VIENNA UNIVERSITY OF TECHNOLOGY
INSTITUTE OF PHOTOGRAMMETRY AND REMOTE SENSING

DIPLOMARBEIT

Metadata Management of Higher Level Remote Sensing Products

Ausgeführt zum Zwecke der Erlangung des akademischen Grades
eines Diplom-Ingenieur unter der Leitung von

Univ.Prof. Dipl.-Ing. Dr.techn. Wolfgang Wagner
Institut für Photogrammetrie und Fernerkundung

eingereicht an der Technischen Universität Wien
Fakultät für Mathematik und Geoinformation
von

Thomas Mistelbauer

Matrikelnummer 0525480
Lichtenau 59
A-3522 Lichtenau

Wien, am 28. Mai 2012 ....

Abstract

Remote sensing products derived from radar satellite data are developed for global monitoring of the hydrosphere. Depending on the instrument, for example Synthetic Aperture Radar (SAR) or Scatterometer (SCAT), the recorded data shows diverse characteristics which enables a great number of applications. With the launch of new satellites with advanced instruments (e.g. Sentinel 1, foreseen launch date around 2013), this range of applications is expected to be extended. Further, existing processing algorithms may have to be adapted or even new algorithms may have to be developed to process the new data. Hence, the interplay of data retrieval and data processing is a steady dynamic process. During the processing from low level- (level 1, raw data) to high level products (level 2, level 3; geophysical products regularly sampled in space and time), various software packages and auxiliary data may be used.

Information about the products characteristics (from acquisition to processing), called metadata, serves as an important source of information for both developers and third-party users. The International Organization for Standardization (ISO) provides metadata standards for describing geographic information. Within the European Union (EU), another important guideline for describing spatial data is the directive for the creation of a geodata-infrastructure in the European Community, called INSPIRE (Infrastructure for Spatial Information in Europe).

The objective of this master thesis is to investigate methods for managing products and the corresponding metadata, and to provide this information for third party users, taking into account the mentioned standards for geographic metadata. The approach is based upon implementing standard compliant XML files, and the design of a relational database for the purpose of metadata storage and management.

Kurzfassung

Fernerkundungsprodukte die aus Radar Satellitendaten abgeleitet werden, werden für die globale Überwachung der Hydrosphäre entwickelt. Je nach verwendetem Messinstrument, wie zum Beispiel Synthetic Aperture Radar (SAR) oder Scatterometer (SCAT) zeigen die aufgezeichneten Daten unterschiedliche Eigenschaften, was eine Vielzahl an Anwendungsmöglichkeiten ermöglicht. Mit dem Start neuer Satelliten (z.B. Sentinel 1, vorgesehener Start um 2013), die neue und besser entwickelte Messinstrumente an Board tragen, solle der Umfang an Anwendungsmöglichkeiten weiter vergrößert werden. Dies kann aber auch bedeuten, dass bestehende Algorithmen zur Prozessierung der neuen Daten an deren Charakteristiken angepasst, oder sogar neue Algorithmen entwickelt werden müssen. Demzufolge ist das Zusammenspiel von Datenbeschaffung und Datenverarbeitung ein stetiger dynamischer Prozess. Im Laufe der Prozessierung von Produkten niedrigeren Levels (Level 1, Rohdaten) bis hin zu Produkten höheren Levels (Level 2, Level 3; georeferenzierte geophysikalische Variablen) können verschiedene Software-Pakete und Hilfsdaten zum Einsatz kommen.

Informationen über die Eigenschaften von Fernerkundungsprodukten (von der Gewinnung bis zur Prozessierung), so genannte Metadaten, dienen als wichtige Informationsquelle sowohl für die Erzeuger selbst als auch für die Drittanwender. Die Internationale Organisation für Normung (ISO) stellt Standards zur Beschreibung von geographischen Metadaten zur Verfügung. In der Europäischen Union (EU) ist die Richtlinie zur Schaffung einer Geodateninfrastruktur in der Europäischen Gemeinschaft, genannt INSPIRE (Infrastructure for Spatial Information in the European Community), ein wichtiger Leitfaden für die Beschreibung von Geodaten.

Das Ziel dieser Diplomarbeit ist es, Methoden für die Verwaltung von Produkten und den anfallenden Metadaten zu untersuchen und diese Informationen für Drittanwender, unter Berücksichtigung der erwähnten Standards, bereitzustellen. Die Herangehensweise ist basierend auf der Implementierung von Standard-konformen XML-Dateien, und der Entwurf einer relationalen Datenbank die zur Speicherung und Verwaltung der Metadaten dient.

Danksagung

Die vorliegende Diplomarbeit wurde am Institut für Photogrammetrie und Fernerkundung (IPF) an der TU Wien verfasst. Ich möchte mich bei allen Kollegen und Kolleginnen für ihre Unterstützung, und vor allem für das sehr angenehme und fruchtbare Arbeitsklima bedanken. Besonders bedanken möchte ich mich bei meinem Betreuer Wolfgang Wagner, der mir die Ausarbeitung dieses für das Institut neuen Themas ermöglicht hat. Weiters danke ich Richard Kidd für die sowohl sprachliche als auch inhaltliche Unterstützung beim Verfassen der Arbeit. Ein besonderer Dank gilt meinem Kollegen Sebastian Hahn, der meine zahlreichen wissenschaftlichen und weniger wissenschaftlichen Fragen stets profund zu beantworten vermochte.

Ein großer Dank gilt meinen Eltern, die mich während des Studium stets unterstützt haben, und ohne die dieser Lebensabschnitt wohl nicht so problemlos abgelaufen wäre. Ich möchte auch meinen Großeltern für die zahlreichen Fahrten von meinem Heimatort im wunderschönen Waldviertel bis zum Bahnhof Krems ein Dankeschön aussprechen.

Schließlich möchte ich auch noch meinen Studienkollegen und Freunden danken, die mir nicht nur immer helfend zur Seite gestanden sind, sonder vielmehr auch dafür gesorgt haben, dass der Spaß in meinem Studium nie zu kurz gekommen ist.

Contents

List of Figures

List of Tables

Listings

List of Acronyms

ASCAT	Advanced Scatterometer on board METOP
CVS	Content Management Systems
CVS	Concurrent Versions Systems
DB	Database
DBMS	Database Management System
DBS	Database System
DOI	Digital Object Identifier
ESA	European Space Agency
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
GHRSST	Group for High-Resolution Sea Surface Temperature
GML	Geography Markup Language
INSPIRE	Infrastructure for Spatial Information in Europe
IPF	Institute of Photogrammetry and Remote Sensing, Vienna University of Technology
ISO	International Organization of Standartisation
LIDAR	Light Detection And Ranging
METOP	Meteorological Operational Satellite
RADAR	Radio Detection And Ranging
SAR	Synthetic Aperture Radar
SCAT	Scatterometer
SSM	Surface Soil Moisture

SWI	Soil Water Index
WARP	Water Retrieval Package
WMO	World Meteorological Organization
XML	Extensible Markup Language

Chapter 1

Introduction

First, let us define what remote sensing is. One definition says that remote sensing is "the collection of information about an object without making physical contact with it" (*Rees*, 2001). Every human being operates remote sensing, even if it is not aware of that. Each person is born with three remote sensing senses, vision (sight), audition (hearing) and olfaction (smell). For each sense, we have specific recording instruments. We see with our eyes, we hear with our ears and we smell with our nose. Although these instruments record a huge amount of data, this information would be useless if we would not have a processor, our brain, which prepares the information in a way that we can interpret it.

Since the humans are inquisitive beings, they started building artificial remote sensing devices to investigate their surroundings, like for example the photo camera. 120 years after the first permanent photograph ever was taken (figure 1.1 left), the first photo of the earth viewed from space, was shot (figure 1.1 middle)(*Newhall*, 1982; *Reichhardt*, 2006).

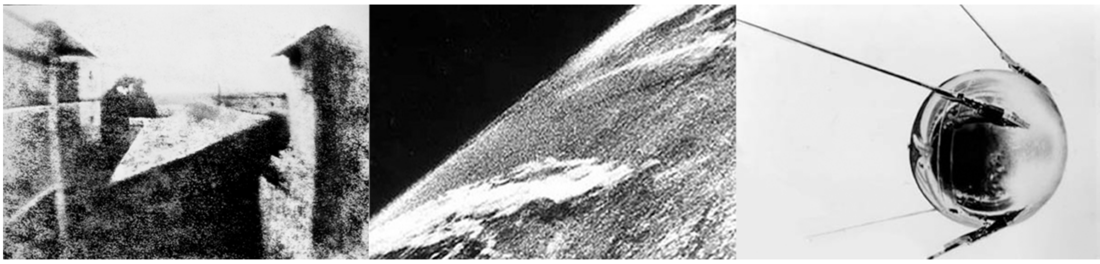


Figure 1.1: ***Left:*** *First successful permanent photograph, "View from the Window at Le Gras", created by Nicéphore Niépce in 1826 (Newhall, 1982)* ***Middle:*** *First photograph from space, taken from an altitude of 65 miles by a 35-millimeter motion picture camera riding on a V-2 missile in 1946 (Reichhardt, 2006)* ***Right:*** *First satellite put into Earth's orbit, Sputnik 1, launched in 1957 (Siddigi, 2003)*

Another decade later, the Soviet Union launched *Sputnik 1*, the first satellite ever put into the Earth's orbit. Not only did this cause the *Sputnik Crisis* (*Clowse*, 1981), it allowed great opportunities in the field of remote sensing. The installation of remote sensing instruments (such as for example multispectral cameras, scatterometers and radiation meters) on satellites allowed the possibility to observe the earth at a much greater scope. While the first satellite mission were driven by military interests, *Explorer 7*[1] was the first successfully launched and fully operable satellite that was solely used for observing the Earth's climate (*SSEC*, 2009). It was carrying an apparatus for comprehensive spatial and temporal monitoring of total cosmic-ray intensity, geomagnetically trapped corpuscular radiation, and solar protons (*Ludwig and Whelpley*, 1960). The functionality of these instruments was based on the same principle: to collect natural radiation that is emitted or reflected by an object. This type is called passive remote sensing. Coming back to the human body, the eyes, the ears and the nose are in this context passive remote sensing sensors.

In contrast to passive remote sensing, active remote sensing aims to collect reflected or backscattered radiation which is emitted by the instrument itself. This principle is for example applied by RADAR and LIDAR, where the time delay between emission and return of a signal is measured (*Skolnik*, 1990).

Both active and passive remote sensing instruments collect a huge amount of data. Analogue to the human perceptual apparatus, this data is useless if it cannot be interpreted properly. Therefore a certain processing has to be performed in order to prepare it for further interpretation. A common way to remove disturbing influences (e.g. noise) is filtering. This method is comparable to the ability of human beings to focus on specific impressions, whereas minor undesired impressions are only noticed subconsciously (it is even said that some people are able to only hear what they want to hear). In the following further processing steps (e.g. georefencing) have to be applied in order to get a usable product. This product usually contains Earth located values converted to geophysical variables. Figure 1.2 shows an example of filtering (a) and a detail of a final product (b).

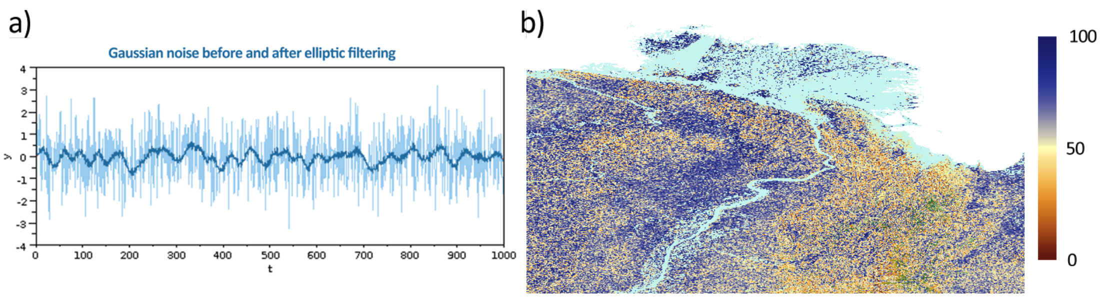


Figure 1.2: ***a)*** *Example of filtering: gaussian noise (light blue) and the signal after elliptic filtering (dark blue)* ***b)*** *Example product: ASAR SSM weekly average, 1 km resolution; Blue areas represent high percentage and red areas represent low percentage of surface soil moisture.*

[1] It is still in orbit: `http://usspaceobjectsregistry.state.gov/registry/dsp_DetailView.cfm?id=16`

Nowadays, numerous satellites are observing the Earth from its orbit, carrying a multiplicity of active and passive remote sensors. The characteristics of these sensors, for example the frequency of the emitted radiation or the ground sampling distance, can vary depending on their design and further on their objective. This enables a wide range of applications and products, and moreover, it brings a huge amount of metainformation with it.

