information for all geobene related information. It is meant as cpi as well as a discussion platform between interested researchers via standard web 2.0 applications like forums, blogs etc.
The CMS is based on "drupal" [URL 2007, 12] which is one of the major open source Content Management Systems (php-suite).
By using the "carto"-module .- a small drupal module allowing the integration of WMS into drupal - it is possible to integrate WMS functionality into our portal. The integration is seamless also for less experienced WMS users. The following figure shows an example:

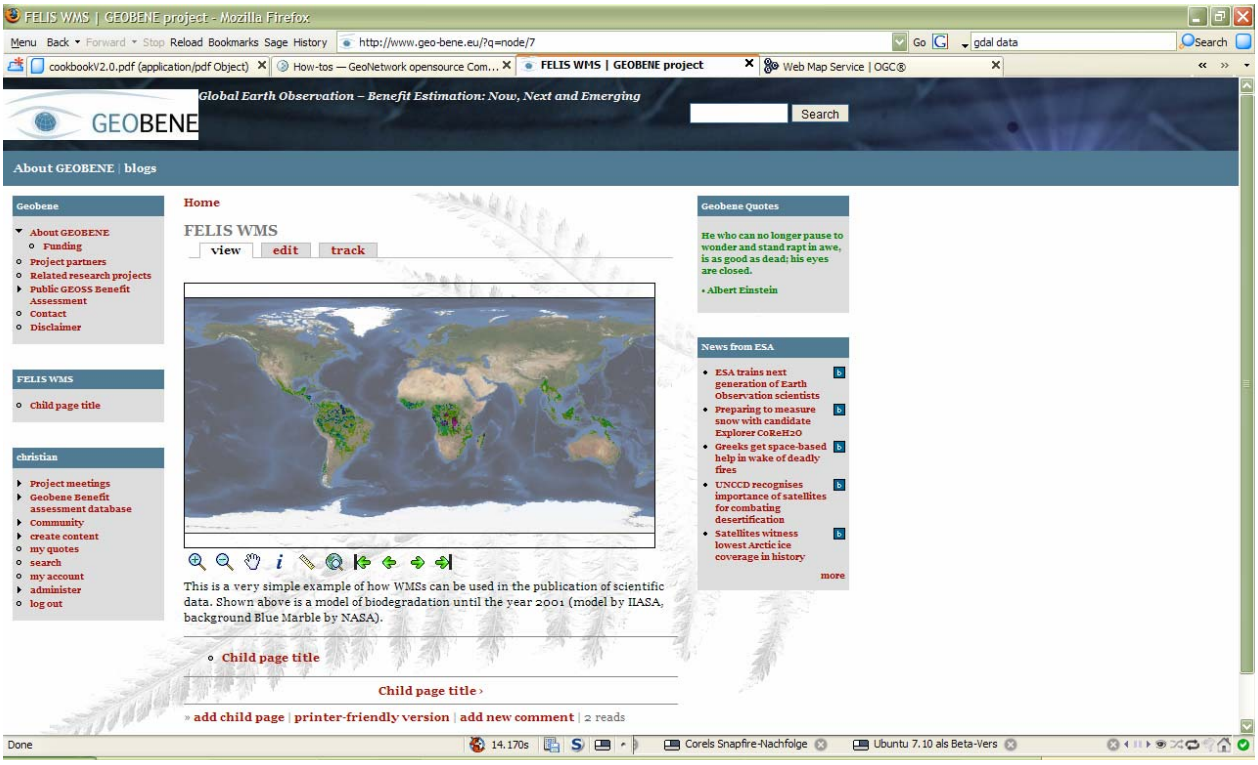


fig. 6 Example of integration of Geobene's WMS layers into the Geobene Web Portal

7 Conclusions

Within Geobene we have analysed the international standards relevant to the project and have developed a system architecture and infrastructure in which to keep our spatial data, register metadata and publish theses metadata to connatural services. We have revised global spatial data initiatives and we are still following the data architecture activities at GEO. Since the discussion is ongoing at GEO we will analyse the GEO data architecture strategy and all it's side linkages with existing initiatives like INSPIRE and UN-SDI within the next deliverable.
Furthermore we have described in detail the data strategy for high data-demanding applications like the global database for the EPIC modelling which is heavily used by our modellers.
Parts of this document are based on earlier works from the authors in different EC projects.
Within in the next project months we will enhance the geodata infrastructure.

8 References

[URL 2007,1] http://www.isotc211.org
[URL 2007, 2] http://www.opengeospatial.org
[URL 2007, 3] http://ec.europa.eu/malta/documents/abc/overviewinspiredirective.pdf
[URL 2007, 4] http://www.earthobservations.org/docs/WP0709.v4.pdf , GEOSSS work plan 2007-2009
[URL 2007, 5] http://www.gsdi.org/docs2004/Cookbook/cookbookV2.0.pdf, The Spatial Data Infrastructure Cookbook
[URL 2007, 7], http://geonetwork-opensource.org/documentation/
[URL 2007, 8], http://portal.opengeospatial.org/files/?artifact_id=20555 , OpenGIS® Catalogue Service Implementation Specification
[URL 2007, 9], http://geonetwork-opensource.org/geonetwork-nodes, collection of geonetwork metadata servers
[URL 2007, 10] http://www.gdal.org, - Geospatial Data Abstraction Library
[URL 2007, 11], http://portal.opengeospatial.org/files/?artifact_id=4756, OGC WMS specification
[URL 2007, 12], http://www.deegree.org/, "deegree - Free Software for Spatial Data Infrastructures"
[URL 2007, 13], http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_tc_browse.htm?commid=54904, ISO's catalogue of available Standards
[URL 2007, 14], http://dublincore.org/documents/dces/ Dublin Core metadata elements
[URL 2007, 15], http://portal.opengeospatial.org/files/?artifact_id=4756, OGC WMS specification
[URL 2007, 16], http://portal.opengeospatial.org/files/?artifact_id=8339, OGC WFS specification
[URL 2007, 17], https://portal.opengeospatial.org/files/?artifact_id=18153, OGC WCS specification
[URL 2007, 18], http://portal.opengeospatial.org/files/?artifact_id=20555 OGC Catalogue Servicves specification
[URL 2007, 19], http://inspire.jrc.it/reports/transposition/INSPIRE_IR_WP2007_2009_en.pdf, INSPIRE Work Programme, Transposition Phase 2007-2009
[Charvat et al, 2005], Charvat, CCSS, FELIS, IGN, MOVIQUITY, JR, IMCS, D2.1, Report on the elaborated service standards and recommendations for the NaturNet-Redime system. NaturNet-Redime project within the 6th framework.
[Kresse, Fadaie, 2004] ISO Standards for Geographic Information, Springer, ISBN 3-540-20130-0
[OGC 2000], The OpenGIS™ Abstract Specification, Topic 11: OpenGIS(tm) Metadata, (ISO/TC 211 DIS 19115)

9 List of figures

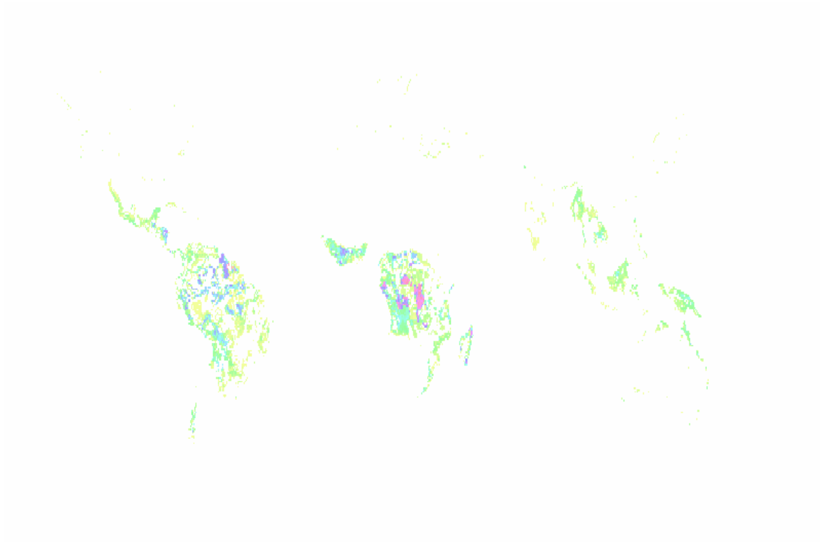
10 Annex I

10.1 Examples for OGC Standards

Example for Geobene's Web Map Service:

http://lyra2.felis.uni-freiburg.de/cgi-bin/feliswms?REQUEST=GETMAP&WMTVER=1.0&WIDTH=800&HEIGHT=600&LAYERS=biomassdegradation&BBOX=-180,-90,180,90&SRS=epsg:4326&

This request renders an image of the form



This example shows a model of biomassdegradation

Example of parts of WFS GetCapabilities Response

This example shows a fractional part of the Web Feature Server of the Dep- of Remote Sensing and Landscape Information Systems (FELIS) of the University of Freiburg. It also shows the filter capabilities of the server. The server is a "deegree" java servlet [URL 2007, 12]

```
<wfs:FeatureTypeList>
-
      <wfs:FeatureType>
<wfs:Name>app:Springs</wfs:Name>
<wfs:Title>Spring in Utah</wfs:Title>
<wfs:Abstract>All Spring in Utah</wfs:Abstract>
-
      <ows:Keywords>
<ows:Keyword>Springs</ows:Keyword>
</ows:Keywords>
<wfs:DefaultSRS>EPSG:26912</wfs:DefaultSRS>
<wfs:OtherSRS>EPSG:4326</wfs:OtherSRS>
-
      <wfs:OutputFormats>
<wfs:Format>text/xml; subtype=gml/3.1.1</wfs:Format>
</wfs:OutputFormats>
-
```

```
      <ows:WGS84BoundingBox>
<ows:LowerCorner>-180.0 -90.0</ows:LowerCorner>
<ows:UpperCorner>180.0 90.0</ows:UpperCorner>
</ows:WGS84BoundingBox>
      <wfs:FeatureType>
<wfs:Name>app:comlV09</wfs:Name>
<wfs:Title/>
<wfs:DefaultSRS>EPSG:4326</wfs:DefaultSRS>
-
      <wfs:Operations>
<wfs:Operation>Query</wfs:Operation>
</wfs:Operations>
-
      <wfs:OutputFormats>
<wfs:Format>text/xml; subtype=gml/3.1.1</wfs:Format>
</wfs:OutputFormats>
-
      <ows:WGS84BoundingBox>
<ows:LowerCorner>-180.0 -90.0</ows:LowerCorner>
<ows:UpperCorner>180.0 90.0</ows:UpperCorner>
</ows:WGS84BoundingBox>
</wfs:FeatureType>
      <ogc:Filter_Capabilities>
      <ogc:Spatial_Capabilities>
      <ogc:GeometryOperands>
<ogc:GeometryOperand>gml:Envelope</ogc:GeometryOperand>
<ogc:GeometryOperand>gml:Point</ogc:GeometryOperand>
<ogc:GeometryOperand>gml:LineString</ogc:GeometryOperand>
<ogc:GeometryOperand>gml:Polygon</ogc:GeometryOperand>
</ogc:GeometryOperands>
-
      <ogc:SpatialOperators>
<ogc:SpatialOperator name="Within"/>
<ogc:SpatialOperator name="Intersects"/>
<ogc:SpatialOperator name="Overlaps"/>
<ogc:SpatialOperator name="BBOX"/>
<ogc:SpatialOperator name="Crosses"/>
<ogc:SpatialOperator name="Contains"/>
<ogc:SpatialOperator name="Disjoint"/>
<ogc:SpatialOperator name="Beyond"/>
<ogc:SpatialOperator name="Equals"/>
<ogc:SpatialOperator name="Touches"/>
</ogc:SpatialOperators>
</ogc:Spatial_Capabilities>
-
      <ogc:Scalar_Capabilities>
<ogc:LogicalOperators/>
-
      <ogc:ComparisonOperators>
<ogc:ComparisonOperator>LessThanEqualTo</ogc:ComparisonOperator>
<ogc:ComparisonOperator>Between</ogc:ComparisonOperator>
<ogc:ComparisonOperator>EqualTo</ogc:ComparisonOperator>
<ogc:ComparisonOperator>GreaterThanEqualTo</ogc:ComparisonOperator>
<ogc:ComparisonOperator>Like</ogc:ComparisonOperator>
<ogc:ComparisonOperator>GreaterThan</ogc:ComparisonOperator>
<ogc:ComparisonOperator>NullCheck</ogc:ComparisonOperator>
<ogc:ComparisonOperator>LessThan</ogc:ComparisonOperator>
</ogc:ComparisonOperators>
-
      <ogc:ArithmeticOperators>
<ogc:SimpleArithmetic/>
</ogc:ArithmeticOperators>
</ogc:Scalar_Capabilities>
-
```

```
      <ogc:Id_Capabilities>
<ogc:EID/>
<ogc:FID/>
</ogc:Id_Capabilities>
</ogc:Filter_Capabilities>
</wfs:WFS_Capabilities>
```