The objective of this thesis is on the one hand to discuss the characteristics of satellite based remote sensing products and their attributes, and to investigate methods to handle the resulting metadata on the other hand. A further aim is to explore ways of how to prepare and provide this metadata, taking into account to achieve the largest possible user range. The approach of this thesis is to provide metadata conformant to existing international standards and data formats.

1.1 Structure of Work

This chapter started with an introduction to remote sensing. Chapter 2 will give an overview of metadata and of corresponding standards for geographic information. A more detailed view on the characteristics and attributes of satellite remote sensing products will be given in Chapter 3, which further discusses the issues of product versioning and naming. Chapter 4 shows how to implement metadata conformant to the standards discussed in chapter 2. Finally, chapter 5 will have a look into how to build up a relational database system which can be used to store and maintain the product metadata. It further demonstrates a possible schema for a metadata information system.

Chapter 2

Metadata

2.1 About Metadata

To get a better understanding of metadata, let us first ask the question: What is metadata? A common definition of metadata is "data about data". Metadata is structured information that describes other information or information services. With well defined metadata, users should be able to get basic information about data, without the need to have knowledge about its entire content.

Think of a steel can (Figure 2.1). Without any label, it is impossible to get information about its content, unless you open it. If the can doesn't have a built in opener, you also need the right tool to open it (Figure 2.2). Imagine a large stack of cans (Figure 2.3). To open all the cans would take some time and take a lot of effort.

Figure 2.1: Steel Can

Figure 2.2: Can Opener

Figure 2.3: Can Pyramid

A label would not only inform you about the content, it usually provides additional information, for example information about the manufacturer, the expiry date, a contact adress etc. All the information the label contains is metadata.

This short example shows two advantages of the use of metadata: it provides information about a certain product in advance and it can save effort.

Example using remote sensing data

The previous example very generally shows the purpose of metadata. Now we will go more into detail by demonstrating an example of metadata for a remote sensing dataset.

2007	1	1	21	0.0	27.0	0.0
2007	1	2	10	0.0	26.5	0.0
2007	1	3	21	18.0	22.5	0.0
2007	1	4	11	11.5	23.5	0.0
2007	1	5	22	42.0	20.5	0.0
2007	1	7	10	21.5	22.0	0.0
2007	1	8	11	32.5	21.5	0.0
2007	1	9	11	64.0	21.5	0.0
2007	1	10	22	99.0	28.5	0.0
2007	1	12	10	34.5	21.0	0.0
2007	1	13	11	90.5	26.5	0.0
2007	1	14	11	80.0	24.0	0.0
2007	1	15	22	47.5	20.5	0.0
2007	1	16	10	44.0	20.0	0.0
2007	1	17	10	23.5	22.0	0.0
2007	1	18	11	50.0	20.5	0.0
2007	1	20	22	69.0	22.0	0.0
2007	1	21	10	70.0	22.0	0.0
2007	1	23	11	45.5	20.0	0.0
2007	1	24	22	41.5	20.5	0.0
2007	1	25	21	44.5	19.5	0.0
2007	1	26	10	21.0	21.5	0.0
2007	1	28	11	60.0	20.5	0.0
2007	1	29	22	35.5	20.0	0.0
2007	1	30	21	100.0	31.5	0.0
2007	1	31	10	100.0	29.5	0.0

Figure 2.4: Sample dataset without description

Figure 2.4 shows a dataset with several columns filled with numeric values. Without having knowledge about the meaning of these columns, it is nearly impossible to figure out what the given numbers stand for. So we first need column headings which will be the first meta information in this example.

Year	Month	Day	Hour	SSM	ERR	SSF
2007	1	1	21	0.0	27.0	0.0
2007	1	2	10	0.0	26.5	0.0
2007	1	3	21	18.0	22.5	0.0
2007	1	4	11	11.5	23.5	0.0
[illegible]	1	5	22	[illegible]	20.5	0.0

Figure 2.5: Sample dataset with column headings

With column headings as given in Figure 2.5, we now know that the values in columns 1-4 provide information about a certain date, and that the columns with heading SSM, ERR and SSF provide some data relevant to that date. The information about what these acronyms stand for, and a description of what they actually mean, is further meta information which is required to be able to interprete the given values. In our case, SSM stands for "Surface

Soil Moisture", ERR for "Error" (Noise of soil moisture) and SSF for "Surface State Flag".

To be able to properly work with this dataset, information about its geographic location, its spatial and temporal extent are of importance. Further information of interest, besides many others, for example would be information about its lineage and processing steps. Some of this information can already be given in the file name, if it follows a certain naming convention. In this example, the file is named as followed:

`TUW_ASCAT_SSM_W52_gp2241731.dat`

With knowledge about the naming convention, it is possible to obtain relevant information from the filename:

TUW	Name of the organisation
ASCAT	Sensor name
SSM	Product
W52	Processing software
gp2241731	Location information

Table 2.1: Example filename for an ASCAT dataset.

With this information, we can presume that the given dataset is an SSM product processed by WARP version 5.2 at the Vienna institute of Techonology (TUW). The data are recorded by ASCAT[1] (Advanced Scatterometer) and are valid for the Grid Point 2241731. Knowledge of the file naming convention and the attributes in the file name also requires some meta information first.

2.1.1 Metadata storage

Meta information can be provided in severel ways, for example:

- in a product guide
- in a seperate file
- in a database
- in the file header of each dataset

More general, metadata can be stored either internally or externally to the dataset.

[1] http://www.esa.int/esaLP/SEMBWEG23IE_LPmetop_0.html

- **Internal storage** allows to transfer your data together with its metadata which ensures that metadata is always available. On the other hand, this method enlarges the size of each file and further can lead to high redundancy.

- **External storage** allows bundling metadata, for example in a database, which facilitates searching. This method avoids redundancy but it must be ensured that the linkage between the data and metadata is always set correctly. A common way to achieve this is to set a key. This key could be for example the file name of the data file and the metadata file. This gets trickier as the number of files increases or metadata is stored externally (e.g. in a database). A solution would be to set a Unique Identifier for both data and files and metadata.

Product guides, seperate metadata files and a databases are examples for external metadata, while metadata given in the file header is internal metadata (embedded metadata).

A very common data format for external metadata is XML. XML (Extensible Markup Language) is a markup language used to illustrate hierarchically structured data as text-data. It is among others used for exchange of data between computer systems, in particular through the internet (*Bray*, 2006).

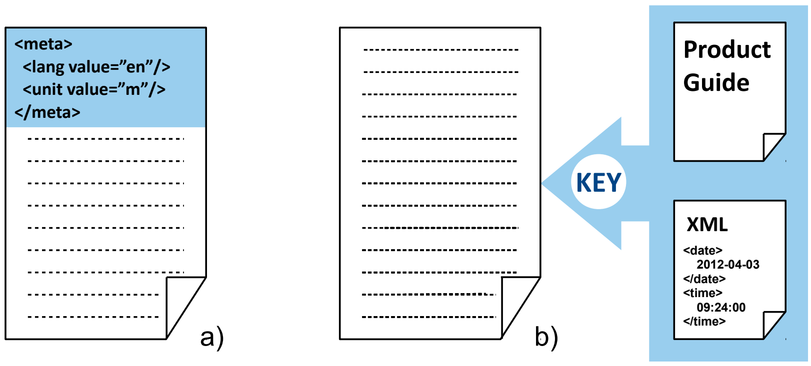


Figure 2.6: Methods of metadata storage: a) embedded metadata, b) external metadata

2.1.2 Benefits of metadata

A priori information

Metadata provides basic information about data in advance. It can be ensured that information about how to work with the data is available and therefore the risk of false data-handling can be reduced.

Management of data

The use of metadata can facilitate the management of large datasets. It can help to organize data, reduce the risk of duplication and makes it more effective to search in large data volumes.

Data Quality

Data quality is driven by a common set (and common understanding) of data standards, domain standards, business rules etc. If the systems follow the common standards (creating same checks, controls, table structure, field definitions...) there can be a big improvement on data quality. Metadata repositories provide the details on the data standards to follow and enforces the adherence to the standards as defined in the repository (*Exceution-MiH*, 2012).

Sharing Resources

The use of quality metadata which is compliant with a certain standard improves the opportunities to share resources with other organisations.

2.2 Metadata Standards

Metadata can be given in any form. To simplify the process of exchanging this information, it is highly advisable to adhere to a certain standard. In this work, ISO standards from the series ISO 19000 will be used to describe metadata.

2.2.1 International Organization for Standardization (ISO)

International Organization for Standardization (ISO) is the world's largest developer and publisher of International Standards. ISO is a network of the national standards institutes of 163 countries, one member per country, with a Central Secretariat in Geneva, Switzerland, that coordinates the system. ISO is a non-governmental organization that forms a bridge between the public and private sectors. On the one hand, many of its member institutes are part of the governmental structure of their countries, or are mandated by their government. On the other hand, other members have their roots uniquely in the private sector, having been set up by national partnerships of industry associations. ISO was founded in February 1947 and has since then developed international standards in nearly all areas.

Because "International Organization for Standardization" would have different acronyms in different languages ("IOS" in English, "OIN" in French for Organization internationale de normalisation), its founders decided to give it also a short, all-purpose name. They chose

"ISO", derived from the Greek isos, meaning "equal". Whatever the country, whatever the language, the short form of the organization's name is always ISO. (extracted from *ISO*, 2012)

2.2.2 ISO-Standard ISO 19115

The ISO 19100 series is a set of standards for geographic information. It establishes a structured set of standards for information concerning objects or phenomena that are directly or indirectly associated with a local relative to the Earth (*WMO*, 2008). It contains standards for describing spatial and temporal schemata (ISO 19107 and ISO 19108), geographic information services (ISO 19119) and Geography Markup Language (GML, ISO 19136), a data format based on XML used for the interexchange of spatial data.

The standard for describing geographic metadata is ISO 19115. It defines how to describe geographical information and associated services, including contents, spatial-temporal purchases, data quality, access and right to use. There are more than 400 metadata elements, but only 22 core elements (*ISO*, 2003, p. 16). If there is a need to add more elements, it is possible to define a community profile for an extension of metadata information (cf. Figure 2.7).

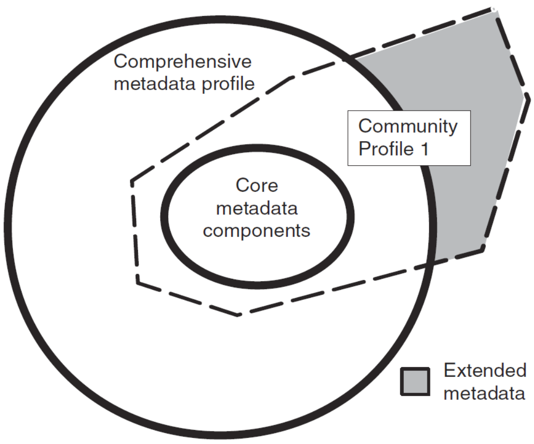


Figure 2.7: Composition of ISO 19115; figure from ISO (2003)

ISO-Standard ISO 19115-2

ISO 19115-2 extends the existing geographic metadata standard by defining the schema required for describing imagery and gridded data. It provides information about the properties of the measuring equipment used to acquire the data, the geometry of the measuring process employed by the equipment, and the production process used to digitize the raw data. This extension deals with metadata needed to describe the derivation of geographic information from raw data, including the properties of the measuring system, and the numerical methods and computational procedures used in the derivation. The metadata required to address coverage data in general is addressed sufficiently in the general part of ISO 19115. (extracted from *ISO*, 2009)

Core Elements

Figure 2.8 shows the core elements of the ISO 19115 Standard for geographic datasets and how it is hierarchically structured. The elements of ISO 19115 can have different obligations: mandatory, conditional and optional. If an element is mandatory, it has to be given in any case. Conditional elements can become mandatory, if certain prerequisites are met. Optional are free to choose, but once they are specified, they may again mandate further elements.

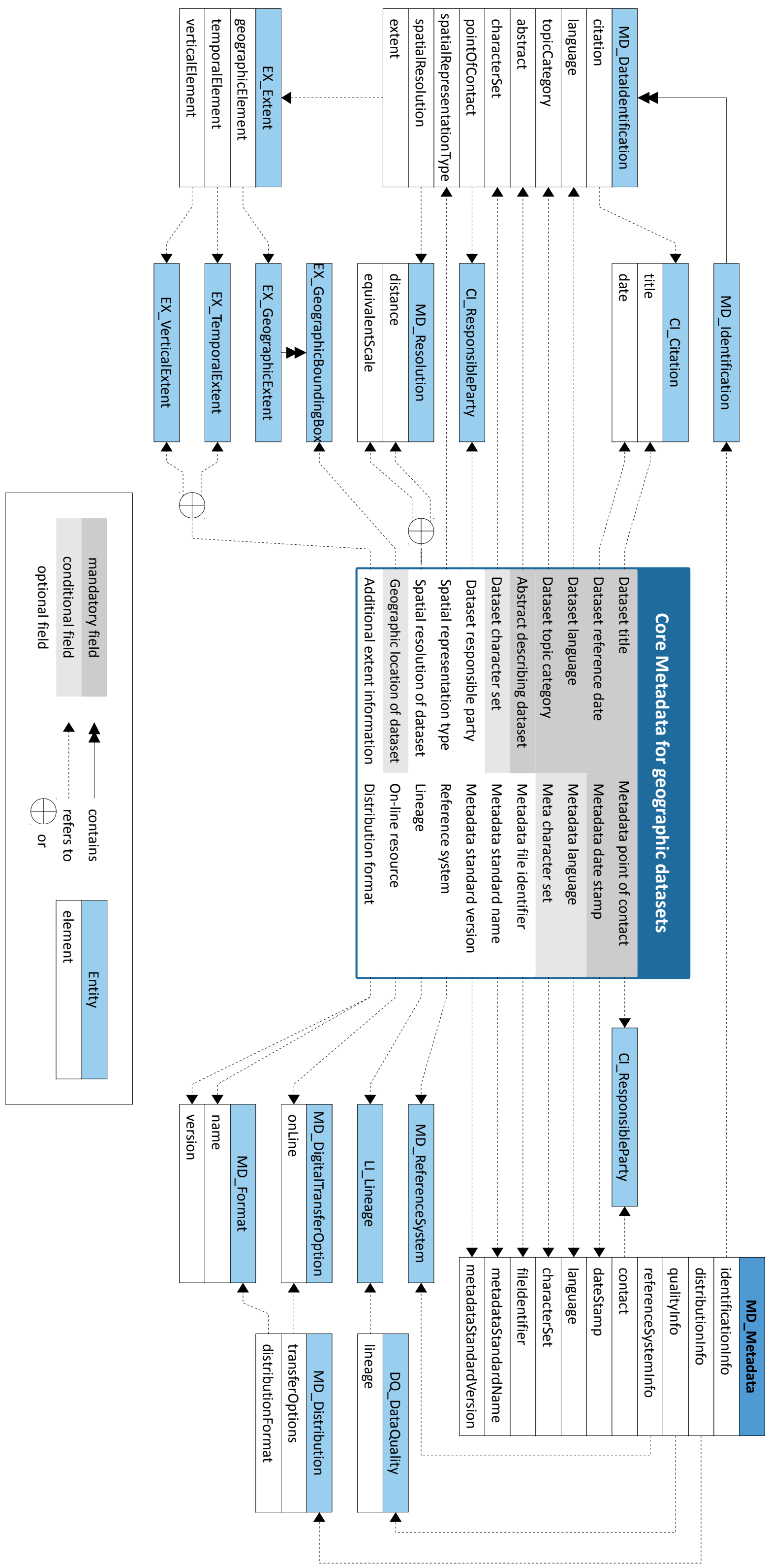


Figure 2.8: ISO 19115 core metadata for geographic datasets

2.2.3 INSPIRE

In March 2007, the European Parliament and the Council of Euro enacted a directive for the creation of a geodata-infrastructure in the European Community, called INSPIRE (Infrastructure for Spatial Information in Europe)(*INSPIRE*, 2007).

The INSPIRE directive aims to create a European Union (EU) spatial data infrastructure. This will enable the sharing of environmental spatial information among public sector organisations and better facilitate public access to spatial information across Europe. INSPIRE is based on a number of common principles (extracted from *INSPIRE*, 2012):

- Data should be collected only once and kept where it can be maintained most effectively.
- It should be possible to combine seamless spatial information from different sources across Europe and share it with many users and applications.
- It should be possible for information collected at one level/scale to be shared with all levels/scales; detailed for thorough investigations, general for strategic purposes.
- Geographic information needed for good governance at all levels should be readily and transparently available.
- Easy to find what geographic information is available, how it can be used to meet a particular need, and under which conditions it can be acquired and used.

In the context of metadata for spatial data and spatial data services, the standards ISO 19115[2], ISO 19119[3], and ISO 15836[4] have been identified as important standards by the European Parliament and the Council of Euro (*INSPIRE*, 2010). All elements of the INSPIRE directive can be expressed by these standards. Figure 2.9 shows the INSPIRE elements and their linkage to the standard ISO 19115.

[2] ISO standard for geographic information metadata
[3] ISO standard for geographic information services
[4] Dublin-Core-Standard; The dublin core is a collection of simple and standardized conventions for describing documents and other objects on the Internet.

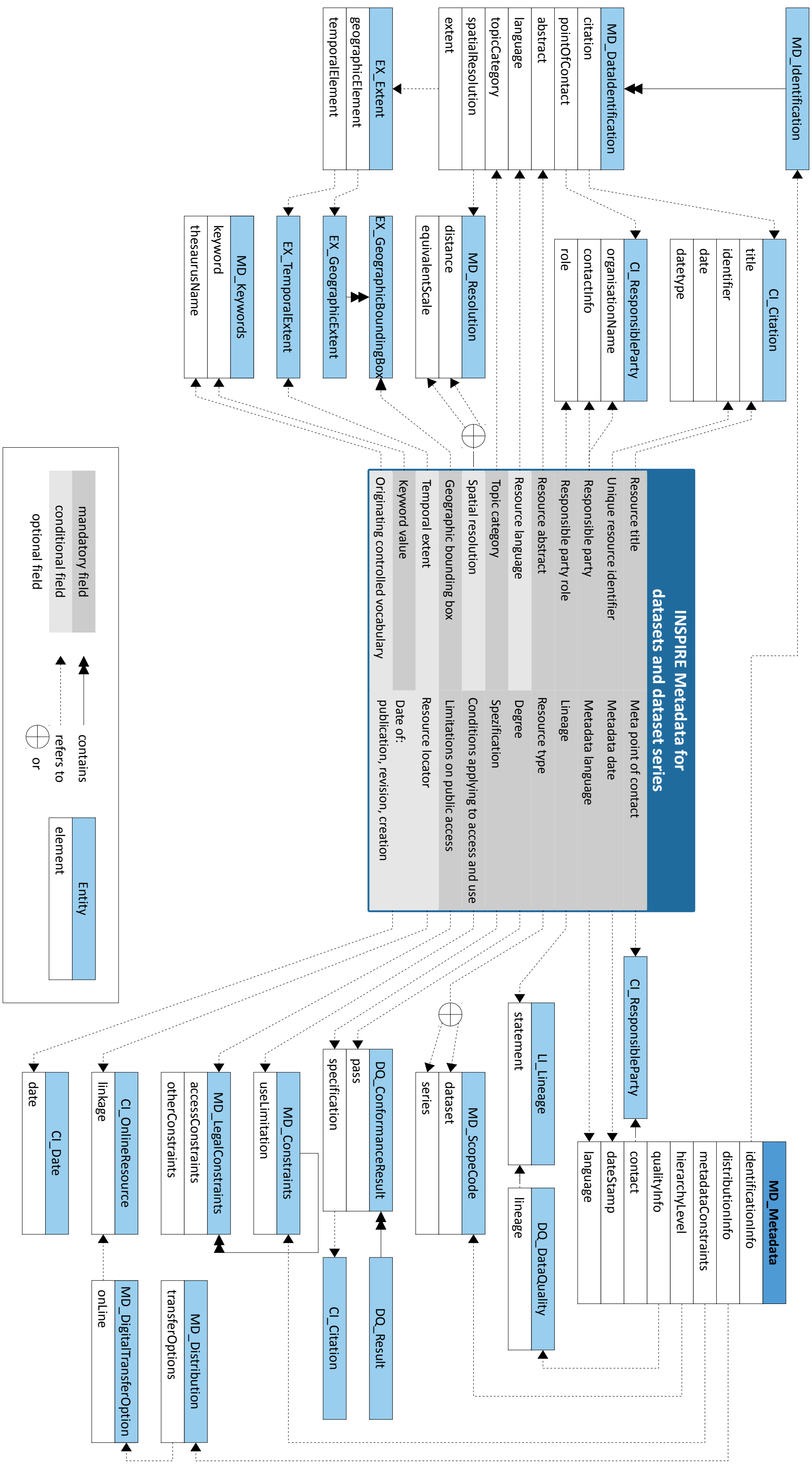


Figure 2.9: *INSPIRE metadata for datasets and dataset series*

2.2.4 Comparison of INSPIRE and ISO 19115 core elements

In the following, we compare the core requirements of ISO 19115 to the requirements of INSPIRE element by element. The Character in brackets next to the name shows the obligations of this element. In this case, M stands for "mandatory", C for "conditional" and O for "optional". The following tables are extracted from the INSPIRE Implementation Rules (*INSPIRE*, 2010).

Identification of the Dataset

ISO 19115	INSPIRE	Definition
Dataset title (M)	Resource Title (M)	Name by which the cited resource is known
Abstract describing the dataset (M)	Resource Abstract (M)	Brief narrative summary of the content of the resource(s)
On-line resource (O)	Resource Locator (C)	Location (address) for on-line access using a Uniform Resource Locator adress or similar addressing scheme.
Dataset language (M)	Resource Language (C)	Language(s) used within the datasets
	Resource Type (M)	Scope to which metadata applies
	Unique Resource Identifier (M)	Value uniquely identifying an object within a namespace
Dataset character set (C)		Full name of the caracter coding standard used for the dataset

Table 2.2: Elements for identification of a dataset

As table 2.2 shows, INSPIRE additionally demandes a Resource Type and a Unique Resource Identifier. Further, a Resource Locator is mandated if a URL is available to obtain more information on the resources and/or related services. Possible values for Resource Type in the scope of the directive are "dataset", "series" and "service". This work focuses mainly on "dataset" and "series".

ISO 19115 demands information about the dataset character set if ISO/IEC 10646-1[5] is not used. Information about the dataset language has to be given in ISO 19115, even if the

[5] ISO standard for Universal Character Set

dataset does not include any textual information. INSPIRE only mandates this information if the dataset includes text.

Classification

ISO 19115	INSPIRE	Definition
Dataset topic category (M)	Topic Category (M)	Main theme(s) of the dataset
	Keyword (M)	Commonly used words or formalised words or phrases used to describe the subject.
	Originating controlled vocabulary (C)	Name of the fromally registered thesaurus or a similar authoritative source of keywords.

Table 2.3: Elements for classifying a dataset

The INSPIRE Implementing rules for metadata mandate the presence of at least one keyword. It shall describe the relevant INSPIRE spatial data theme originating from the general environmental multilingual thesaurus (GEMET)[6]. In addition to the INSPIRE spatial data theme, also other keywords might be added. These may be described as a free text or may originate from any controlled vocabulary. Information about the controlled vocabulary is mandated by INSPIRE.