Example for GML

Example for a PointType:

```
<complexType name="PointType">
 <complexContent>
  <extension base="gml:AbstractGeometricPrimitiveType">
   <sequence>
    <choice>
     <element ref="gml:pos" />
     <element ref="gml:coordinates" />
     <element ref="gml:coord" />
    </choice>
   </sequence>
  </extension>
 </complexContent>
</complexType>
```

Example for ISO19115-Metadata in XML format (according to ISO 19139)

This example shows the result of a query in Geobene’s metadata server:


```
<?xml version="1.0" encoding="UTF-8"?>
<gmd:MD_Metadata xmlns:gmd="http://www.isotc211.org/2005/gmd"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:gml="http://www.opengis.net/gml"
xmlns:gts="http://www.isotc211.org/2005/gts" xmlns:gco="http://www.isotc211.org/2005/gco">
        <gmd:fileIdentifier>
        <gco:CharacterString>8d362098-202e-4d45-a9b2-fb34374dc852</gco:CharacterString>
        </gmd:fileIdentifier>
        <gmd:language>
        <gco:CharacterString>eng</gco:CharacterString>
        </gmd:language>
        <gmd:characterSet>
        <gmd:MD_CharacterSetCode                                codeListValue="utf8"
        codeList="./resources/codeList.xml#MD_CharacterSetCode"/>
        </gmd:characterSet>
        <gmd:contact>
        <gmd:CI_ResponsibleParty>
        <gmd:individualName>
        <gco:CharacterString>Christian Schill</gco:CharacterString>
        </gmd:individualName>
        <gmd:organisationName>
        <gco:CharacterString>FELIS University of Freiburg</gco:CharacterString>
        </gmd:organisationName>
        <gmd:positionName>
        <gco:CharacterString/>
        </gmd:positionName>
        <gmd:contactInfo>
        <gmd:CI_Contact>
        <gmd:phone>
        <gmd:CI_Telephone>
        <gmd:voice>
        <gco:CharacterString/>
        </gmd:voice>
        <gmd:facsimile>
        <gco:CharacterString/>
        </gmd:facsimile>
        </gmd:CI_Telephone>
        </gmd:phone>
        <gmd:address>
```

```
<gmd:CI_Address>
<gmd:deliveryPoint>
<gco:CharacterString/>
</gmd:deliveryPoint>
<gmd:city>
<gco:CharacterString/>
</gmd:city>
<gmd:administrativeArea>
<gco:CharacterString/>
</gmd:administrativeArea>
<gmd:postalCode>
<gco:CharacterString>79106</gco:CharacterString>
</gmd:postalCode>
<gmd:country>
<gco:CharacterString>GERMANY</gco:CharacterString>
</gmd:country>
<gmd:electronicMailAddress>
<gco:CharacterString>christian.schill@felis.uni-freiburg.de</gco:CharacterString>
</gmd:electronicMailAddress>
</gmd:CI_Address>
</gmd:address>
</gmd:CI_Contact>
</gmd:contactInfo>
<gmd:role>
<gmd:CI_RoleCode codeList="./resources/codeList.xml#CI_RoleCode"
codeListValue="pointOfContact"/>
</gmd:role>
</gmd:CI_ResponsibleParty>
</gmd:contact>
<gmd:dateStamp>
<gco:DateTime>2007-09-18T16:16:55</gco:DateTime>
</gmd:dateStamp>
<gmd:metadataStandardName>
<gco:CharacterString>ISO 19115:2003/19139</gco:CharacterString>
</gmd:metadataStandardName>
<gmd:metadataStandardVersion>
<gco:CharacterString>1.0</gco:CharacterString>
</gmd:metadataStandardVersion>
<gmd:referenceSystemInfo>
<gmd:MD_ReferenceSystem>
<gmd:referenceSystemIdentifier>
<gmd:RS_Identifier>
<gmd:code>
<gco:CharacterString>WGS 1984</gco:CharacterString>
</gmd:code>
</gmd:RS_Identifier>
</gmd:referenceSystemIdentifier>
</gmd:MD_ReferenceSystem>
</gmd:referenceSystemInfo>
<gmd:identificationInfo>
<gmd:MD_DataIdentification>
<gmd:citation>
<gmd:CI_Citation>
<gmd:title>
<gco:CharacterString>Biomass degradation</gco:CharacterString>
</gmd:title>
<gmd:date>
<gmd:CI_Date>
<gmd:date>
<gco:DateTime>2007-09-13T15:06:00</gco:DateTime>
</gmd:date>
<gmd:dateType>
<gmd:CI_DateTypeCode codeList="./resources/codeList.xml#CI_DateTypeCode"
codeListValue="publication"/>
</gmd:dateType>
</gmd:CI_Date>
</gmd:date>
<gmd:edition>
<gco:CharacterString/>
</gmd:edition>
<gmd:presentationForm>
<gmd:CI_PresentationFormCode codeList="./resources/codeList.xml#CI_PresentationFormCode"
codeListValue="mapDigital"/>
</gmd:presentationForm>
</gmd:CI_Citation>
</gmd:citation>
<gmd:abstract>
<gco:CharacterString>Model run from Georg Kindermann, IIASA, 2007</gco:CharacterString>
</gmd:abstract>
<gmd:purpose>
```

```xml
<gco:CharacterString>To support the GEOBENE project</gco:CharacterString>
</gmd:purpose>
<gmd:status>
<gmd:MD_ProgressCode codeList="./resources/codeList.xml#MD_ProgressCode"
codeListValue="onGoing"/>
</gmd:status>
<gmd:pointOfContact>
<gmd:CI_ResponsibleParty>
<gmd:individualName>
<gco:CharacterString>Kindermann, Georg</gco:CharacterString>
</gmd:individualName>
<gmd:organisationName>
<gco:CharacterString>IIASA</gco:CharacterString>
</gmd:organisationName>
<gmd:contactInfo>
<gmd:CI_Contact>
<gmd:phone>
<gmd:CI_Telephone>
<gmd:voice>
<gco:CharacterString/>
</gmd:voice>
<gmd:facsimile>
<gco:CharacterString/>
</gmd:facsimile>
</gmd:CI_Telephone>
</gmd:phone>
<gmd:address>
<gmd:CI_Address>
<gmd:deliveryPoint>
<gco:CharacterString/>
</gmd:deliveryPoint>
<gmd:city>
<gco:CharacterString>Laxenburg</gco:CharacterString>
</gmd:city>
<gmd:administrativeArea>
<gco:CharacterString/>
</gmd:administrativeArea>
<gmd:postalCode>
<gco:CharacterString/>
</gmd:postalCode>
<gmd:country>
<gco:CharacterString>AUSTRIA</gco:CharacterString>
</gmd:country>
<gmd:electronicMailAddress>
<gco:CharacterString/>
</gmd:electronicMailAddress>
</gmd:CI_Address>
</gmd:address>
</gmd:CI_Contact>
</gmd:contactInfo>
<gmd:role>
<gmd:CI_RoleCode codeList="./resources/codeList.xml#CI_RoleCode"
codeListValue="originator"/>
</gmd:role>
</gmd:CI_ResponsibleParty>
</gmd:pointOfContact>
<gmd:resourceMaintenance>
<gmd:MD_MaintenanceInformation>
<gmd:maintenanceAndUpdateFrequency>
<gmd:MD_MaintenanceFrequencyCode
codeList="./resources/codeList.xml#MD_MaintenanceFrequencyCode"
codeListValue="asNeeded"/>
</gmd:maintenanceAndUpdateFrequency>
</gmd:MD_MaintenanceInformation>
</gmd:resourceMaintenance>
<gmd:graphicOverview>
<gmd:MD_BrowseGraphic>
<gmd:fileName>
<gco:CharacterString/>
</gmd:fileName>
<gmd:fileDescription>
<gco:CharacterString>thumbnail</gco:CharacterString>
</gmd:fileDescription>
</gmd:MD_BrowseGraphic>
</gmd:graphicOverview>
<gmd:graphicOverview>
<gmd:MD_BrowseGraphic>
<gmd:fileName>
<gco:CharacterString/>
</gmd:fileName>
```

```
<gmd:fileDescription>
<gco:CharacterString>large_thumbnail</gco:CharacterString>
</gmd:fileDescription>
</gmd:MD_BrowseGraphic>
</gmd:graphicOverview>
<gmd:descriptiveKeywords>
<gmd:MD_Keywords>
<gmd:keyword>
<gco:CharacterString/>
</gmd:keyword>
<gmd:type>
<gmd:MD_KeywordTypeCode codeList="./resources/codeList.xml#MD_KeywordTypeCode"
codeListValue="theme"/>
</gmd:type>
</gmd:MD_Keywords>
</gmd:descriptiveKeywords>
<gmd:descriptiveKeywords>
<gmd:MD_Keywords>
<gmd:keyword>
<gco:CharacterString>Model</gco:CharacterString>
</gmd:keyword>
<gmd:type>
<gmd:MD_KeywordTypeCode codeList="./resources/codeList.xml#MD_KeywordTypeCode"
codeListValue="theme"/>
</gmd:type>
</gmd:MD_Keywords>
</gmd:descriptiveKeywords>
<gmd:resourceConstraints>
<gmd:MD_LegalConstraints>
<gmd:accessConstraints>
<gmd:MD_RestrictionCode codeList="./resources/codeList.xml#MD_RestrictionCode"
codeListValue="copyright"/>
</gmd:accessConstraints>
<gmd:useConstraints>
<gmd:MD_RestrictionCode codeList="./resources/codeList.xml#MD_RestrictionCode"
codeListValue=""/>
</gmd:useConstraints>
<gmd:otherConstraints>
<gco:CharacterString/>
</gmd:otherConstraints>
</gmd:MD_LegalConstraints>
</gmd:resourceConstraints>
<gmd:spatialRepresentationType>
<gmd:MD_SpatialRepresentationTypeCode
codeList="./resources/codeList.xml#MD_SpatialRepresentationTypeCode"
codeListValue="vector"/>
</gmd:spatialRepresentationType>
<gmd:spatialResolution>
<gmd:MD_Resolution>
<gmd:equivalentScale>
<gmd:MD_RepresentativeFraction>
<gmd:denominator>
<gco:Integer/>
</gmd:denominator>
</gmd:MD_RepresentativeFraction>
</gmd:equivalentScale>
</gmd:MD_Resolution>
</gmd:spatialResolution>
<gmd:language>
<gco:CharacterString>eng</gco:CharacterString>
</gmd:language>
<gmd:characterSet>
<gmd:MD_CharacterSetCode codeList="./resources/codeList.xml#MD_CharacterSetCode"
codeListValue="utf8"/>
</gmd:characterSet>
<gmd:topicCategory>
<gmd:MD_TopicCategoryCode>boundaries</gmd:MD_TopicCategoryCode>
</gmd:topicCategory>
<gmd:extent>
<gmd:EX_Extent>
<gmd:temporalElement>
<gmd:EX_TemporalExtent>
<gmd:extent>
<gml:TimePeriod>
<gml:beginPosition/>
<gml:endPosition/>
</gml:TimePeriod>
</gmd:extent>
</gmd:EX_TemporalExtent>
</gmd:temporalElement>
```

```
</gmd:EX_Extent>
</gmd:extent>
<gmd:extent>
<gmd:EX_Extent>
<gmd:geographicElement>
<gmd:EX_GeographicBoundingBox>
<gmd:westBoundLongitude>
<gco:Decimal>-180</gco:Decimal>
</gmd:westBoundLongitude>
<gmd:eastBoundLongitude>
<gco:Decimal>180</gco:Decimal>
</gmd:eastBoundLongitude>
<gmd:southBoundLatitude>
<gco:Decimal>-90</gco:Decimal>
</gmd:southBoundLatitude>
<gmd:northBoundLatitude>
<gco:Decimal>90</gco:Decimal>
</gmd:northBoundLatitude>
</gmd:EX_GeographicBoundingBox>
</gmd:geographicElement>
</gmd:EX_Extent>
</gmd:extent>
<gmd:supplementalInformation>
<gco:CharacterString></gco:CharacterString>
</gmd:supplementalInformation>
</gmd:MD_DataIdentification>
</gmd:identificationInfo>
<gmd:distributionInfo>
<gmd:MD_Distribution>
<gmd:transferOptions>
<gmd:MD_DigitalTransferOptions>
<gmd:onLine>
<gmd:CI_OnlineResource>
<gmd:linkage>
<gmd:URL>http://lyra2.felis.uni-freiburg.de/cgi-bin/feliswms</gmd:URL>
</gmd:linkage>
<gmd:protocol>
<gco:CharacterString>OGC:WMS-1.1.1-http-get-map</gco:CharacterString>
</gmd:protocol>
<gmd:name>
<gco:CharacterString>FELIS Web Map Server</gco:CharacterString>
</gmd:name>
<gmd:description>
<gco:CharacterString/>
</gmd:description>
</gmd:CI_OnlineResource>
</gmd:onLine>
<gmd:onLine>
<gmd:CI_OnlineResource>
<gmd:linkage>
<gmd:URL>http://localhost:8080/geonetwork/srv/en/resources.get?id=12&fname=&access=privat
e</gmd:URL>
</gmd:linkage>
<gmd:protocol>
<gco:CharacterString>WWW:DOWNLOAD-1.0-http--download</gco:CharacterString>
</gmd:protocol>
<gmd:name>
<gco:CharacterString/>
</gmd:name>
<gmd:description>
<gco:CharacterString/>
</gmd:description>
</gmd:CI_OnlineResource>
</gmd:onLine>
<gmd:onLine>
<gmd:CI_OnlineResource>
<gmd:linkage>
<gmd:URL/>
</gmd:linkage>
<gmd:protocol>
<gco:CharacterString>OGC:WMS-1.1.1-http-get-map</gco:CharacterString>
</gmd:protocol>
<gmd:name>
<gco:CharacterString/>
</gmd:name>
<gmd:description>
<gco:CharacterString/>
</gmd:description>
</gmd:CI_OnlineResource>
</gmd:onLine>
```

```
        </gmd:MD_DigitalTransferOptions>
        </gmd:transferOptions>
        </gmd:MD_Distribution>
        </gmd:distributionInfo>
        <gmd:dataQualityInfo>
        <gmd:DQ_DataQuality>
        <gmd:scope>
        <gmd:DQ_Scope>
        <gmd:level>
        <gmd:MD_ScopeCode                     codeList="./resources/codeList.xml#MD_ScopeCode"
        codeListValue="dataset"/>
        </gmd:level>
        </gmd:DQ_Scope>
        </gmd:scope>
        <gmd:lineage>
        <gmd:LI_Lineage>
        <gmd:statement>
        <gco:CharacterString/>
        </gmd:statement>
        </gmd:LI_Lineage>
        </gmd:lineage>
        </gmd:DQ_DataQuality>
        </gmd:dataQualityInfo>
</gmd:MD_Metadata>
```

10.2 Overview of registered Geonetwork metadata servers

See [URL 2007, 9] for details

OWNER	WEB LINK	DESCRIPTION
FAO Headquarters	http://www.fao.org/geonetwork	FAO-GeoNetwork
UNEP Headquarters	http://www.ecomundus.net/	ecoMundus - Network for Environmental Information and Data
WFP VAM-SIE Headquarters	http://vam.wfp.org/geonetwork	GeoNetwork SIE Headquarters
WFP Latin America – Regional Bureau	http://201.224.73.132	Wold Food Programme Vulnerability Analysis and Mapping VAM SIE ODPC Operations Department City
WFP West Africa - Regional Bureau	http://213.154.77.158/geonetwork	Wold Food Programme Vulnerability Analysis and Mapping VAM SIE ODD Operations Department Dakar
WFP Middle East and North Africa Regional Bureau	http://vamodc.wfp.org/geonetwork	Wold Food Programme Vulnerability Analysis and Mapping VAM SIE ODC Operations Department Cairo
WFP Eastern and Central Africa Regional Bureau	http://193.108.214.8/geonetwork	Wold Food Programme Vulnerability Analysis and Mapping VAM SIE ODC Operations Department Kampala
WFP Southern Africa Regional Bureau	http://196.36.132.196/geonetwork	Wold Food Programme Vulnerability Analysis and Mapping VAM SIE ODJ Operations Department Johannesburg
WFP Asia Regional Bureau	http://203.146.113.37/geonetwork	Wold Food Programme Vulnerability Analysis and Mapping ODB SIE Operations Department Bangkok
WFP Sudan Country Office	http://212.0.146.203/geonetwork	Wold Food Programme Vulnerability Analysis and Mapping SIE Khartoum
WFP Ethiopia Country Office	http://10.11.157.8/geonetwork	Wold Food Programme Vulnerability Analysis and Mapping SIE
WFP Afghanistan Country Office	http://82.205.204.77/geonetwork	Wold Food Programme Vulnerability Analysis and Mapping SIE
WHO Headquarters	http://www.who.int/geonetwork	World Health Organization GIS Resources
OCHA Headquarters	http://geonetwork.unocha.org/mapsondemand/	United Nations Office for the Coordination of Humanitarian Affaires
CGIAR-CSI	http://geonetwork.csi.cgiar.org/geonetwork	CGIAR-CSI - Consortium for Spatial Information - Main node
CGIAR-CSI IWMI	http://geonetwork.iwmi.org/geonetwork	CGIAR-CSI - Consortium for Spatial Information IWMI - International Water Management Institute
CGIAR-CSI CIAT	http://gisweb.ciat.cgiar.org:8080/geonetwork pending installation	CGIAR-CSI - Consortium for Spatial Information Centro CIAT - Internacional de Agricultura Tropical

CGIAR-CSI Tsunami	http://tsdc.iwmi.org/geonetwork	CGIAR-CSI - Consortium for Spatial Information
CGIAR-CSI CIP	http://tsdc.iwmi.org/geonetworkcip	CGIAR-CSI - Consortium for Spatial Information
CGIAR-CSI ILRI	http://tsdc.iwmi.org/geonetworkilri	CGIAR-CSI - Consortium for Spatial Information
CGIAR-CSI CIMMYT	http://tsdc.iwmi.org/geonetworkcimmyt/	CGIAR-CSI - Consortium for Spatial Information
CGIAR-CSI IRRI	http://tsdc.iwmi.org/geonetworkirri/ pending installation	CGIAR-CSI - Consortium for Spatial Information
CGIAR-CSI CIFOR	http://webdev.cifor.cgiar.org:8080/geonetwork It is not working at the moment	CGIAR-CSI - Consortium for Spatial Information
SETSAN	http://www.setsan.org.mz/geonetwork	SETSAN Technical Secretariat for Food Security and Nutrition
SANDRE	http://sandre.eaufrance.fr/geonetwork/srv/fr/main.home	Le Service d'Administration Nationale des Données et Référentiels sur l'Eau
ICIMOD	http://arcsde.icimod.org.np:8080/geonetwork/	International Centre for Integrated Mountain Development (ICIMOD) Kathmandu,
MMA	http://mapas.mma.gov.br/geonetwork/	Ministerio do Meio Ambiente, Brasil Geo processamento
SIGLAB INPA	http://200.17.53.49:8080/geonetwork	Laboratório de Análise e Modelagem Espacial do INPA – SIGLAB INPA Instituto National de Pesquisas da Amazônia -Brasil
RCMRD	http://www.rcmrd.org/geonetwork	Regional Center for Mapping of Resources for Development
AGHRYMET	http://geonetwork.agrhymet.ne	Centre Regional Aghrymet
SADC	http://www.sadc.int/geonetwork	Southern African Development Community
SWALIM	http://geonetwork.faoswalim.org:8080/geonetwork	Somalia Water and Land Information Management
The NCO of UNSDI	http://www.unsdi.nl/index.php	The Netherlands Coordination Office of UNSDI

10.3 Web Map Service supported raster formats

Our WMS uses GDAL, the - Geospatial Data Abstraction Library. See [URL 2007, 10] for details

Depending on compilation it supports the following formats:
GRASS (ro): GRASS Database Rasters (5.7+)
VRT (rw+): Virtual Raster
GTiff (rw+): GeoTIFF
NITF (rw+): National Imagery Transmission Format
HFA (rw+): Erdas Imagine Images (.img)
SAR_CEOS (ro): CEOS SAR Image
CEOS (ro): CEOS Image
ELAS (rw+): ELAS
AIG (ro): Arc/Info Binary Grid
AAIGrid (rw): Arc/Info ASCII Grid
SDTS (ro): SDTS Raster
DTED (rw): DTED Elevation Raster
PNG (rw): Portable Network Graphics
JPEG (rw): JPEG JFIF
MEM (rw+): In Memory Raster
JDEM (ro): Japanese DEM (.mem)
GIF (rw): Graphics Interchange Format (.gif)
ESAT (ro): Envisat Image Format
BSB (ro): Maptech BSB Nautical Charts
XPM (rw): X11 PixMap Format
BMP (rw+): MS Windows Device Independent Bitmap
AirSAR (ro): AirSAR Polarimetric Image
RS2 (ro): RadarSat 2 XML Product
PCIDSK (rw+): PCIDSK Database File
PCRaster (rw): PCRaster Raster File
ILWIS (rw+): ILWIS Raster Map
RIK (ro): Swedish Grid RIK (.rik)
PNM (rw+): Portable Pixmap Format (netpbm)
DOQ1 (ro): USGS DOQ (Old Style)
DOQ2 (ro): USGS DOQ (New Style)
ENVI (rw+): ENVI .hdr Labelled
EHdr (rw+): ESRI .hdr Labelled
PAux (rw+): PCI .aux Labelled
MFF (rw+): Atlantis MFF Raster
MFF2 (rw+): Atlantis MFF2 (HKV) Raster
FujiBAS (ro): Fuji BAS Scanner Image
GSC (ro): GSC Geogrid
FAST (ro): EOSAT FAST Format
BT (rw+): VTP .bt (Binary Terrain) 1.3 Format
LAN (ro): Erdas .LAN/.GIS
CPG (ro): Convair PolGASP
IDA (rw+): Image Data and Analysis
NDF (ro): NLAPS Data Format
ECW (rw): ERMapper Compressed Wavelets

JP2ECW (rw+): ERMapper JPEG2000
L1B (ro): NOAA Polar Orbiter Level 1b Data Set
FIT (rw): FIT Image
RMF (rw+): Raster Matrix Format
USGSDEM (rw): USGS Optional ASCII DEM (and CDED)
GXF (ro): GeoSoft Grid Exchange Format

10.4 Roadmap for implementing rules of the INSPIRE Directive

see [URL 2007, 19] for details

The table below provides the roadmap for the development and implementation of the Implementing Rules. The first column gives the date of the milestone. The second column indicates the article of the INSPIRE Directive that is relevant for the milestone.