Geographic location

ISO 19115	INSPIRE	Definition
Geographic location of the dataset (C)	Geographic Bounding Box (M)	Coordinates of the limit of the dataset extent, expressed in longitude in decimal degrees (positive east)

Table 2.4: Geographic location of a dataset

INSPIRE is more restrictive, a geographic bounding box is mandated. ISO 19115 only mandates this information if the hierarchy level ("Resource Type" in INSPIRE) is "dataset". Both INSPIRE and ISO 19115 demand four coordinates, the west- and east-bound longitudes and the south- and north-bound latitudes in decimal values.

[6] http://www.eionet.europa.eu/gemet/inspire_themes

Temporal reference

ISO 19115	INSPIRE	Definition
Additional extent information (O)	Temporal Extent (C)	Time period covered by the content of the dataset.
Dataset reference date (M)	Temporal Reference (M)	Reference date for the cited resource.

Table 2.5: Temporal references for a dataset

The INSPIRE implementing rules for metadata require at least one temporal reference chosen from one of these four categories: temporal extent, data of publication, date of last revision or the date of creation. ISO 19115 is more demanding and requires that at least one of the following is provided: date of publication, the date of last revision or the date of creation. Therefore, whilst providing a temporal extent would suffice to satisfy the INSPIRE implementing rules for metadata it is not enough to be compliant with ISO 19115. To be compliant with ISO 19115 it is necessary to use at least one among date of publication, date of last revision, or the date of creation.

Quality and Validity

ISO 19115	INSPIRE	Definition
Lineage (O)	Lineage (M)	General explanation of the data producer's knowledge about the lineage of a dataset.
Spatial resolution of the dataset (O)	Spatial Resolution (C)	Ground sample distance or level of detail expressed as the scale denominator of a comparable hardcopy map or chart.
Spatial representation type (O)		Digital representation of spatial information in the dataset.
Distribution format (O)		Provides a description of the format of the data to be distributed.

Table 2.6: Elements describing Quality and Validity

While lineage information is optional in ISO 19115, a general lineage statement is mandated by INSPIRE. Apart from describing the process history, the overall quality of the dataset (series) should be included in the lineage metadata element. This statement should contain any quality information required for interoperability and/or valuable for use and evaluation of the data set (series).

Information about spatial resolution is mandatory in INSPIRE if an equivalent scale or a resolution distance can be specified. This information is optional in ISO 19115. The ISO 19115 core additionally provides information about the spatial representation type and the distribution format, both elements are optional.

Conformity

ISO 19115	INSPIRE	Definition
	Degree (M)	Indication of the conformance result.
	Specification (M)	Citation of the product specification or user requirement against which data is being evaluated.

Table 2.7: Conformity elements

In conformance to INSPIRE, the metadata shall include information on the degree of conformity with the implementing rules. These rules for metadata define three degrees of conformity: conformant, not conformant and not evaluated. When the conformity to an INSPIRE Specification has been evaluated, it shall be reported as a domain consistency element in ISO 19115 metadata. In that case, if the evaluation has passed, the degree is conformant, otherwise it is not conformant. Both, degree and specification are mandatory in INSPIRE.

Constraints related to access and use

ISO 19115	INSPIRE	Definition
	Limitations on public access (M)	Restrictions and legal prerequisites for accessing and using the resource or metadata.
	Conditions for access and use (M)	Restrictions on the access and use of a resource or metadata.

Table 2.8: Elements describing Constraints related to acces and use of data

There are two major requirements expressed in Directive 20007/2/EC (INSPIRE) in terms of documentation of the constraints as part of the metadata:

- The conditions applying to access and use of the resource, and where applicable, the corresponding fees.
- The limitations on public access.

INSPIRE mandates at least one ISO 19115 element for each.

Responsible organisation

ISO 19115	INSPIRE	Definition
Dataset responsible party (O)	Responsible party (M)	Identification of, and means of communication with, persons and organizations associated with the resources.
	Responsible party role (M)	Function performed by the responsible party.

Table 2.9: Information about the responsible organisation

INSPIRE is more demanting by mandating both the name of the organisation, and a contact e-mail address. Further, INSPIRE mandates information about the function of the responsible party.

Metadata on metadata

Both INSPIRE and ISO 19115 mandate a point of contact and a date for the metadata. As for the element "responsible organisation", INSPIRE is more demanding by mandating a contact e-mail address in addition. ISO 19115 on the other hand is more restrictive concerning the metadata date. This element shall contain the date that the metadata was created while INSPIRE may contain the date when the metadata record was created or uploaded.

ISO 19115 mandates information about the metadata character set if ISO/IEC 10646-1 is not used or if it is not defined by the encoding. Further, the ISO 19115 core includes a metadata file identifier, metadata standard name and standard version, which are optional elements.

In terms of metadata language, INSPIRE is more restrictive by mandating metadata language even if it is defined by the encoding.

2.2.4.1 Conclusion

The comparison shows that the ISO 19115 Core and INSPIRE differ in both number and obligation of demanded elements. Generally speaking, INSPIRE mandates more information whereas ISO 19115 is more demanding concerning metadata on metadata. This leads

ISO 19115	INSPIRE	Definition
Metadata point of contact (M)	Metadata point of contact (M)	Party responsible for the metadata information.
Metadata date stamp (M)	Metadata date (M)	Date that the metadata was created.
Metadata language (C)	Metadata language (M)	Language used for documenting metadata.
Metadata character set (C)		Full name of the character coding standard used for the metadata set.
Metadata file identifier (O)		Unique identifier for the metadata file.
Metadata standard name (O)		Name of the metadata standard (including profile name) used.
Metadata standard version (O)		Version of the used metadata standard.

Table 2.10: Metadata on Metadata

to the conclusion, that conformance of an ISO 19115 metadata set to the ISO 19115 Core does not all guarantee the conformance to INSPIRE and vice versa (*INSPIRE*, 2010). In order to achieve both ISO 19115 and INSPIRE conformance, all elements, either mandated by ISO 19115 or INSPIRE, should be given.

Chapter 3

Remote Sensing Products

In chapter 2 standards for geographic data have been presented. Both the standard ISO 19115 and the EU initiative "INSPIRE" aim to cover metadata for geographic products in general. This chapter will focus on satellite remote sensing products and have a closer look at their attributes and metadata.

Generally speaking, a remote sensing product can be described as a collection of datasets representing a geophysical parameter, obtained from a certain source, processed by a certain software for a certain purpose. They can be divided into various categories, differ in level of detail and coverage, or have a different data structure.

3.1 Product Attributes

Category and Geophysical Parameter

Products can be classified into several categories, for example Soil Moisture- or Freeze Thaw products. They can further be determined by the geophysical parameters they represent, such as for example Surface Soil Moisture (SSM) or Soil Water Index (SWI). In some cases, products can contain more than one parameter. An example was given in figure 2.4 in section 2.1, there are SSM, ERR (noise of soil moisture) and SSF (Surface State Flag), where SSM is the main product and ERR and SFF are auxiliary parameters.

Product Level

Products can exist in different levels, depending on their processing status. The definition of these levels can vary from organisation to organisation. Table 3.1 shows a comparison between definitions from NASA (extracted from NASA's Earth Science Reference Handbook (*Parkinson et al.*, 2006)) and EUMETSAT (*EUMETSAT*, 2012).

Comparison shows that both level definitions are specified for the processing chain applied by each organisation, hence they differ slightly in both quantity and detail. But it further shows that both share a basic structure.

	NASA	EUMETSAT
Level 0	Reconstructed, unprocessed instrument/payload data at full resolution; any and all communications artifacts, e.g., synchronization frames, communications headers, duplicate data removed.	Raw data after restoration of the chronological data sequence for each instrument, i.e. after demultiplexing of the data by instrument, removal of any data overlap due to the data dump procedure and relevant quality checks. Raw instrument data information (telemetry packets) is maintained during this process.
Level 1a	Reconstructed, unprocessed instrument data at full resolution, time-referenced, and annotated with ancillary information, including radiometric and geometric calibration coefficients and georeferencing parameters, e.g., platform ephemeris, computed and appended but not applied to the Level 0 data.	Instrument data in full resolution with radiometric and geometric (i.e. Earth location) calibration computed and appended but not applied.
Level 1b	Level 1a data that have been processed to sensor units (not all instruments have Level 1b data products).	Calibrated, Earth located and quality controlled product, in the original pixel location, and packaged with needed ancillary, engineering and auxiliary data.
Level 1c		In case of the IASI spectra, level 1b data after application of the apodization function.
Level 2	Derived geophysical variables at the same resolution and location as the Level 1 source data.	Earth located values converted to geophysical parameters, at the same spatial and temporal sampling as the level 1b and 1c data.
Level 3	Variables mapped on uniform space-time grids, usually with some completeness and consistency.	Gridded point geophysical products on a multi-pass basis, or resampled products (applicable for some Satellite Application Facility (SAF) products).
Level 4	Model output or results from analyses of lower level data, e.g., variables derived from multiple measurements.	

Table 3.1: Comparison between Product Level definitions from NASA and EUMETSAT

Source

Product data can be recorded by passive and active microwave sensors on a satellite in a certain mode. Common instruments for data retrieval are Synthetic Aperture Radar (SAR) and Scatterometer (SCAT). Relevant information for instance would be the name of the satellite, the recording sensor and (if present) the recording mode used to obtain data. Generally speaking, source describes the method of obtaining data and furthermore the used instruments.

Satellite	**Sensor**	**Mode**
ENVISAT	ASAR	Global Mode
ALOS	PalSAR	ScanSAR
MetOp	ASCAT	High-res Mode

Table 3.2: Examples for sources of Remote Sensing Products

Resolution

The generic term *Resolution* describes the level of detail of the Product data. It can be divided into four attributes, the Spatial-, Temporal- and Radiometric Resolution and the Spatial- and Temporal Sampling Rate.

a) Spatial Resolution

The spatial resolution can be seen as the size of the pixel over land which is covered for each measurement. This value can vary from some few hundred meters (e.g. ENVISAT ASAR[1] 150 m) up to several kilometers (e.g. MetOp ASCAT[2] 50 km). Figure 3.1 demonstrates different resolutions with two Surface Soil Moisture (SSM) maps of the Lena-Delta Area in Siberia. Figure 3.1a shows the ASCAT SSM monthly average product with a resolution of 50 km and figure 3.1b shows the ASAR SSM weekly average product with a resolution of 1 km. It is clearly visible that the 1 km product provides greater level of detail, while the 50 km product looks pixelated at this zoom level.

b) Temporal Resolution

The temporal resolution represents an absolute value of a time period for which one measurement is valid. This can either be a snapshot (e.g. single measurement for a certain

[1] http://earth.esa.int/object/index.cfm?fobjectid=1525
[2] http://www.esa.int/esaLP/ESA266094UC_LPmetop_0.html

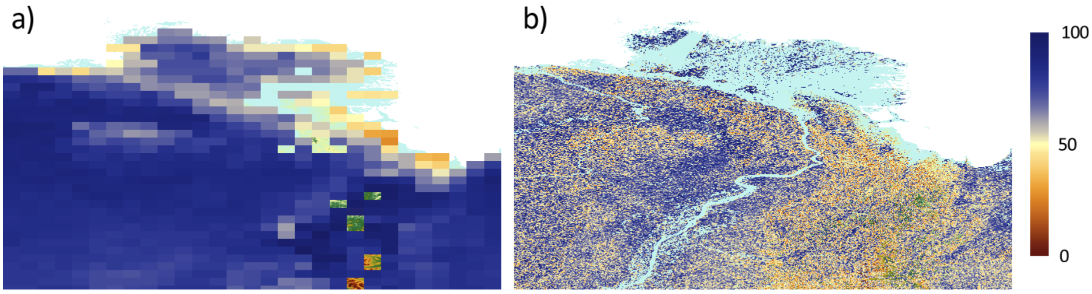


Figure 3.1: Surface Soil Moisture maps of Lena Delta Area
a) ASCAT SSM monthly average, 50 km resolution b) ASAR SSM weekly average, 1 km resolution; Blue areas represent high percentage and red areas represent low percentage of surface soil moisture.

point in time) or a value representing measurements aggregated over a certain time period (e.g. weekly average).