Milestone date Relevant
article Description
2007-05-15 Entry into force of INSPIRE Directive
2007-08-15 22§2 Establishment of the INSPIRE Committee
2008-05-15 5§4 Adoption of IR for the creation and updating of metadata
2008-05-15 21(4) Adoption of IR for monitoring and reporting
2008-05-15* 16 Adoption of IR for discovery and view services
2008-11-15* 16 Adoption of IR for download services
2008-11-15* DS-D2.7 Adoption of IR for data exchange
2008-11-15* 16(a) Adoption of IR for Coordinates Transformation Service
2009-05-15* 17(8) Adoption of IR governing the access rights of use to spatial data sets and services for Community institutions and bodies
2009-05-15 9(a) Adoption of IRs for the interoperability and harmonisation of spatial data sets and services for Annex I spatial data themes
2009-05-15 24§1 Provisions of Directive are brought into force in MS (transposition date)
2010-05-15 21§1 Implementation of provisions for monitoring
2010-05-15 6(a) Metadata available for spatial data corresponding to Annex I and Annex II spatial data themes
2010-05-15* 16 Discovery and view network services operational
2010-05-15 15 The EC establishes and runs a geo-portal and Community level
2010-05-15 21§2 Member States' first report to the Commission. From then onwards Member States have to present reports every 3 years
2010-11-15* 16 Download services operational
2010-11-15* 16(a) Coordinates Transformation Services operational
2010-11-15* 16 Adoption of IR for invoke "spatial data service" Network Service
2011-05-15 7§3, 9(a)
Newly collected and extensively restructured spatial data sets available in accordance with the IRs for interoperability and harmonisation of spatial data sets and services for Annex I spatial data themes
2012-05-15 9(b)
Adoption of the IR s for the interoperability and harmonisation of spatial data sets and services for Annex II and III spatial data themes
Infrastructure for Spatial Information in Europe Reference: INSPIRE IR WP2007-2009-v1 0.doc
CT Implementing Rules development 2007-05-15 Page 11 of 46
Milestone date Relevant
article Description
2012-11-15* 16 Invoke "spatial data service" service operational
2013-05-15 21§2 Member States' second report to the Commission
2013-05-15 6(b) Metadata available for spatial data corresponding to Annex III spatial data themes
2014-05-15 23 Commission's report to the EP and the Council. From then onwards the Commission has to present reports every 6 years.
2014-05-15 7§3, 9(b)
Newly collected and extensively restructured spatial data sets available in accordance with IRs for interoperability and harmonisation of spatial data sets and services for Annex II and III spatial data themes
2016-05-15 7§3, 9(a)

Other spatial data sets available in accordance with IR s for
interoperability and harmonisation of spatial data sets and
services for Annex I spatial data themes
2019-05-15 7§3, 9(b)
Other spatial data sets available in accordance with IR s for
interoperability and harmonisation of spatial data sets and
services for Annex II and III spatial data themes
2020-05-15 23 Commission's second report to the EP and the Council.
* = date proposed by Commission

ANNEX II

10.5 *“Detailed description of data strategy for high data-demanding application - global database for the EPIC modelling”*

Paper header:

Detailed description of the data strategy for global EPIC modeling

Rastislav Skalský[1], Erwin Schmid[2], Zuzana Tarasovičová[1] & Juraj Balkovič[1]

[1] Soil Science and Conservation Research institute, Bratislava, Slovakia
[2] University of Natural Resources and Applied Life Sciences, Vienna, Austria

Outline

A pilot geographical database of data inputs for global EPIC applications is introduced. General data strategy for the high-data demanding application is addressed and results of the database design and testing and data processing and integration in pilot area are briefly described in the article (spatial data representation and visualization frame and data set on topography, soil land cover/land use). General aim of the article is to justify the data interpretation approach and the database structure (Chapter 1 and 2.1) as well as to provide the sound basis for the detailed metadata description of geographical database for global agro-ecosystem modeling (Chapter 2.2 and 2.3). The article is open-structured and various refinements in number and content of paragraphs or chapters can be done following the coming advances in data processing.

Detailed description of the data strategy for global EPIC modeling

Rastislav Skalský[1], Erwin Schmid[2], Zuzana Tarasovičová[1] & Juraj Balkovič[1]

[1] *Soil Science and Conservation Research institute, Bratislava, Slovakia*
[2] *University of Natural Resources and Applied Life Sciences, Vienna, Austria*

Outline

A pilot geographical database of data inputs for global EPIC applications is introduced. General data strategy for the high-data demanding application is addressed and results of the database design and testing and data processing and integration in pilot area are briefly described in the article (spatial data representation and visualization frame and data set on topography, soil land cover/land use). General aim of the article is to justify the data interpretation approach and the database structure (Chapter 1 and 2.1) as well as to provide the sound basis for the detailed metadata description of geographical database for global agro-ecosystem modeling (Chapter 2.2 and 2.3). The article is open-structured and various refinements in number and content of paragraphs or chapters can be done following the coming advances in data processing.

List of content

1. Global EPIC modeling

1.1. Modeling of a global agro-ecosystem

The goal of the global agro-ecosystem modeling is to employ the EPIC model to simulate dynamic changes (50 – 100 years) of bio-physical indicators in agro-ecosystems (biomass production, soil organic matter accumulation, soil erosion, greenhouse gasses emission and sequestration – further referred as environmental indicators) for observed and alternative agricultural management systems and under climate change.
The EPIC modeling should bring an accurate view on global agro-ecosystems and to come up with a general model which includes all relevant natural and socio-economic conditions and processes influencing the global agro-ecosystem functions. **The EPIC model and global environmental data fusion** is supposed to yield the bio-physical data for alternative land use and management strategies. The data is used in economic land use optimization models to find the optimal land use and management mixes at global scales.

1.1.1. Model and data

The abstraction and formalization (model) of many bio-physical and –chemical processes important in agro-ecosystem modelling (i.e. **functional aspect** of the agro-ecosystem model) is already implemented in EPIC.
In general, data represents a model of state and/or dynamics of the elements of agro-ecosystem and also vertical and horizontal spatial arrangement or relations among these elements (**structural and qualitative aspect** of the agro-ecosystem model). Regarding this, a dataset can be considered as an abstract space (or frame) within which the functional relationships expressed in a model take a place in a data-model fusion. The situation in the data part of a supposed data model fusion is, however, a bit different when compared to the functional part. Although the data on global agro-ecosystem is more or less available, it is not organized in such a form that it can be immediately used as an input to a selected simulation model and it has to be **interpreted and re-organized** prior to it is applied in simulation modeling.

1.1.2. Database tasks

The most suitable way how to treat such a huge amount of the data needed for global agro-ecosystem modeling is to organize it in the well designed database. The database should follow some general requirements:

- It has to reflect all mandatory data inputs required by the EPIC model and provide a necessary set of attributes on individual qualities of all required elements of the system being simulated (global agro-ecosystem);
- It has to provide an appropriate spatial frame for representing the structural aspect of the agro-ecosystem being modeled and this mostly with regard to those qualities important for the simulation model. A tool for communicating the attribute information and the model results in the geographical context have to be integrated (i.e. geographical frame is required to fully represent the structural aspect of all the EPIC model input/outputs in the global modeling);

There are some additional facts which make above mentioned more particular:

- Because the global agro-ecosystems are the major focus of modeling, the global data availability is the main limiting factor, which also influences the final quality of the global database (overall appropriateness of the data and spatial resolution and attribute depth);
- Above mentioned general assumptions have to be modified with regard to the quality of available data and some kind of compromises must be done to provide a model with as accurate spatial and attribute data as possible (not necessarily the optimal one from a functional model point of view);
- A common interface has to be established to link with the economic land use optimization models.

1.2. EPIC model

1.2.1. General model description

The Erosion Productivity Impact Calculator (EPIC) was developed by a USDA modelling team in the early 80s to asses the status of U.S. soil and water resources (Williams et al., 1984; Williams, 1990; Jones et al., 1991). The first major application of EPIC was to evaluate soil erosion impacts for 135 U.S. land resource regions (Putnam et al., 1988). EPIC compounds various components from CREAMS (Knisel, 1980), SWRRB (Williams et al., 1985), GLEAMS (Leonard et al., 1987), and has been continuously expanded and refined to allow simulation of many processes important in agricultural and forest land management (Sharpley and Williams, 1990; Williams, 1995, Williams et al. 2000). This development resulted in the model name being changed to ***E****nvironmental* ***P****olicy* ***I****ntegrated* ***C****limate* (Williams et al., 1996). A major carbon cycling routine was performed by Izaurralde et al. (2001, 2004) based on the approach used in CENTURY (Parton et al., 1994).
The elemental spatial modelling unit area considered by EPIC is generally a field-size area - up to 100 ha - where weather, soil, topography, and management system are assumed to be homogeneous. The major components in EPIC are weather simulation, hydrology, erosion-sedimentation, nutrient and carbon cycling, pesticide fate, plant growth and competition, soil temperature and moisture, tillage, cost accounting, and plant environment control. EPIC operates on a daily time step, and is capable of simulating hundreds of years if necessary. EPIC can be used to compare management systems and their effects on crop yields, on water, nitrogen, phosphorus, pesticides, organic carbon, and sediment transport, on organic carbon sequestration, and eventually on green house gas emissions. The management components that can be changed are crop rotations, crop/grass mixes, tillage operations, irrigation scheduling, drainage, furrow diking, liming, grazing, burning operations (e.g., on prairies), tree pruning, thinning and harvest, manure handling (e.g., lagoons), and fertilizer and pesticide application rates and timing.

1.2.2. Input data requirements

Like for any biophysical process model, the quality and completeness of input data is of substantial importance. Data on four major input data components (weather, soil, topography and management practices) is essential to run EPIC.

The **weather variables** necessary for running EPIC are precipitation (in mm), minimum and maximum air temperature (in degree Celsius), and solar radiation (in MJ/m^2). If the Penman methods are used to estimate potential evapo-transpiration, wind speed (in m/sec measured at 10 m height), and relative humidity (in %) are also required. If measured daily weather data is available it can be directly input into EPIC. In addition, monthly statistics of this daily weather (mean, standard deviation, skew coefficient, probabilities of wet-dry and wet-wet days, etc.) need to be computed and input in the model. EPIC provides a statistical support program to compute the statistics of relevant variables based on daily weather records. Consequently, long historical daily weather records (20-30 years) for all weather variables are desirable for statistical parameter calculations. Based on the weather variable statistics, EPIC can generate weather patterns for long-run analyses (100+ years), or as indicated above, daily weather records (e.g., from world climate models with downscaling procedures) can be input directly. There is also the option of reading a sequence of actual daily weather and use generated weather afterwards or before within a simulation run.

A large number of physical and chemical **soil parameters** to describe each soil layer and subsequently an entire soil profile can be input in EPIC. These soil parameters (Table 1 in Appendix) are separated between general and layer-specific as well as between essential and useful soil information requirement. The essential soil parameters are mandatory input while the remaining ones would help to further describe a soil specific profile situation. In EPIC, a soil profile can be split in up to 10 soil layers of which each is described by a specific set of chemical and physical soil parameters. If, for instance, the description of only two soil layers is available (top-soil and sub-soil), EPIC still allows (optional) to split the soil profile into ten soil layers. It assures that e.g., soil temperature and soil moisture can be appropriately estimated through soil layers and time.

The **topography** of a field or HRU is described by average field size (ha), slope length (m), and slope (%). In addition, elevation, longitude and latitude are needed for each site or HRU. Information on elevation, longitude and latitude is usually provided by a climate station of a digital elevation map.

Wide range of **management scheduling** in EPIC allows flexibility in modelling different cropping, forestation, and tillage systems (including crop rotations and crop mixes). However, it requires reliable information of the actual management practice for a given region or site. Generally, information on

- Date of planting (including potential heat units the crop needs to reach maturity);
- Date, type (commercial; dairy, swine, etc. manure), and amount of fertilizer (elemental NPK) in kg/ha; if manure is applied, information on application rate (in case of grazing the stocking rate) and nutrient composition (orgN, minN (NO3-N + NH3-N), orgP, minP, minK, orgC, and fraction of NH3-N on minN) is needed;
- Date and amount of irrigation (including NO_3 and salt concentration in irrigation water) in mm;
- Date and type of tillage operation (plough, harrow spike, field cultivator, thinning, etc.); and
- Date and type of harvesting (combine, hay cutting, grazing, etc.) is needed.

2. Geographical database for global EPIC modeling

2.1. Conceptual level of the database proposal

2.1.1. Homogenous response units

The EPIC model is a typical plot-scale model and requires a homogenous set of input data. Therefore, a land unit is characterized by homogenous weather, topography, soil, land use and

management system. Consequently, in characterizing a heterogeneous region, one can derive many individual land units to portray the spatial variability of a landscape.
The global agro-ecosystem modeling (see the paragraph 1.1.) requires a comprehensive geographical database that fulfills certain standards. Moreover, the geographical database serves also as a tool for land cover/land use scenarios implementation. This leads to the definition of the **homogenous response units** (HRU) which represents the conceptual framwork of the geographical database.
The concept of HRU used here was adopted from earlier works (Schmid et al. 2006, Balkovič et al. 2006) as a general concept for delineation of the spatial units. **Only those parameters of landscape, which are relatively stable over time (even under climate change) and hardly adjustable by farmers were selected to create the raster of HRUs.** The HRU coverage was obtained by mere intersection of reclassified and categorized rasters such as **altitude, slope** and **soil texture** (Table 2.1).

Table 2.1: *Common criteria for HRU definition (EPIC model and land use optimization models)*

LAND CHARACTERISTIC	UNIT	CLASSES DEFINITION (INTERVALS)
altitude	meters above see level	0 – 300, 300 – 600, 600 – 1100, > 1100 (to be implemented: 1100 – 2500, > 2500),
slope inclination	degree	0 – 3, 3 – 6, 6 – 10, 10 – 15, 15 – 30, 30 – 50, > 50
soil texture	-	sandy, loamy, clay, stony, peat

The class intervals definitions for HRU delineation followed the common line of communication with higher-level optimization models so that the EPIC model derived and consecutively HRU level aggregated information on environmental indicators can be easily conveyed to the economic optimization steps.

2.1.2. Simulation units

The integration of the HRU layer with further data (weather, land use, crop management, political boundary layers, etc.) leads to individual **simulation units** (SimU). Consequently, the spatial delineation of a SimU is based on the original HRU delineation which is further sub-divided by defined interpretation frame (see the paragraph 2.1.3) and administrative region border (country level). The interpretation frame secures the local detail in the data analysis and interpretation across the delineated SimU and it makes the spatial reference to the weather or other low-resolution data. The administrative units provide important reference to statistical data on land use and management (agricultural census data) necessary for estimation of the human inputs into an agro-ecosystem. Based on averaging or most frequent class of underlying spatial data across the SimU delineation the most probable scenario of the altitude, slope and soil genetic unit combination is identified for each of the SimU.
Because of the absence of accurately spatially represented data on land cover/cropland land use (paragraph 2.2., paragraph 2.3.5) being essential for down-scaling the agricultural census data a land cover/land use information is for each SimU assigned only as a geo-coded information on area portion of the SimU delineation occupied by the defined land cover/cropland land use type. A SimU spatial delineation attributed with information on area portion of a land cover type and land use for cropland represent **full semantic content** of the SimU used in the EPIC model for global agro-ecosystem simulations.

2.1.3. Representation and communication of the spatial information

A nested set of two hierarchical geographical grids of 5 arc minutes and 30 arc minutes spatial resolution is used to unify various sources data.
The resolution level of 5 arc minutes for the data interpretation and spatial representation follows the spatial resolution of the least detailed source of essential data inputs required for the global-ecosystem characterization. More spatially detailed data can be generalized to this resolution using raster algebra or vector data based analytical tools available in GIS software toolboxes. The 30 arc minutes lower-resolution interpretation frame was selected with regard to some practical reasons which are to secure a local detail in the data analysis (secure the data unified data interpretation procedures for both the small and big countries) and to enable some simplifications in the SimU related calculations.

Grid zones represented by HRU or SimU delineations are used instead of the pixel-by-pixel identification to provide a direct spatial reference to the data records in the relational database. This is to simplify as much as possible the relational data manipulation and decrease the number of EPIC runs. Individual pixel based identification is used only for the data visualization purposes. In a global coverage of the 5 and 30 arc minutes spatial resolution grids only those pixels are considered for analysis which represents land surface (including glaciers and inland waters); all other pixels which represent oceans are excluded as being not relevant for global agro-ecosystem modeling.

Interoperability and **communication** of the geographical data/information is secured by x and y centroid coordinates information assigned to each pixel of the 5 arc minutes resolution grid. Pixel centroid coordinates can be easily used to geographically display visualization frame and any data or information can be easily attributed to the visualized points via database identification keys stored in the database tables.

2.2. Input data sources identification

Brief description of relevant data sources which were used in data analysis and interpretations for the global EPIC inputs geographical database presented in this article is given in the following paragraphs. It does not apply to bring exhaustive information on global ecosystem data available worldwide.

2.2.1. Topography

GTOPO30 digital elevation model (*http://edc.usgs.gov/products/elevation/gtopo30.html*) was used as a source of global elevation data was. GTOPO30 is a global digital elevation model available in 30 arc seconds horizontal resolution (approximately 1 km); altitude data are expressed in meters above the see level. It was derived from several raster and vector sources of topographic information. GTOPO30, completed in late 1996, was developed over a three year period through a collaborative effort led by staff at the U.S. Geological Survey's EROS Data Center.

The high-resolution global Shuttle Radar Topography Mission (SRTM) digital elevation model derived by NASA (*http://www2.jpl.nasa.gov/srtm/*) was used as a source for the derivation of the global slope data was a. NASA-SRTM digital elevation model is available in 3 arc second (approximately 90 m) horizontal resolution for areas between the latitudes from

N 60 to S 60 degrees, altitude data are expressed in meters above see level. The source for the slope data derivation in higher latitudes served also GTOPO30 digital elevation model.

2.2.2. Soil

The digital version of the 1:5 000 000 scale Soil map of the world (further referred as DSMW) version 3.6 (*http://www.fao.org/geonetwork/srv/en/metadata.show?id=14116&currTab=simple*) was used as a source of data on distribution of major soil units across the world. DSMW soil mapping units delineations (available both in vector or 5 arc minutes resolution raster) are attributed with information on soil mapping unit soil components (soil typological units, no-soil bodies) and information on their area portion (in percent) within the soil mapping unit delineation.

The World Inventory of Soil Emission Potential project soil profile database derived 5 arc minutes resolution grid of soil properties (further referred as WISE) processed by ISRIC (*http://www.isric.org/UK/About+Soils/Soil+data/Geographic+data/Global/WISE5by5minutes.htm*) was used as a source of the soil typological unit soil profile analytical data for 5 depth interval (20 cm for total depth of 1m) was. The list and detailed description of available analytical characteristic can be found in publication of Batjes (2006).

2.2.3. Administrative regions

Global Administrative Regions Layer (further referred as GAUL) version 2007 (*http://www.fao.org/geonetwork/srv/en/metadata.show?id=12691&currTab=simple*) processed under the authority of the FAO and European Commission was used as a source of the country-level administrative regions data.
Countries are in GAUL vector layer identified both by country name and country code which can be easily compared with official UN coding of countries and world regions (*http://unstats.un.org/unsd/methods/m49/m49.htm*).

2.2.4. Land cover and land use

The Global Land Cover for year 2000 (*http://www-gvm.jrc.it/glc2000/defaultGLC2000.htm*) produced as a common activity of several national and international institutions coordinated by JRC(further referred as GLC2000) was used as a basic source on land cover information on global level. GLC2000 is available as a raster layer with global coverage and spatial resolution of 30 arc seconds (approximately 1 km). In total 21 land cover classes (including natural and seminatural areas, cultivated and urbanized areas, inland waters, glaciers and oceans) are included in GLC2000 legend (see the Table 2 in Apendix).

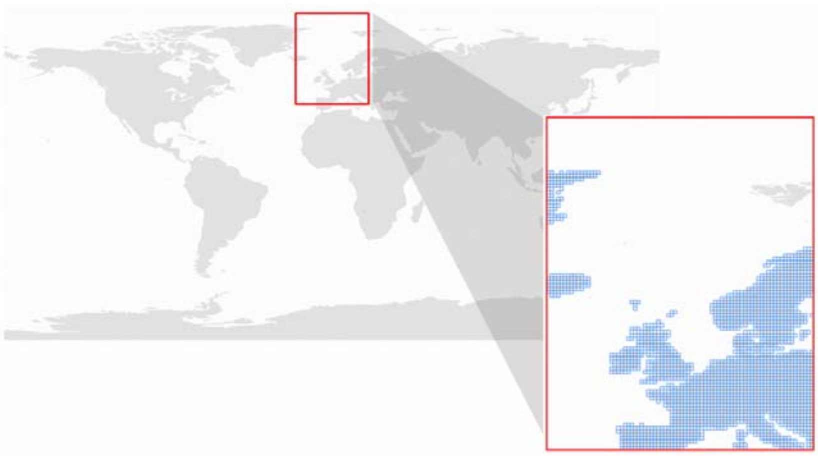
Global crop distribution data processed by International Food Policy Research Institute, Whasington D.C. (URL: www.ifpri.org) was used as a source of basic information on land cover and crop and crop management distribution on global level (data set has no been officially published yet, the data set was provided after personal communication with dr. Liangzhi You). The data (further referred as IFPRI-GLU) represents a global coverage of a regular-shaped statistical units (pixels with spatial resolution of 5 arc minutes) attributed with estimation of cultivated and harvested areas (physical area in ha) and crop production (tons/ha) for 20 most important crops managed under the four agricultural production systems (irrigated, high input, low input and subsistence). The data estimation (down-scaling) was based on analysis and interpretation of national and sub-national agricultural census data and spatial data on global land cover (GLC2000), crop suitability of area data and population density data. All the data details and methodological rationales are given in the article of You et Wood (2006).