To demonstrate temporal resolution and temporal sampling for time series products, let us have again a look at the ASCAT SSM Product from the example in chapter 2.1. Each row in figure 2.4 represents a measurement at a certain point in time. The time distance between these measurements is the temporal sampling rate. This value shows variations by several hours up to more than day. Therefore, it is not possible to give an absolute value for the temoral sampling rate in this example, hence it is irregular. Since each measurement is a snapshot, the temporal resolution would be 0 hours. If you would now aggregate all measurements on a weekly basis and for example calculate the average, the temporal resolution would be seven days. If this is done for example in a weekly interval, the temporal sampling rate also is seven days.

An example for raster data is the Pan-Arctic Weekly Land Surface Temperature product from the ESA DUE Permafrost Project[3] which is available for all satellite overpasses within a seven-day period based on aggregation of daily products. It is available for each day based on a seven-day sliding time window. In this case, the temporal sampling rate is one day, with a resolution of seven days.

c) Radiometric Resolution

Radiometric resoltuion is the sensitivity of a sensor to incoming reflectance. It refers to the number of divisions of bit depth in data collected by a sensor, for example 255 for 8-bit and 65,536 for 16-bit. (*ESRI*, 2012)

[3] http://www.ipf.tuwien.ac.at/permafrost/

d) Spatial Sampling

The spatial sampling distance is the ground distance between each measurement points.

e) Temporal Sampling

The definition of temporal sampling can vary depending on the data structure of the product.

- For **time series** data, the temporal sampling rate shows the time interval between two measurements over the same point. This interval depends on factors like the orbital period of the satellite or the swath with of the recording sensor. For satellites in sun-synchronous polar orbit the temporal sampling rate usually is irregular whereas geostationary satellites continuously cover the same area (see figure 3.2).

- For **raster data** such as images, it describes the absolute value of the consecutive time interval in which the data is available. This definition is also valid for time series data averaged over a certain time on a regular basis.

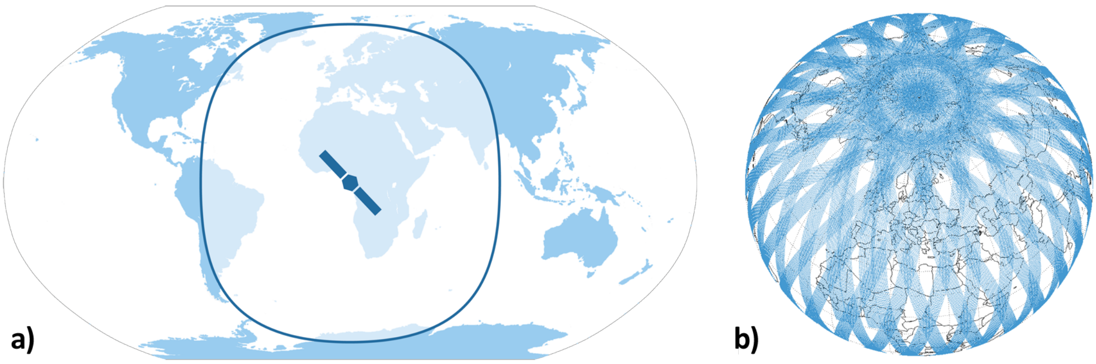


Figure 3.2: a) Geostationary Satellites continuously cover the same area
b) Daily coverage of MetOp satellite following a polar sun-synchronous orbit (figure from Hahn et al. (2012))

Extent

Each product is valid within a certain area. This area is called the extent or coverage and has a spatial and a temporal component. The temporal extent is the time period which is covered by the product data, the spatial component is the geographic area which is covered by the product, represented by a bounding box.

Data Structure and Format

Remote sensing products data can have either a raster or vector structure. Typical raster products are images, common vector products are for example time series, which are measured or calculated values over a period of time for a single point. Furthermore, the corresponding files can be stored in various formats, depending on how and for which purpose they are processed.

Processing

In order to be able to work with satellite data, several processing steps are necessary (e.g. removing duplicate data/noise). Hence, to achieve a certain Product Level (c.f. table 3.1) the use of a certain software is inevitable. Further, in some cases auxiliary data is essential to produce a certain product. Figure 3.3 shows a process flow applied by the WARP 6.0 software package. It demonstrates the interplay of raw data (external data), model paramters, auxilliary data and the applied processing steps (represented by the letters A-D) to produce a final product. In this example, the output is a SSM time series dataset obtained from raw scatterometer data at 25 km resolution.

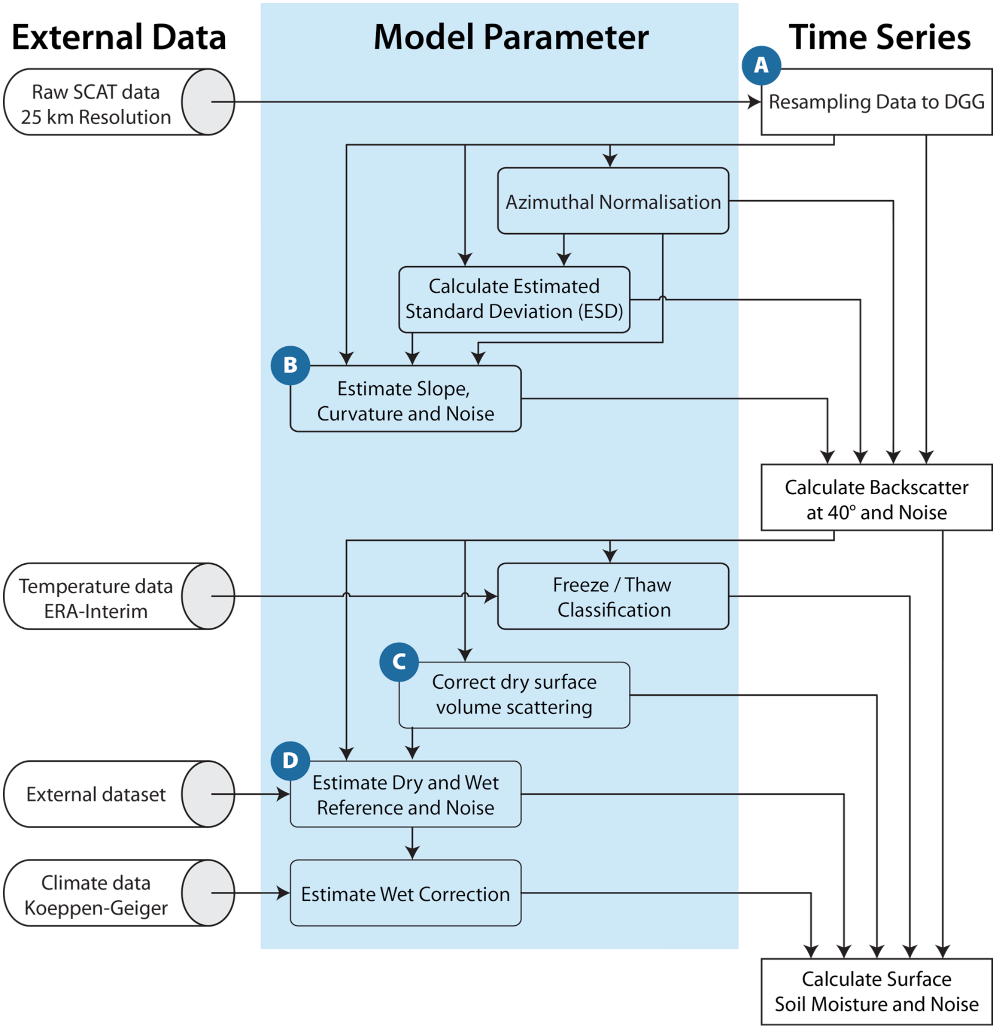


Figure 3.3: WARP 6.0 Data Flow Chart (A) Resampling Data to a Discrete Global Grid (B) Estimation of Slope Curvature Noise (C) Correction of Dry Surface Volume Scattering (D) Estimation of Dry and Wet Reference; Figure from Hahn et al. (2012)

3.2 Product Versioning

The interplay of data retrieval and data processing is a steady dynamic process. Processing software and packages like the Water Retrieval Package (WARP) are in continuous development and therefore, products can exist in several versions.

In general, a process step consists of three parts: input data, auxiliary data and a processing software. A typical work flow for producing remote sensing products is described in figure 3.4.

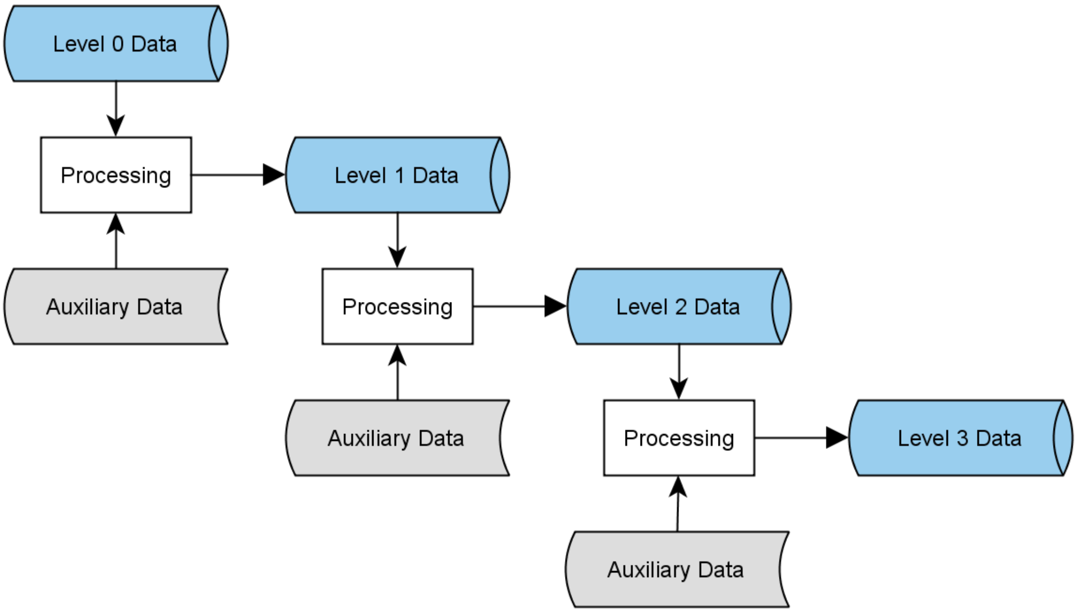


Figure 3.4: Low level to high level data processing

In order to facilitate reproduction of certain product versions, it is inevitable to have knowledge about all previous processing steps, including input data, auxiliary data and the applied processes. A possible approach is the use of revision control systems, for example CVS[4]. This system allows to replicate the version of the code repository which has been used to process the product. This can be achieved either by giving date (when the processing has taken place) or by the use of tags.

Once the desired repository version is restored, the user now needs to know the name of the procedure which is was used to process the product. If multiple processes were used, knowledge about the starting routine is inevitable. The prerequisite for this method is that all input and auxiliary data must exist in the same directory and in the same version as they were when they were used during processing. To meet this prerequisite may be tricky.

[4] `http://savannah.nongnu.org/projects/cvs`

Figure 3.5 shows the version history of the procedure *cal_sigma40_ascat.pro*. It shows, for example that version 1.5 is related to the tag *WARP5_2_FINAL*.