2.3. Metadata description of the geographical database

2.3.1. Pilot area selection

A randomly selected rectangular-shaped area (Fig. 2.1) with spatial extent of Xmin = -20 dd (decimal degree), Ymin = 40 dd, Xmax = 20 dd and Ymax = 90 dd was used as a pilot area for design, construction and testing the geographical database for global EPIC model input (and potentially output) data organization. Within the pilot area all analysis and data interpretations were done using globally available data and obtained the results therefore represents **full global-level operability of the geographical database**.

Figure 2.1: *Pilot area for design and testing the geographical database for global EPIC modeling*



2.3.2. General description

To the date, pilot geographical database for global agro-ecosystem modeling addresses mostly the **geographical aspect of the database task** (it provides a geographical frame for the EPIC model input/output data representation). **Attribute data task** is not fully covered and only essential data on topography and soil required by the EPIC model have been included so far. Although other essential attribute data for the EPIC model runs (management data - crop calendar, fertilization, irrigation, etc.) has not been included yet, the crop and agricultural management system specific land use data assigned to SimU delineation provide a link to the agricultural management dataset being actually processed.

The geographical database comprises four separate datasets addressing various aspects of a global agro-ecosystem structural model:

- Geographical and visualization data frame (**GeoFrame**);
- SimU specific EPIC model topography and soil input data for the crop land area (**EPIC_CrLnInput**);
- SimU specific data on land cover types areas (**LandCover**)

- SimU specific data on cultivated crop areas separately for each management system (**CrSh_H, CrSh_I, CrSh_L, CrSh_S**);

The datasets are **implemented** in commercial Microsoft® Office Access 2003 database system. Simple database structure enables **export** the datasets in arbitrary interoperable data format and makes it open for different database systems.

2.3.3. GeoFrame dataset

GeoFrame dataset represents the frame for a global data geographical representation and visualization. Regular lattice of pixel centroid with spatial resolution of 5 arcmin is attributed with all necessary attributes for HRU/SimU identification and definition. GeoFrame dataset also contains data on land cover specific visualization mask. The list and identification of the GeoFrame dataset attributes is given in the Table 2.2.

Table 2.2: *List of GeoFrame dataset attributes*

ATTRIBUTE	DATA TYPE	SHORT DESCRIPTION
COL_ROW	Text(12)	**primary database key**, column and row based indexing of 5min resolution grid;
XCOORD	Number(Double)	geographic projection field; x coordinate [decimal degree] of 5 arcmin resolution pixel centroid;
YCOORD	Number(Double)	geographic projection field; y coordinate [decimal degree] of 5 arcmin resolution pixel centroid;
Pxarea	Number(Double)	real physical area of 5 arcmin resolution pixel [ha];
HRU	Number(Integer)	**foreign database key**, code of homogenous response unit (3-digit combination of 1. altitude class (vlues 1 – 5), 2. slope class (values 1 – 7) and 3. soil texture class (values 1 – 5));
COL_ROW30	Text(12)	**foreign database key**, identification of 30min reference grid pixel (data interpretation frame);
COUNTRY	Number(LongInteger)	**foreign database key**, identification of country according to official UN list of country codes;
Crop_mask	Number(Byte)	cropland mask for spatial representation and visualization of the data;
OthAg_mask	Number(Byte)	other agricultural land mask for spatial representation and visualization of the data;
Grass_mask	Number(Byte)	grassland mask for spatial representation and visualization of the data;
Forest_mask	Number(Byte)	forestland mask for spatial representation and visualization of the data;
OthNat_mask	Number(Byte)	other natural vegetation land mask for spatial representation and visualization of the data;
NotRel_mask	Number(Byte)	not relevant areas mask for spatial representation and visualization of the data;

The additional metadata comments have to be given to the following attribute fields or attribute fields groups:

COL_ROW, COL_ROW30
The column and row identification is based on global 5 (30) arc minutes resolution grid with extent from 90 dd to -90 dd for a latitude and from 180 dd to – 180 dd for a longitude. The COL_ROW (COL_ROW30) is count from upper left corner of the global extent grid. The irrelevant pixels were excluded (see the paragraph 2.1.3) only after each pixel was assigned with the COL_ROW value.

COL_ROW serves the unique identification for each of the 5 arc minutes resolution pixel. COL_ROW30 identification is an essential component for SimU spatial delineation and identification.

PXarea

To get correct area of the pixels displayed in a geographic co-ordinate system the input data had to be converted into an equal-area projection. A routine described by Lethcoe and Klaver (1998) was used to transform geographical projected data into equal-area Goode-Homolosine equal area projection system which is most frequently used for global data.

Because of high demand on real area computation at global level the calculation was simplified and geographic coordinate system transformation and real area calculation was done only for the 30 arc minutes resolution grid. Real area value for the final 5 arc minutes resolution pixels yielded from dividing the 30 arc minutes resolution pixel real area by 36 (total number of 5 arc minutes resolution pixels falling to the 30 arc minutes one).

This simplification should not significantly influence the real area estimation for SimU or HRU in latitudes up to 60 dd or – 60 dd where agricultural land use mostly occurs. Nevertheless, an additional and detailed analyze should be done to verify the correctness of the real area calculation algorithm.

HRU

HRU is spatially delineated as a set (zone) of individual pixels (5 arc min resolution grid) characterized by the same class of altitude, slope and soil. HRU is an essential component for SimU spatial delineation and identification.

HRU specific **altitude class** (Table 2.1) represent most frequent altitude class calculated across the 5 arc minutes resolution pixel. The most frequent value was calculated as a zonal majority of an underlying 30 arc seconds resolution GTOPO30 raster based altitude classification. Totally, 5 altitude classes are considered for HRU definition.

HRU specific **slope class** (Table 2.1) represent most frequent slope class calculated across the 5 arc minutes resolution pixel. The most frequent value was calculated as a zonal majority of an underlying 30 arc seconds resolution NASA-SRTM and GTOPO30 data derived slope data. Prior to it was used in final calculations the original 3 arc seconds resolution NASA-SRTM data set was transformed to a lower (30 arc seconds) resolution using zonal majority procedure. For missing latitudes (from 60 to 90 deg N) adjusted GTOPO30 derived slope class values were used (original GTOPO classes were modified using look up table coming from analysis of the 3 arc second resolution slope classes frequencies across the 30 arc second resolution pixel in the latitudes where both NASA-SRTM and GTOPO30 data were available). Totally, 7 slope classes are considered for HRU delineation.

HRU specific **soil texture class** (Table 2.1) represent the most frequent soil texture class identified across the 5 arc minutes resolution pixel. The most frequent value for the defined pixel was calculated as a sum of relative to pixel area (%) areas of the pre-classified DSMW soil typological unit data yielded from the DSMW soil mapping unit composition data. WISE analytical table data on the soil typological unit specific soil texture distribution in a soil profile was used for the DSMW soil typological unit pre-classification into a 5 textural classes (soil profile dominant texture class was taken and organic and stony soils were identified). Totally, 5 slope texture classes are considered for HRU delineation.

COUNTRY

The 5 arc minutes pixel representation of GAUL data was used to identify administrative region on a country level for each pixel. COUNTRY is an essential component for SimU spatial delineation and identification.

Crop_mask, OthAg_mask, Grass_mask, Forest_mask, OthNat_mask, NotRel_mask
Because of the spatially explicit identification of land cover type is absent in SimU delineation Crop_mask, OthAg_mask, Grass_mask, Forest_mask, OthNat_mask, and NotRel_mask attribute fields serve the land cover type specific EPIC modeling input/output data visualization frame. The 1 or 0 value was assigned to the pixel if more (1) or less (0) than 275 ha of the defined land cover type is present in the pixel (more details on land cover type interpretation are given in paragraph 2.3.5). The 99 value was assigned to all pixels with significant inconsistence in input land cover data used for the land cover type interpretation (see the paragraph 2.3.5). If the defined land cover type specific data (EPIC inputs/outputs, other landscape qualities) is to be geographically displayed only 1-valued pixels should be considered as a relevant for the visualization.

2.3.4. EPIC_CrLnInput dataset

EPIC_CrLnInput dataset is the source of a cropland related input data for the EPIC model simulations. The HRU, COL_ROW30 and COUNTRY based unique identification is used to interpret and attribute all mandatory soil and topography data required by the EPIC model to the delineated SimU. List and identification of the EPIC_CrLnInput dataset attributes is given in the Table 2.3.

Table 2.3: *List of EPIC_CrLnInput dataset attributes*

ATTRIBUTE	DATA TYPE	SHORT DESCRIPTION
HRU	Number(Integer)	**primary database key**, code of homogenous response unit (3-digit combination of 1. altitude class (vlues 1 – 5), 2. slope class (values 1 – 7) and 3. soil texture class (values 1 – 5));
COL_ROW30	Text(12)	**primary database key**, identification of 30min reference grid pixel (data interpretation frame);
COUNTRY	Number(LongInteger)	**primary database key**, identification of country according to official UN list of country codes;
SimU_lon	Number(Double)	longitude (decimal degrees) asigned to SimU delineation;
SimU_lat	Number(Double)	latitude (decimal degrees) asigned to SimU delineation;
SimU_corArea	Number(Double)	SimU information field, physical area (ha) of SimU delineation (corrected);
CrLn_area	Number(Single)	SimU information field, physical area (ha) of a cropland within the SimU delineation;
SimU_alti	Number(Double)	mean altitude (m) for SimU delineation;
SimU_slp	Number(Double)	dominat slope (%) within SimUdelineation;
SimU_STU	Text(2)	SimU information field, area dominant DSMW soil typological unit within the SimU;
HG	Number(Byte)	USDA NRCS soil hydrological group;
ALB	Number(Single)	albedo of moist soil surface, a constant value set to 0,15;
*) DEPTH_	Number(single)	depth of soil layer bottom (m), soil layers 1 - 5;
*) VS_	Number(Integer)	volume of stones (%); content of soil fragments > 2mm in diameter, soil layers 1 – 5;
*) SAND_	Number(Integer)	sand content (%), content of particles of 0.063 – 2 mm, soil layers 1 - 5;
*) SILT_	Number(Integer)	silt content (%), content of particles of 0.002 - 0.063 mm, soil layers 1 - 5;
*) CLAY_	Number(Integer)	silt content (%), content of particles of 0.002 - 0.063 mm, soil layers 1 - 5;
*) BD_	Number(Single)	bulk density (t/m3) of soil, soil layers 1 - 5;
*) CEC_	Number(Single)	soil cation exchange capacity (cmol/kg) , soil layers 1 - 5;
*) SB_	Number(Single)	sum of bases (cmol/kg) , soil layers 1 - 5;
*) PH_	Number(Single)	soil pH in H2O, soil layers 1 - 5;

ATTRIBUTE	DATA TYPE	SHORT DESCRIPTION
*) CARB_	Number(Single)	(calcium) carbonate content (%) in a soil, soil layers 1 - 5;
*) CORG_	Number(Single)	soil organic carbon content (%), soil layers 1 - 5;
*) FWC_	Number(Double)	water content in soil (mm/mm) at field water capacity, soil layer 1 - 5;
*) WP_	Number(Double)	water content in soil (mm/mm) at wilting point, soil layer 1 - 5;
*) KS_	Number(Double)	saturated soil hydraulic conductivity (mm/hour), soil layer 1 - 5;

*) To represent all of the five depth interval of the soil profile, attribute field is in the dataset multiplied; symbol "_" is in the attribute field name replaced by real number of a soil layer (1 – 5).

The additional metadata comments have to be given to the following attribute fields or attribute fields groups:

HRU, COL_ROW30, COUNTRY
The attribute fields combination stands for unique identification of the SimU delineation. For detailed description see the paragraph 2.3.3.