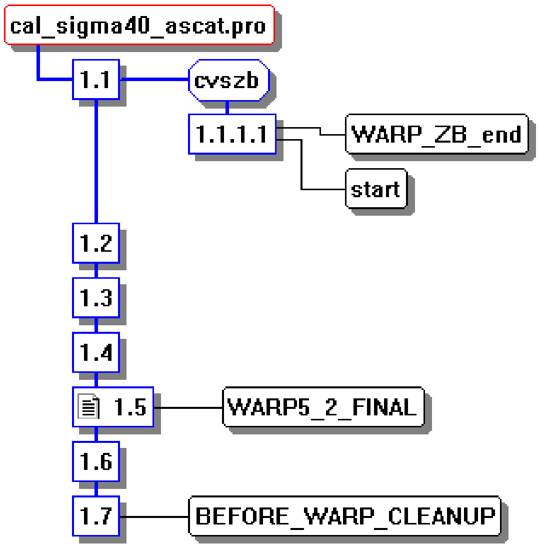


Figure 3.5: Version history of the procedure "cal_sigma40_ascat.pro", showing the versions (blue box) and the related tags (black box)

3.3 Product and File Naming

As mentioned in the example in chapter 2.1, the product- respectively file name is an important carrier of meta information. Not only should the name give a short description about the product, but also uniquely identify it. To avoid ambiguity, products and their corresponding files should be uniquely determined by their name, hence, either all attributes listed in chapter 3 should be represented in the name, or an unique identifier is used (e.g. Digital Object Identifier - DOI[5]). Nonetheless, the name should contain enough information to answer the questions:

- Where is the data from?
- What is its content?
- What is its format?

To keep the name short, acronyms representing the naming components should be used, as given in the example in chapter 2.1:

`TUW_ASCAT_SSM_W52_gp2241731.dat`

In this case, `TUW` indicates the origin of the data (TU Wien). The other parts give information about the content of the data. It is a SSM product, recorded by ASCAT, processed by WARP version 5.2 valid for the Grid Point 2241731. The file extension `dat` indicates that it is a data file, but does not give an allusion to the actual data format (it may contain data in text or binary format) (*FileInfo.com*, 2009).

Possible naming components may for example be (without warranty of completeness):

- Unique identifier
- Data type and format (may be sufficiently described already by the file extension)
- Recording instrument
- Platform (that carries the instrument)
- Processing level (see table 3.1)
- Processing software
- Product/Software version

[5] `http://www.doi.org/`

- Indicative date & time
- Temporal coverage information
- Temporal interval (e.g. daily, monthly)
- Location information
- Originator/Organisation; Informtation that states where the data originated from
- Project where the data originated from
- ...

File naming The file name has to uniquely identify each dataset within a certain dataset series. This can be achieved by setting naming components, whose values differ depending on the particular dataset. Let us have again a look at the example given above:

`TUW_ASCAT_SSM_W52_gp2241731.dat`

This file is part of a dataset representing SSM values for specific grid points. In this case the grid point information (gp2241731) is the unique identifier.

Product naming The product name has to cover all associated files, it can be seen as the name of the whole dataset series. Generally speaking it should contain all the information of the corresponding filenames, except the components that uniquely identify each file. A proper product name derived from the example given above for example would be:

`TUW_ASCAT_SSM_W52`

Naming conventions To facilitate product and file management, it is inevitable to harmonize the product names by implementing a file naming convention. These conventions exist in a huge number nowadays, each optimized for a certain purpose. The following example shows the file naming convention defined by the The World Meteorological Organization. In this convention coded naming components are used, in order to keep the filenames short.

```
pflag_productidentifier_oflag_originator_yyyyMMddhhmmss[_freeformat].type[.compression]
```

Listing 3.1: WMO file naming convention; from WMO (2009)

In this convention, *pflag* and *oflag* are strings that indicate how to decode the subsequent field. The *productidentifier* describes the product generally. The field *originagor* provides information that states where the data originated from. *yyyyMMddhhmmss* is a date and time stamp field. *freeformat* is an optional field containing further description of the data, which may be required by the given originator. *type* describes the format type of the data. If the data is compressed, the WMO file naming convention mandates the presence of the field *compression*, which specifies the used compression technique. For further information about this convention, see *WMO* (2009).

Chapter 4

Implementation of Metadata Standards

In chapter 2 the standards for geographic metadata were presented, chapter 3 discussed the attributes of remote sensing products and their version- and naming-management. In the following, the implementation of these standards for remote sensing products, using Extensible Markup Language (XML), will be demonstrated. For further information about the XML syntax see *Refsnes-Data* (2012), as it is not addressed in this work.

4.1 Guidance and Terminology

Geographic Metadata XML (gmd) Since ISO 19115 does not specify a certain encoding, implementations of geoinformation data can vary depending on the interpretation of the generator (*ISO*, 2009). To simplify the standardisation of implementation of metadata, the standard ISO 19139 defines rules for encoding using XML schemata. The outcome of this standard is "Geographic MetaData XML"-encoding (gmd-encoding), an XML based implementation of the ISO 19115 standard. The common prefix used to refer to that namespace within XML tags is *gmd*[1].

Geographic Markup Language grammar (GML) In some cases, the XML code is extended with GML grammar (cf. listing 4.12). GML is an XML grammar written in XML Schema for the description of application schemes as well as the transport and storage of geographic information modeled in accordance with the conceptual modeling framework used in the ISO 19100 series of International Standards and including both the spatial and non-spatial properties of geographic features (extracted from *ON*, 2009a, p. 2). The common prefix used to refer to that namespace within XML tags is gml[2].

Geographic Common XML (gco) The implementation of ISO 19115 requires the encoding of required basic types from ISO/TS 19103 (Geographic information – Conceptual

[1]The full gmd-namespace specifications can be found at `http://www.isotc211.org/2005/gmd/`
[2]The full GML-namespace specifications can be found at `http://www.isotc211.org/2005/gml/`

schema language) and conceptual elements from ISO 19118 (Encoding). As these elements are not specific to geographic metadata, a separate namespace is created to contain the XML schemes. The common prefix used to refer to that namespace within XML tags is *gco*[3]. (extracted from *ON*, 2009b, p. 62)

Metadata properties

Class The ISO 19115 standard can be seen as set of metadata classes (respectively packages) and their elements, code lists and enumerations. A metadata class is a set of elements, belonging to a particular topic (eg. data quality, lineage, etc.). In XML code, it is represented by a prefix, followed by an underscore and a name (eg. MD_Metadata). Table 4.1 gives an overview over the classes and their prefixes, used in this work.

Prefix	Information-type	Standard
CI	Citation	ISO 19115
DQ	Data Quality	ISO 19115
EX	Extent	ISO 19115
LE	Lineage Extended	ISO 19115-2
LI	Lineage	ISO 19115
MD	Metadata	ISO 19115
MI	Metadata for Imagery	ISO 19115-2
MX	Metadata – XML schema	ISO/TS 19139
QE	Data Quality Extended	ISO 19115-2

Table 4.1: List of prefixes of the various types of information. Table from ISO (2009)

Element An element is part of a class. It has a specific data type, like for example *codeList*, enumeration or character string. Further, it can represent its own class. Example being the element *characterSet* of the class *MD_Metadata*, which refers to the CodeList-class *MD_CharacterSetCode*. The full list of data types is given in the class *MD_DatatypeCode* in ISO 19115 (B.5.13).

Enumeration The conceptual schema language of the ISO 19100 series supports two enumerated types, *Enumeration* and *CodeList*. A class of type Enumeration can contain only simple attributes which represent the enumeration values (see ISO/TS 19103). The list of

[3] The full gco-namespace specifications can be found at `http://www.isotc211.org/2005/gco/`

values is not extensible. An enumeration, using the list item "geoscientificInformation" is encoded as followed:

```
<gmd:MD_TopicCategoryCode>geoscientificInformation</gmd:MD_TopicCategoryCode>
```

Codelist The second type of enumerations is *CodeList*, which is an extensible class. Table 4.2 shows an example of a CodeList, as given in ISO 19115.

	Name	Domain code	Definition
1.	CI_DateTypeCode	DateTypeCd	identification of when a given event occurred
2.	creation	0001	date identifies when the resource was brought into existence
3.	publication	0002	date identifies when the resource was issued
4.	revision	0003	date identifies when the resource was examined or re-examined and improved or amended

Table 4.2: CI_DateTypeCodeList CodeList table from ISO 19115

Since *CodeList* classes are extensible, it is necessary to give information about the used CodeList. This is done with the attribute *codeList*, which requires a valid URL that references a codeList definition within a registry of a codeList catalogue. The value after the #-sign points to the class used. Additionally, a codeListValue attribute has to be set. Listing 4.1 shows an example of encoding a CodeList class, using the *DateTypeCode* class as given in table 4.2.

```
<gmd:CI_DateTypeCode codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards
    /ISO_19139_Schemas/resources/Codelist/ML_gmxCodelists.xml#CI_DateTypeCode"
    codeListValue="creation">
    creation
</gmd:CI_DateTypeCode>
```

Listing 4.1: XML encoding of a CodeList class; used class: DateTypeCode, used codelist value: creation

Text and numbers Element values, which can be expressed by free text or numbers, are of data type *characterString* respectively *integer*. Both types are defined by the *gco* namespace.

Encoding example

Figure 4.1 shows an example of encoding of a dataset title. The tables on top show the metadata classes and their relations to each other, the code in the bottom demonstrates the implemented XML code. The dataset title is expressed by the *MD_Metadata* element *identificationInfo*. This element further refers to the class *MD_Identification*, which contains two specified classes, *MD_DataIdentification* and *MD_ServiceIdentification* (not shown in the figure). Since a dataset title has to be set, the class *MD_DataIdentification* hast to be used (*MD_Identification* refers to information about metadata itself, *MD_ServiceIdentification* to information about data services).

The element needed for setting a title is *citation*, which refers to the class *CI_Citation*. This class mandates the presence of the two elements *title* and *date* (other elements like eg. *alternateTitle* or *edition* are optional). The element *title* now represents our dataset title. It is of data type *CharacterString*, which is defined in the gco-namespace. The second mandatory field is *date* which further refers to the *CI_Date*-class. This class contains two elements the *date* itself and *dateType*, which is a *CodeList* with allowed values given in the class *CI_DateTypeCode* (see table 4.2. The valid format for *date* is specified by ISO 8601.

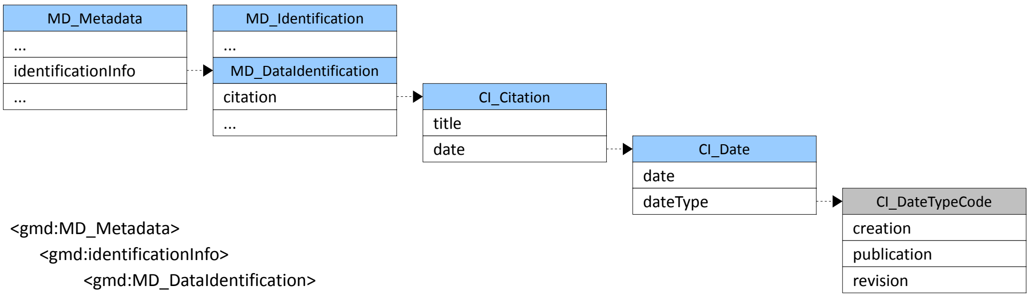
4.2 INSPIRE and ISO 19115 compliant implementation

The following chapter will demonstrate how to build up an INSPIRE and ISO 19115 conformal XML file following the INSPIRE implementing rules and the implementing instructions of the ISO standards ISO 19118 (Encoding), ISO 19136 (Geography Markup Language) and ISO 19139 (XML Schema implementation derived from ISO 19115). For further information about these implementation rules please see *INSPIRE* (2010), *ISO* (2011), *ISO* (2007a) and *ISO* (2007b), as they will not be discussed in this work in detail.

INSPIRE nomenclature will be used, following the order given in section 2.2.4. The code listings are based on the ASCAT AMSRE Level 4 SSM product[4]. If not further described, please see chapter 2.2.4 to get information about the obligations of the following properties. To obtain a more "graphical overview" about the hierarchical structure of the XML code, you can compare the following code listings with the figures 2.8 and 2.9 which cover most of the entities given in the following.

A complete INSPIRE compliant XML file can be found in annex A. Note that this file only includes properties required by INSPIRE and the ISO 19115 core.

[4] http://wacmos.itc.nl/?q=node/5




```
<gmd:MD_Metadata>
    <gmd:identificationInfo>
        <gmd:MD_DataIdentification>
            <gmd:citation>
                <gmd:CI_Citation>
                    <gmd:title>
                        <gco:CharacterString>Example Title</gco:Characterstring>
                    </gmd:title>
                    <gmd:date>
                        <gmd:CI_Date>
                            <gmd:date>
                                <gco:Date>2012-12-12</gco:Date>
                            </gmd:date>
                            <gmd:dateType>
                                <gmd:CI_DateTypeCode codeList=”http://urltocodelist.com#CI_DateTypeCode” codeListValue=”creation”>
                                    creation
                                </gmd:CI_DateTypeCode>
                            </gmd:dateType>
                        </gmd:CI_Date>
                    </gmd:date>
                </gmd:CI_Citation>
            </gmd:citation>
        </gmd:MD_DataIdentification>
    </gmd:identificationInfo>
</gmd:MD_Metadata>
```

Figure 4.1: Example of XML encoding of the property dataset title

4.2.1 Identification of the dataset

Resource Title

The dataset title is expressed by the element *title*, which is part of the class *CI_Citation*. It is mandated by ISO 19115 and INSPIRE and should always be given. Additionally, the class *CI_Citation*, requires information about the date. For further explanations see figure 4.1.

```
<gmd:MD_Metadata>
...
  <gmd:identificationInfo>
    <gmd:MD_DataIdentification>
    ...
      <gmd:citation>
        <gmd:CI_Citation>
          <gmd:title>
            <gco:CharacterString>ASCAT AMSRE L4 SSM</gco:CharacterString>
          </gmd:title>
          <gmd:date>
            <gmd:CI_Date>
              <gmd:date>
                <gco:Date>2010-08-03</gco:Date>
              </gmd:date>
              <gmd:dateType>
                <gmd:CI_DateTypeCode codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/Codelist/ML_gmxCodelists.xml#CI_DateTypeCode" codeListValue="creation">creation</gmd:CI_DateTypeCode>
              </gmd:dateType>
            </gmd:CI_Date>
          </gmd:date>
        </gmd:CI_Citation>
      </gmd:citation>
    ...
    </gmd:MD_DataIdentification>
  </gmd:identificationInfo>
...
</gmd:MD_Metadata>
```

Listing 4.2: XML code for Resource Title; mandated by ISO 19115 and INSPIRE

Resource Abstract

Similarly for the resource title is mandated by ISO 19115 and INSPIRE, it should describe the whole product in general.

```
<gmd:MD_Metadata>
...
  <gmd:identificationInfo>
    <gmd:MD_DataIdentification>
    ...
      <gmd:abstract>
        <gco:CharacterString>Combining information derived from satellite-based passive
             and active microwave sensors has the potential to offer improved
            estimates of surface soil moisture at global scale. We develop and
            evaluate a methodology that takes advantage of the retrieval
            characteristics of passive (AMSR-E) and active (ASCAT) microwave satellite
             estimates to produce an improved soil moisture product in flat_earth
            projection.
        </gco:CharacterString>
      </gmd:abstract>
    ...
    </gmd:MD_DataIdentification>
  </gmd:identificationInfo>
...
</gmd:MD_Metadata>
```

Listing 4.3: XML code for Resource Abstract; mandated by ISO 19115 and INSPIRE

Resource Locator

This information is mandated by INSPIRE, if a URL is available to obtain more information on the resource. Although this information is optional in ISO 19115, it should be given in order to achieve both ISO and INSPIRE compliance. The data type "URL" demands a semantically correct uniform resource locator.

```
<gmd:MD_Metadata>
...
  <gmd:distributionInfo>
    <gmd:MD_Distribution>
    ...
      <gmd:transferOptions>
        <gmd:MD_DigitalTransferOptions>
          <gmd:onLine>
            <gmd:CI_OnlineResource>
              <gmd:linkage>
                <gmd:URL>http://wacmos.itc.nl/?q=node/5</gmd:URL>
              </gmd:linkage>
            </gmd:CI_OnlineResource>
          </gmd:onLine>
```

```
          </gmd:MD_DigitalTransferOptions>
        </gmd:transferOptions>
      ...
      </gmd:MD_Distribution>
    </gmd:distributionInfo>
  ...
</gmd:MD_Metadata>
```

Listing 4.4: XML code for Resource Locator; mandated by INSPIRE, optional in ISO 19115

Resource Language

An instance of the language property is mandated by ISO 19115. It can be defaulted to the value of the metadata language when the dataset or the dataset series does not contain textual information. If this is the case, the element *language* has to be set outside the *identificationInfo* property (see section Metadata Language). In INSPIRE, this property is only mandated if the resource includes textual information.

Possible values are given in the codelist based on alpha-3 codes of ISO 639-2 (bibliographic codes). It is recommended to use only three-letter codes from in ISO 639-2B[5], since the two-letter codes do not cover all languages.

```
<gmd:MD_Metadata>
...
  <gmd:identificationInfo>
    <gmd:MD_DataIdentification>
      <gmd:language>
        <gmd:LanguageCode codeList="http://www.loc.gov/standards/iso639-2/"
              codeListValue="eng">eng</gmd:LanguageCode>
      </gmd:language>
    </gmd:identificationInfo>
  </gmd:MD_DataIdentification>
...
</gmd:MD_Metadata>
```

Listing 4.5: XML code for Resource Language; mandated by ISO 19115, conditional in INSPIRE

Resource Type

This property is not mandated by ISO 19115, but should be given to achieve conformance to the INSPIRE metadata implementing rules. Possible values for this property are given in the codelist *MD_ScopeCode* in Annex B of ISO 19115 (B.5.25)[6]. Although this list contains

[5] The full code list is given at http://www.loc.gov/standards/iso639-2/php/code_list.php
[6] The full list is available at http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/Codelist/gmxCodelists.xml#MD_ScopeCode

almost 20 entries, the only INSPIRE compliant values are "dataset", "series" and "service".

```
<gmd:MD_Metadata>
...
  <gmd:hierarchyLevel>
    <gmd:MD_ScopeCode codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/Codelist/gmxCodelists.xml#MD_ScopeCode" codeListValue="dataset">dataset
    </gmd:MD_ScopeCode>
  </gmd:hierarchyLevel>
...
</gmd:MD_Metadata>
```

Listing 4.6: *XML code for Resource Type; mandated by INSPIRE*

Unique Resource Identifier

A Unique Resource Identifier is only required by INSPIRE. This property could for example be the file name, if a single dataset file should be described. For describing a whole dataset series, it makes sense to use the product name extended with the version number (if present).

The *code* element is required at a minimum, and, if a certain code space or naming convention is used, a *codeSpace* element may be provided. If this value is provided, then the data type for the identifier should be *RS_Identifer*, which is substitutable for the usual *MD_Identifier*.

```
<gmd:MD_Metadata>
...
  <gmd:identificationInfo>
    <gmd:MD_DataIdentification>
      <gmd:citation>
        <gmd:CI_Citation>
        ...
          <gmd:identifier>
            <gmd:RS_Identifier>
              <gmd:code>
                <gco:CharacterString>ASCAT_AMSRE_L4_SSM_V01</gco:CharacterString>
              </gmd:code>
              <gmd:codeSpace>
                <gco:CharacterString>http://wacmos.itc.nl</gco:CharacterString>
              </gmd:codeSpace>
            </gmd:RS_Identifier>
          </gmd:identifier>
```

```
        ...
        </gmd:CI_Citation>
      </gmd:citation>
    </gmd:MD_DataIdentification>
  </gmd:identificationInfo>
...
</gmd:MD_Metadata>
```

Listing 4.7: XML code for Unique Resource Identifier; mandated by INSPIRE

Dataset Character Set

ISO 19115 mandates information about the used character set if ISO/IEC 10646-1 (eg: UTF8) is not used. Possible values for this property are listed in the codelist *MD_CharacterSetCode* in ISO 19115 (B.5.10).

```
<gmd:MD_Metadata>
...
  <gmd:identificationInfo>
    <gmd:MD_DataIdentification>
      <gmd:characterSet>
        <gmd:MD_CharacterSetCode codeList="http://standards.iso.org/ittf/
            PubliclyAvailableStandards/ISO_19139_Schemas/resources/Codelist/
            gmxCodelists.xml#MD_CharacterSetCode" codeListValue="utf8">utf8
        </gmd:MD_CharacterSetCode>
      </gmd:characterSet>
    </gmd:MD_DataIdentification>
  </gmd:identificationInfo>
...
</gmd:MD_Metadata>
```

Listing 4.8: XML code for Dataset Character Set; mandated by ISO 19115

4.2.2 Classification

Topic Category

The property topic category is mandated by both INSPIRE and ISO 19115. This element is an enumeration, with allowed values listed in ISO 19115 (B.5.27). If necessary several categories can be given. In our example, the topic category is "geoscientific information".

```
<gmd:MD_Metadata>
...
  <gmd:identificationInfo>
```

```
    <gmd:MD_DataIdentification>
    ....
      <gmd:topicCategory>
        <gmd:MD_TopicCategoryCode>geoscientificInformation</gmd:MD_TopicCategoryCode>
      </gmd:topicCategory>
    ...
    </gmd:MD_DataIdentification>
  </gmd:identificationInfo>
...
</gmd:MD_Metadata>
```

Listing 4.9: XML code for Topic Category; mandated by INSPIRE and ISO 19115

Keyword and Originating Controlled Vocabulary

The INSPIRE implementing rules for metadata mandate the presence of at least one keyword, originating from the general environmental multilingual thesaurus (GEMET)[7] (*INSPIRE*, 2010, p. 23).

In this example, the keyword "Soil" originating from "GEMET - INSPIRE themes, version 1.0" is used. Listing 4.10 shows the corresponding XML code. Each keyword is described within *descriptiveKeywords* property. In order to be consistent with ISO 19115, all keyword values originating from a single version of a single controlled vocabulary shall be grouped in a single instance of the ISO 19115 this property. The keyword value itself is represented by a *keyword* tag. Originating controlled vocabularies (if present) are represented within the *thesaurusName* element. This element refers to the class *CI_Citation*, which mandates a *title* and a *date*, and further a *dateType*.

```
<gmd:MD_Metadata>
...
  <gmd:identificationInfo>
    <gmd:MD_DataIdentification>
      ...
      <gmd:descriptiveKeywords>
        <gmd:MD_Keywords>
          <gmd:keyword>
            <gco:CharacterString>Soil</gco:CharacterString>
          </gmd:keyword>
          <gmd:thesaurusName>
            <gmd:CI_Citation>
              <gmd:title>
                <gco:CharacterString>
                  GEMET - INSPIRE themes, version 1.0
```

[7] http://www.eionet.europa.eu/gemet/inspire_themes

```
                </gco:CharacterString>
              </gmd:title>
              <gmd:date>
                <gmd:CI_Date>
                  <gmd:date>
                    <gco:Date>2008−06−01</gco:Date>
                  </gmd:date>
                  <gmd:dateType>
                    <gmd:CI_DateTypeCode codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/Codelist/ML_gmxCodelists.xml#CI_DateTypeCode" codeListValue="publication">
                      publication
                    </gmd:CI_DateTypeCode>
                  </gmd:dateType>
                </gmd:CI_Date>
              </gmd:date>
            </gmd:CI_Citation>
          </gmd:thesaurusName>
        </gmd:MD_Keywords>
      </gmd:descriptiveKeywords>
      ...
    </gmd:MD_DataIdentification>
  </gmd:identificationInfo>
...
</gmd:MD_Metadata>
```

Listing 4.10: XML code for Keyword and Originating Controlled Vocabulary; mandated by INSPIRE

4.2.3 Geographic Location

While INSPIRE always mandates a bounding box, this property is only mandated by ISO 19115 if the element *hierarchyLevel* equals "dataset" (see section Resource Type). As discussed in chapter 2.2.4, the bounding box has to be described by west- and east-bound longitudes and south- and north-bound latitudes in decimal values.

```
<gmd:MD_Metadata>
...
  <gmd:identificationInfo>
    <gmd:MD_DataIdentification>
    ...
      <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:MD_DataIdentification>
 </gmd:identificationInfo>
...
</gmd:MD_Metadata>
```