SimU_corArea
For detailed description see the paragraph 2.3.5.

SimU_Lon, SimU_lat
To represent a general geographical position of SimU for EPIC modeling purposes, longitude and latitude values of the 30 arc minute resolution pixel centroid (x and y coordinates) are assigned to each of the SimU delineated falling in its border. This simplification should not influence significantly the EPIC modeling results.

SimU_Alti, SimU_Slp, SimU_STU
Because of the altitude, slope and soil characteristics used for HRU delineation are too general and rough to be directly applied as a SimU specific input values for the EPIC model simulation to get locally specific data a detailed input data interpretation was done for a **simplified SimU delineation** (HRU and 30 arc minutes spatial resolution interpretation frame defined SimU delineations).
Zonal mean procedure of the underlying 30 arc seconds resolution GTOPO30 altitude values was used to get the SimU representative value for the **altitude**.
Table 2.4: *Representative slope values assigned to the slope classes*

CLASS INTERVAL	REPRESENTATIVE VALUE (DEG)
0 - 3	1
3 - 6	4
6 - 10	8
10 - 15	12
15 - 30	23
30 - 50	40
> 50	60

To get the SimU specific value of **slope** a representative value was assigned to each of the slope class interval used for HRU delineation (Table 2.4). Consecutively, the values were re-calculated by simple formula to get a slope values expressed in per-cent value (*slope in % = tg(slope in deg) * 100*). In a case a quasi-homogenity is supposed for SimU (see the paragraph 2.1.1) the simple-defined algorithm for SimU specific slope estimation as it was applied seems to be more appropriate as an averaging procedure because the averaging can result in

non-realistic (artificial) values if a wide rage of slopes is presented within a SimU spatial delineation.
A slightly modified methodology firstly introduced by Batjes at al. (1995) was applied to get SimU representative **soil typological unit** from DSMW data. As a difference from the above mentioned methodology, a simplified SimU delineation was used as an interpretation frame instead of 30 arc minute resolution pixel and only dominant soil typological unit was considered in the further data processing. The SimU dominant soil typological unit was calculated as a sum of relative to pixel area (%) areas of the soil typological unit yielded from the DSMW soil mapping unit composition data and the DSMW soil mapping unit area portion (%) of total area of the simplified SimU delineation. The dominant soil typological unit area portion within the simplified SimU delineation obtained for dominant soil typological units vary in extreme cases from 14 to 100%, in most cases, however, the area portion of the dominant soil typological unit is between 40 – 60% of the simplified SimU delineation area. The SimU spatial delineation related dominant soil typological unit provided the reference to all other soil analytical data.

HG
Hydrological soil group was interpreted for each soil typological unit from WISE analytical data on soil texture using simplified rules given in official USDA-NRCS engineering manual (USDA-NRCS 2007).

DEPTH_, VS_, SAND_, SILT_, CLAY_, BD_, CEC_, SB_, PH_, CARB_, CORG_
Original WISE soil analytical data was used to provide all necessary soil characteristics for the dominant soil typological units. Only small refinements were done to the data to convert original units of the soil characteristics values to the units required by the EPIC model (SB_, CARB_, CORG_).

FWC_, WP_, KS_
DSMW soil typological units specific hydro-physical characteristics of soil required by the EPIC model were estimated from WISE analytical data on soil texture and bulk density using neural-network based pedotransfer model Rosetta (Shaap et Bouten 1996); water content (mm/mm) at field water capacity and wilting point was calculated by Van Genuchten equation for the soil water retention curve (Wösten et Van Genuchten 1988) using Rosetta model derived parameters (residual moisture, water content in saturated soil and alpha and n shape parameters of the soil water retention curve).

2.3.5. LandCover dataset

LandCover dataset represent non-spatial semantic content of delineated SimU. Land cover types in more detail specify the characteristics of each delineated SimU. Physical and relative to total delineated SimU area (%) express the area portion of each land cover type within the defined SimU spatial delineation. List and identification of the LandCover dataset attributes is given in the Table 2.5.

Table 2.5: *List of LandCover dataset attributes*

ATTRIBUTE	DATA TYPE	SHORT DESCRIPTION
HRU	Number(Integer)	**primary database key**, code of homogenous response unit (3-digit combination of 1. altitude class (vlues 1 – 5), 2. slope class (values 1 – 7) and 3. soil texture class (values 1 – 5));
COL_ROW30	Text(12)	**primary database key**, identification of 30min reference grid

ATTRIBUTE	DATA TYPE	SHORT DESCRIPTION
		pixel (data interpretation frame);
COUNTRY	Number(LongInteger)	**primary database key**, identification of country according to official UN list of country codes;
SimU_totArea	Number(Double)	SimU information field; total area (ha) of the SimU delineation;
SimU_corArea	Number(Double)	SimU information field; total area (ha) of the SimU delineation (corrected);
InCon_p	Number(Double)	inconsistent land cover data pixels physical area (ha) within SimU delineation;
CrLn_p	Number(Double)	physical area (ha) of cropland within the SimU delineation;
OthAg_p	Number(Double)	physical area (ha) of other agricultural land within the SimU delineation;
Grass_p	Number(Double)	physical area (ha) of grassland within the SimU delineation;
Forest_p	Number(Double)	physical area (ha) of forestland within the SimU delineation;
OthNat_p	Number(Double)	physical area (ha) of other natural vegetation within the SimU delineation;
NotRel_p	Number(Double)	physical area (ha) of not relevant land covers within the SimU delineation;
InCon_r	Number(Double)	inconsistent land cover data pixels relative to total SimU area (%) within SimU delineation;
CrLn_r	Number(Double)	relative to total SimU area (%) of cropland within the SimU delineation;
OthAg_r	Number(Double)	relative to total SimU area (%) of other agricultural land within the SimU delineation;
Grass_r	Number(Double)	relative to total SimU area (%) of grassland within the SimU delineation;
Forest_r	Number(Double)	relative to total SimU area (%) of forestland within the SimU delineation;
OthNat_r	Number(Double)	relative to total SimU area (%) of other natural vegetation within the SimU delineation;
NotRel_r	Number(Double)	relative to total SimU area (%) of not relevant land covers within the SimU delineation;

The additional metadata comments have to be given to the following attribute fields or attribute fields groups:

HRU, COL_ROW30, COUNTRY
The attribute fields combination stands for unique identification of the SimU delineation. For detailed description see the paragraph 2.3.3.

All other attribute fields
Land cover types interpretation for the further characterization of SimU delineations was based on GLC 2000 and IFPRI-GLU datasets. The combined data sources interpretation was necessary because none of the input data sources fully covered two of the most important requirements for land cover information:

- Cropland has to be exactly identified as the ago-ecosystem is in the focus of the global EPIC modeling (information absent in GLC 2000 data);
- Also other land cover types (grasslands, forest land, other natural vegetation) have to be exactly identified so that HRU/SimU delineation related land cover/use change scenarios can take a place in a economic-optimization steps (information absent in IFPRI-GLU data).

The basis for the land cover interpretation IFPRI-GLU dataset was selected because it represent source of complex information on spatial distribution of land use systems within

cropland and crop share data (see the paragraph 2.3.6) important for interpretation of the possible crop rotations which are necessary for the EPIC model simulations. The GLC 2000 dataset was decided to fill the gap in the information on other important land cover types absent in IFPRI-GLU data.
Pixel-by-pixel approach was applied to analyze IFPRI-GLU cropland area and GLC 2000 derived (pixel specific) look up table with land cover types area portion of total pixel area (land cover types definitions are given in Table 2.6). Interpretation rules were set to get the pixel specific area of the defined land cover types:

- If the cropland area reported in IFPRI-GLU was less than sum of GLC 2000 reported agricultural land cover classes (16,17,18), IFPRI-GLU reported value represent the total cropland type area for the defined pixel and other agricultural land type area is calculated as the subtract of the IFPRI-GLU derived cropland type area and total agricultural area reported by GLC 2000. Areas for all other land cover types came directly from GLC 2000 derived look up table;
- If the cropland area reported in IFPRI-GLU was more than sum of GLC 2000 reported agricultural land cover classes (16,17,18) and the difference did not exceed an arbitrary set threshold of 500 ha, IFPRI-GLU reported value represent the total cropland type area for the defined pixel and other agricultural land area is equal to zero. Areas for all other land cover types are recalculated to yield 100% of the total pixel area after cropland area portion is added;
- If the cropland area reported in IFPRI-GLU was more than sum of GLC 2000 reported agricultural land cover classes (16,17,18) and the difference exceeded an arbitrary set threshold of 500 ha the pixel was assigned as inconsistent (value 99 for land cover visualization frame). Only 0.65% of the total GeoFrame dataset pixel count was assigned as irrelevant and omitted from further analyses because of land cover data inconsistencies.

A zonal count procedure was used to obtain final SimU related area portion or physical area of all land cover types falling within its border and to calculate **total area** of the SimU spatial delineation (SimU_totArea attribute field). In a case of some inconsistent land cover data pixels fall within the SimU spatial delineation the area portion (or physical area) of such pixels was recorded in a specific class (InCon_r or InCon_p attribute fields) and total area of the SimU spatial delineation was corrected by subtracting the inconsistent pixel areas from total SimU delineation area and **corrected area** was recorded in a specific field (SimU_corArea attribute field).

Table 2.6: *GLC2000 classes derived land cover types for further characterization of the SimU delineations*

LAND COVER TYPE	ABBREVIATION	ORIGINAL GLC2000 CLASSESS*)
cropland	CrLn	- (IFPRI-GLU cropland)
other agricultural land	OthAg	16,17,18, (IFPRI-GLU reported cropland subtracted)
grassland	Grass	13,
forestland	Forest	1, 2, 3, 4, 5, 6, 9, 10,
other natural vegetation	OthNat	7, 8, 11, 12, 14, 15,
not relevant land covers	NotRel	19, 20, 21, 22,

*) GLC 2000 land cover classes description is given in appendix (Table 2)

2.3.6. CrSh_I, CrSh_H, CrSh_L and CrSh_S datasets

CrSh_, CrSh_H, CrSh_L and CrSh_S datasets serve the land use specification for cropland relevant SimU. Agricultural management (irrigated, high input, low input and subsistence) specific crop shares including 20 most frequent crops or group of crops are reported in four separate datasets with the same structure. List and identification of the CrSh_, CrSh_H, CrSh_L and CrSh_S datasets attributes is given in the Table 2.7.

Table 2.7: *List of CrSh_I, CrSh_H, CrSh_L and CrSh_S datasets attributes*

ATTRIBUTE	DATA TYPE	SHORT DESCRIPTION
HRU	Number(Integer)	**primary database key**, code of homogenous response unit (3-digit combination of 1. altitude class (vlues 1 – 5), 2. slope class (values 1 – 7) and 3. soil texture class (values 1 – 5));
COL_ROW30	Text(12)	**primary database key**, identification of 30min reference grid pixel (data interpretation frame);
COUNTRY	Number(LongInteger)	**primary database key**, identification of country according to official UN list of country codes;
*) Phys_	Number(Double)	total area (ha) within SimU delineation where crops are cultivated within the defined management system;
WHEA	Number(Double)	relative (%) area of crop within the defined management system, wheat;
RICE	Number(Double)	relative (%) area of crop within the defined management system, rice (paddy);
MAIZ	Number(Double)	relative (%) area of crop within the defined management system, maize;
BARL	Number(Double)	relative (%) area of crop within the defined management system, barley;
MILL	Number(Double)	relative (%) area of crop within the defined management system, millet;
SORG	Number(Double)	relative (%) area of crop within the defined management system, sorghum;
POTA	Number(Double)	relative (%) area of crop within the defined management system, potato;
SWPY	Number(Double)	relative (%) area of crop within the defined management system, sweet potato and yams;
CASS	Number(Double)	relative (%) area of crop within the defined management system, cassava;
BANP	Number(Double)	relative (%) area of crop within the defined management system, plantain and banana;
SOYB	Number(Double)	relative (%) area of crop within the defined management system, soybean;
BEAN	Number(Double)	relative (%) area of crop within the defined management system, beans;
OPUL	Number(Double)	relative (%) area of crop within the defined management system, other Pulse;
SUGB	Number(Double)	relative (%) area of crop within the defined management system, sugar beets;
SUGC	Number(Double)	relative (%) area of crop within the defined management system, sugar cane;
COFF	Number(Double)	relative (%) area of crop within the defined management system, coffee;
COTT	Number(Double)	relative (%) area of crop within the defined management system, cotton;
OFIB	Number(Double)	relative (%) area of crop within the defined management system, other fibres;
GROU	Number(Double)	relative (%) area of crop within the defined management system, groundnuts;

ATTRIBUTE	DATA TYPE	SHORT DESCRIPTION
OOIL	Number(Double)	relative (%) area of crop within the defined management system, other oil crops;
OTHE	Number(Double)	relative (%) area of crop within the defined management system, other not-listed crop;

*) Irrigated, high input, low input or subsistence agriculture management system can be addressed and identified. Abbreviations “I”, “H”, “L” or “S” stands for the defined management system identification in the attribute field name (suffix position).

The additional metadata comments have to be given to the following attribute fields or attribute fields groups:

HRU, COL_ROW30, COUNTRY
The attribute fields combination stands for unique identification of the SimU delineation. For detailed description see the paragraph 2.3.3.

All other attribute fields
The relative to total cropland area within SimU areas (%) of the individual crops and total physical area (ha) occupied by the defined management system were re-calculated from data on physical crop areas coming from IFPRI-GLU dataset.

References and web links

Balkovič, J., Schmid, E., Bujnovský, R., Skalský, R., Poltárska, K., 2006. Biophysical modelling for evaluating soil carbon sequestration potentials on arable land in the pilot area Baden-Württemberg (Germany). Agriculture, Vol. 52, No. 4, p. 169 - 176

Batjes, N., H., 2006. ISRIC-WISE derived soil properties on a 5 by 5 arc-minutes global grid (ver. 1.1). Report 2006/2, ISRIC - World Soil Information, Wageningen, 46 pp., Available on internet: http://www.isric.org/isric/webdocs/Docs/ISRIC_Report_2006_02.pdf

Batjes, N. H., Bridges, E.M., Nachtergaele, F.O., 1995. World Inventory of Soil Emission Potentials: Development of a Global Soil Data Base of Process-Controlling Factors. In Peng, S., Ingram, K.T., Neue, H.U., Ziska, L.H. (Eds.) 1995. Climate change and rice. Springer-Verlar, Berlin, Heidelberg, p. 102 - 115

Izaurralde, R.C., J.R. Williams, W.B. McGill, and N.J. Rosenberg. 2001. Simulating soil carbon dynamics, erosion, and tillage with EPIC. Presented at the First National Conference on Carbon Sequestration sponsored by the U.S. Department of Energy – National Energy Technology Laboratory, 14-17 May, Washington, D.C. Available on internet: http://www.netl.doe.gov/publications/proceedings/01/carbon_seq/5c2.pdf (accessed 18 September 2007)

Jones, C.A., P.T. Dyke, J.R. Williams, J.R. Kiniry, V.W. Benson, and R.H. Griggs. 1991. EPIC: an operational model for evaluation of agricultural sustainability. Agric. Syst. 1991, 37:341-350.

Kinsel, W.G. 1980. A field scale model for chemicals, runoff, and erosion from agricultural management systems. U.S. Dept. Agric. Conserv. Res. Rept. No. 26.

Leonard, R.A., Knisel, W.G., and Still, D.A. 1987. *GLEAMS:* Ground water loading effects on agricultural management systems. Trans. ASAE, 1987, 30(5), pp 1403-1428.

Lethcoe, K.J., Klaver, R.W., 1998. Simulating the Interrupted Goode Homolosine Projection With ArcInfo. 18th Annual ESRI International User Conference in San Diego Proceedings, California, Available on internet: http://gis.esri.com/library/userconf/proc98/PROCEED/TO850/PAP844/P844.HTM#KJ (Accessed: 18 September 2007)

Parton, W.J., D.S. Ojima, C.V. Cole, and D.S. Schimel. 1994. A general model for soil organic matter dynamics: senstivity to litter chemistry, texture and management. In Quantitative modeling of soil forming processes, SSSA Spec. Public., Madison, WI: SSSA, 1994, No. 39., pp 147-167

Putnam, J., J. Williams, and D. Sawyer. 1988. Using the erosion-productivity impact calculator (EPIC) to estimate the impact of soil erosion for the 1985 RCA appraisal. In Journal Soil Water Cons. 1988, 43(4):pp 321-326.

Schmid, E., Balkovič, J., Moltchanova E., Skalský, R., Poltárska, K., Müller, B., Bujnovský, R., 2006: Biophysical process modeling for EU25: concept, data, methods, and results. Deliverable D3 (T30), Final Report, Appendix II., EU FP 6 Project INSEA – Integrated Sink Enhancement Assessment (SSPI-CT-2003/503614 with DG RTD), 2006, 76 s.

Schaap, M.G., Bouten, W., 1996. Modelling Water Retention Curves of Sandy Soils Using Neural Networks. Water Resour. Res., Vol. 32, s. 3033 – 3040.

Sharpley, A.N., Williams, J.R., (eds). 1990. EPIC - Erosion/Productivity Impact Calculator: 1. Model Documentation. U.S. Dept. Agric. Tech. Bull. No. 1768.

USDA-NRCS, 2007. Hydrologic Soil Groups. National Engineering Handbook : Part 630 Hydrology : Chapter 7. 210-VI-NEH-630.7, USDA, Washington DC. Available on internet: http://directives.sc.egov.usda.gov/media/pdf/H_210_630_7.pdf (Accessed 18 September 2007)

Williams, J. R. 1990. The erosion productivity impact calculator (EPIC) model: a case history. Phil. Trans. R. Soc. London, 1990, 329: pp 421-428.

Williams, J.R. 1995. The EPIC Model. In Computer Models of Watershed Hydrology (Ed.: V.P. Singh). Water Resources Publications, Highlands Ranch, Colorado, 1995, pp 909-1000.

Williams, J.R., C.A. Jones, and P.T. Dyke. 1984. A modeling approach to determining the relationship between erosion and soil productivity. Trans. ASAE, 1984, 27(1), pp 129-144.

Williams, J.R., A.D. Nicks, and J.G. Arnold. 1985. SWRRB, a simulator for water resources in rural basins. In Journal ASCE Hydr., 1985, 111(6): pp 970-986.

Williams, J.R., J.G. Arnold, and R. Srinivasan. 2000. The APEX model. BRC Report No. 00-06. Temple, TX: Texas Agric. Expt. Station, Texas Agric. Exten. Service, Texas A&M University.

Wösten, J. H. M., Van Genuchten, M. T., 1988. Using Texture and Other Soil Properties to Predict the Unsaturated Soil Hydraulic Functions. Soil. Sci. Soc. Am. J., Vol. 52, p. 1762 - 1770

You, L. and S. Wood. 2006. An entropy approach to spatial disaggregation of agricultural production. Agricultural Systems Vol.90, Issues1-3 p.329-347.

DSMW – *http://www.fao.org/geonetwork/srv/en/metadata.show?id=14116 &currTab=simple*

GAUL - *http://www.fao.org/geonetwork/srv/en/metadata.show?id=12691&currTab= simple*

GLC2000 - *http://www-gvm.jrc.it/glc2000/defaultGLC2000.htm*

GTOPO30 – *http://edc.usgs.gov/products/elevation/gtopo30.html*

NASA-SRTM – *http://www2.jpl.nasa.gov/srtm/*

UN Administrative regions - *http://unstats.un.org/unsd/methods/m49/m49.htm*

WISE - *http://www.isric.org/UK/About+Soils/Soil+data/Geographic+data/Global/ WISE5by 5minutes.htm*

Appendix

Table 1: *List of physical and chemical soil parameters needed by EPIC*

ESSENTIAL SOIL INFORMATION	USEFUL SOIL INFORMATION
general soil and hydrologic information	
soil albedo (moist)	initial soil water content (fraction of field capacity)
hydrologic soil group (A, B, C, or D)	minimum depth to water table in m
	maximum depth to water table in m
	initial depth to water table in m
	initial ground water storage in mm
	maximum ground water storage in mm
	ground water residence time in days
	return flow fraction of water percolating through root zone
	soil weathering (CaCO3 soils; non-CaCO3 soils that are slightly, moderately or highly weathered)
	number of years of cultivation
	soil group (kaolinitic, mixed, or smetitic)
	fraction of org C in biomass pool
	fraction of humus in passive pool
	soil weathering code
by soil layer	
depth from surface to bottom of soil layer in m	bulk density of the soil layer (oven dry) in t/m3
bulk density of the soil layer (moist) in t/m3	wilting point (1500 kPa for many soils) in m/m
sand content in %	field capacity (33 kPA for many soils) in m/m
silt content in %	Initial organic N concentration in g/t
soil pH	sum of bases in cmol/kg
organic carbon in %	cation exchange capacity in cmol/kg
calcium carbonate content in %	coarse fragment content in %vol.
	initial soluble N concentration in g/t
	initial soluble P concentration in g/t
	initial organic P concentration in g/t
	exchangeable K concentration in g/t
	crop residue in t/ha
	saturated conductivity in mm/h
	fraction of storage interacting with NO3 leaching
	phosphorous sorption ratio
	lateral hydraulic conductivity in mm/h
	electrical conductivity in mm/cm
	structural litter kg/ha
	metabolic litter kg/ha
	lignin content of structural litter in kg/ha
	carbon content of structural litter in kg/ha
	C content of metabolic litter in kg/ha
	C content of lignin of structural litter in kg/ha
	N content of lignin of structural litter in kg/ha
	C content of biomass in kg/ha
	C content of slow humus in kg/ha
	C content of passive humus kg/ha
	N content of structural litter in kg/ha
	N content of metabolic litter in kg/ha
	N content of biomass in kg/ha
	N content of slow humus in kg/ha
	N content of passive humus in kg/ha
	observed C content at the end of simulation

Table 2: *Legend to Global Land Cover 2000 (GLC2000)*

LAND COVER CLASS	CLASS DESCRIPTION
1	Tree Cover, broadleaved, evergreen
2	Tree Cover, broadleaved, deciduous, closed
3	Tree Cover, broadleaved, deciduous, open
4	Tree Cover, needle-leaved, evergreen
5	Tree Cover, needle-leaved, deciduous
6	Tree Cover, mixed leaf type
7	Tree Cover, regularly flooded, fresh water
8	Tree Cover, regularly flooded, saline water
9	Mosaic: Tree Cover / Other natural vegetation
10	Tree Cover, burnt
11	Shrub Cover, closed-open, evergreen
12	Shrub Cover, closed-open, deciduous
13	Herbaceous Cover, closed-open
14	Sparse herbaceous or sparse shrub cover
15	Regularly flooded shrub and/or herbaceous cover
16	Cultivated and managed areas
17	Mosaic: Cropland / Tree Cover / Other natural vegetation
18	Mosaic: Cropland / Shrub and/or grass cover
19	Bare Areas
20	Water Bodies
21	Snow and Ice
22	Artificial surfaces and associated areas
23	No data