Listing 4.11: XML code for the geographic location; mandated by INSPIRE

4.2.4 Temporal Reference

Temporal Extent

The temporal extent is described by the class *EX_TemporalExtent*, which requests a begin- and end date. Both dates are of data type *TM_Primitive*, which is an abstract class that represents a non-decomposed element of geometry or topology of time *ISO* (2002), described in the standard ISO 19108.

```
<gmd:MD_Metadata>
...
 <gmd:identificationInfo>
  <gmd:MD_DataIdentification>
   ...
    <gmd:extent>
      <gmd:EX_Extent>
        ...
        <gmd:temporalElement>
```

```
            <gmd:EX_TemporalExtent>
              <gmd:extent>
                <gml:TimePeriod gml:id="extent">
                  <gml:beginPosition>1978−01−01</gml:beginPosition>
                  <gml:endPosition>2010−12−31</gml:endPosition>
                </gml:TimePeriod>
              </gmd:extent>
            </gmd:EX_TemporalExtent>
          </gmd:temporalElement>
          ...
        </gmd:EX_Extent>
      </gmd:extent>
    ...
    </gmd:MD_DataIdentification>
  </gmd:identificationInfo>
...
</gmd:MD_Metadata>
```

Listing 4.12: XML code for the temporal extent; conditional in INSPIRE, optional for ISO 19115

Date of Publication/Creation/Revision

The three properties, "date of publication", "date of creation" and "date of last revision" are parts of the *CI_Date* class which is composed of the mandated elements *date* and *dateType*.

```
<gmd:MD_Metadata>
...
  <gmd:identificationInfo>
    <gmd:MD_DataIdentification>
    ...
      <gmd:citation>
        <gmd:CI_Citation>
        ...
          <gmd:date>
            <gmd:CI_Date>
              <gmd:date>
                <gco:Date>2007−09−15</gco:Date>
              </gmd:date>
              <gmd:dateType>
                <gmd:CI_DateTypeCode codeList="http://standards.iso.org/ittf/
                    PubliclyAvailableStandards/ISO_19139_Schemas/resources/Codelist/
                    ML_gmxCodelists.xml#CI_DateTypeCode" codeListValue="publication">
                    publication</gmd:CI_DateTypeCode>
              </gmd:dateType>
            </gmd:CI_Date>
```

```
            </gmd:date>
           ...
          </gmd:CI_Citation>
        </gmd:citation>
      ...
    </gmd:MD_DataIdentification>
  </gmd:identificationInfo>
...
</gmd:MD_Metadata>
```

Listing 4.13: XML code for setting a certain date and type; conditional for INSPIRE, mandated by ISO 19115

4.2.5 Quality and Validity

Lineage

Information about lineage is given in the class *LI_Lineage*, which is part of *DQ_DataQuality*. INSPIRE mandates a general lineage statement, which is expressed by the element *statement*.

```
<gmd:MD_Metadata>
...
  <gmd:dataQualityInfo>
    <gmd:DQ_DataQuality>
     ...
      <gmd:lineage>
        <gmd:LI_Lineage>
          <gmd:statement>
            <gco:CharacterString>Volumetric soil water content (m\^3 m\^-3) from AMSR-E
                 and degree of saturation (in percent) from ASCAT are rescaled against
                 a reference land surface model data set using a cumulative
                distribution function matching approach. While this imposes any bias
                of the reference on the rescaled satellite products, it adjusts them
                to the same range and preserves the dynamics of original satellite-
                based products. Comparison with in situ measurements demonstrates that
                 where the correlation coeffcient between rescaled AMSR-E and ASCAT is
                 greater than 0.65 (transitional regions), merging the different
                satellite products increases the number of observations while
                minimally changing the accuracy of soil moisture retrievals. The error
                 characterisation is represented by the variable sm_noise in the
                header of the NetCDF file. For full description of the errors see
                Dorigo et al. (2010; doi:10.5194/hess-14-2605-2010)
            </gco:CharacterString>
          </gmd:statement>
```

```
          </gmd:LI_Lineage>
        </gmd:lineage>
      ...
      </gmd:DQ_DataQuality>
    </gmd:dataQualityInfo>
  ...
</gmd:MD_Metadata>
```

Listing 4.14: XML code for Lineage, mandated by INSPIRE

Spatial Resolution

Spatial resolution of a dataset can be expressed as an equivalent scale or as a resolution distance (*INSPIRE*, 2012). Both possibilities are covered by the class *MD_Resolution*. However, only the latter case will be discussed, since this is most commonly used for remote sensing products. This case is implemented as *Distance* type. The unit of measure is expressed by the attribute "uom" within the distance tag and is either a conventional unit of measure symbol (e.g. `<gco:Distance uom="km">`) or a link to a definition. The latter case is illustrated in listing 4.15.

```
<gmd:MD_Metadata>
  ...
    <gmd:identificationInfo>
      <gmd:MD_DataIdentification>
      ...
        <gmd:spatialResolution>
                <gmd:MD_Resolution>
                    <gmd:distance>
                        <gco:Distance uom="http://standards.iso.org/ittf/
                            PubliclyAvailableStandards/ISO_19139_Schemas/resources/uom
                            /ML_gmxUom.xml#m">25000</gco:Distance>
                    </gmd:distance>
                </gmd:MD_Resolution>
            </gmd:spatialResolution>
      ...
      </gmd:MD_DataIdentification>
    </gmd:identificationInfo>
  ...
</gmd:MD_Metadata>
```

Listing 4.15: XML code for spatial resolution, conditional in INSPIRE, optional in ISO 19115

Spatial Representation Type

The spatial representation type is an optional property in ISO 19115. It is represented by the element *spatialRepresentationType*, which is part of the *MD_DataIdentification* class. It allows values given in the code list *MD_SpatialRepresentationTypeCode* (ISO 19115, B.5.26).

```
<gmd:MD_Metadata>
...
  <gmd:identificationInfo>
    <gmd:MD_DataIdentification>
    ...
      <gmd:spatialRepresentationType>
        <gmd:MD_SpatialRepresentationTypeCode codeList="http://standards.iso.org/ittf/
            PubliclyAvailableStandards/ISO_19139_Schemas/resources/Codelist/
            gmxCodelists.xml#MD_SpatialRepresentationTypeCode">grid</
            gmd:MD_SpatialRepresentationTypeCode>
      </gmd:spatialRepresentationType>
    ...
    </gmd:MD_DataIdentification>
  </gmd:identificationInfo>
...
</gmd:MD_Metadata>
```

Listing 4.16: XML code for the spatial representation type

Distribution format

Information about the distribution format is covered within the class *MD_Distribution* by the element *distributionFormat*. This element refers to the class *MD_Format*, which mandates at least information about the format name and version. This property is optional in ISO 19115.

```
<gmd:MD_Metadata>
...
  <gmd:distributionInfo>
    <gmd:MD_Distribution>
    ...
      <gmd:distributionFormat>
        <gmd:MD_Format>
          <gmd:name>
            GeoTiff
          </gmd:name>
          <gmd:version>
            1.0
```

```
            </gmd:version>
          </gmd:MD_Format>
        </gmd:distributionFormat>
      ...
      </gmd:MD_Distribution>
    </gmd:distributionInfo>
  ...
</gmd:MD_Metadata>
```

Listing 4.17: XML code for distribution format, optional in ISO 19115

4.2.6 Conformity

INSPIRE mandates information about the degree conformance to a certain INSPIRE specification. This information can be given within the *DQ_ConformanceResult* class, which mandates the presence of the elements *specification*, *explanation*) and *pass*. The specification is represented by the *CI_Citation* class, which mandates a title and a date, and further a datetype.

The INSPIRE implementing rules define three degrees of conformity: "conformant", "not conformant" and "not evaluated". Since the ISO 19115 element *pass*, by which the degree of conformance is expressed, only allows the values "true" (eq. conformant) and "false" (eq. not conformant), it is only possible to report about conformity to a specification when it has been evaluated. The only ISO 19115 conformant way for not evaluated conformity would be not to report anything in the metadata. Further ISO 19115 mandates an explanation of the meaning of the conformance for this result. The INSPIRE implementing rules recommend to use a default explanation such as "See the referenced specification" (*INSPIRE*, 2010, p. 52).

```
<gmd:MD_Metadata>
...
  <gmd:dataQualityInfo>
    <gmd:DQ_DataQuality>
      ...
      <gmd:report>
        <gmd:DQ_DomainConsistency xsi:type="gmd:DQ_DomainConsistency_Type">
          <gmd:result>
            <gmd:DQ_ConformanceResult xsi:type="gmd:DQ_ConformanceResult_Type">
              <gmd:specification>
                <gmd:CI_Citation>
                  <gmd:title>
                    <gco:CharacterString>COMMISSION REGULATION (EC) No 1205/2008 of 3
                        December 2008 implementing Directive 2007/2/EC of the European
```

```
                        Parliament and of the Council as regards metadata</
                        gco:CharacterString>
                  </gmd:title>
                  <gmd:date>
                    <gmd:CI_Date>
                      <gmd:date>
                        <gco:Date>2008-12-04</gco:Date>
                      </gmd:date>
                      <gmd:dateType>
                        <gmd:CI_DateTypeCode codeList="http://standards.iso.org/ittf/
                            PubliclyAvailableStandards/ISO_19139_Schemas/resources/
                            Codelist/ML_gmxCodelists.xml#CI_DateTypeCode"
                            codeListValue="publication">publication</
                            gmd:CI_DateTypeCode>
                      </gmd:dateType>
                    </gmd:CI_Date>
                  </gmd:date>
                </gmd:CI_Citation>
              </gmd:specification>
              <gmd:explanation>
                <gco:CharacterString>See the referenced specification</
                    gco:CharacterString>
              </gmd:explanation>
              <gmd:pass>
                <gco:Boolean>true</gco:Boolean>
              </gmd:pass>
            </gmd:DQ_ConformanceResult>
          </gmd:result>
        </gmd:DQ_DomainConsistency>
      </gmd:report>
      ...
    </gmd:DQ_DataQuality>
  </gmd:dataQualityInfo>
  ...
</gmd:MD_Metadata>
```

Listing 4.18: XML code for conformity and specification

4.2.7 Constraints related to access and use

Information about constraints which apply to the resources are represented by the element *resourceConstraints* given in *MD_DataIdentification*.

Conditions applying to access and use are expressed by the element *useLimitation* within the subclass *MD_Constraints*, access constraints (*accessConstraints*) and use limitations (*otherConstraints*) within *MD_LegalConstraints*. If there are no limitations on public access, the metadata file shall indicate that fact (*INSPIRE*, 2007). Therefore, the element *accessConstraints* has to be set to "otherRestrictions", and *otherConstraints* to "no limitations".

```
<gmd:MD_Metadata>
...
  <gmd:identificationInfo>
    <gmd:MD_DataIdentification>
    ...
      <gmd:resourceConstraints>
        <gmd:MD_Constraints>
          <gmd:useLimitation>
            <gco:CharacterString>No conditions apply</gco:CharacterString>
          </gmd:useLimitation>
        </gmd:MD_Constraints>
      </gmd:resourceConstraints>
      <gmd:resourceConstraints>
        <gmd:MD_LegalConstraints>
          <gmd:accessConstraints>
            <gmd:MD_RestrictionCode codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/Codelist/gmxCodelists.xml#MD_RestrictionCode" codeListValue="otherRestrictions">>otherRestrictions</gmd:MD_RestrictionCode>
          </gmd:accessConstraints>
          <gmd:otherConstraints>
            <gco:CharacterString>no limitations</gco:CharacterString>
          </gmd:otherConstraints>
        </gmd:MD_LegalConstraints>
      </gmd:resourceConstraints>
    ...
    </gmd:MD_DataIdentification>
  </gmd:identificationInfo>
...
</gmd:MD_Metadata>
```

Listing 4.19: XML code for access constraints and limitations

4.2.8 Responsible organisation

The responsible organisation is represented by *CI_ResponsibleParty*, which is a subclass of *CI_Citation*. INSPIRE mandates a name (*organisationName*) and a contact e-mail address(*contactInfo*). Further, the role of the responsible organisation has to be given (*role*), represented by a value from the codelist *CI_RoleCode*.

```
<gmd:MD_Metadata>
...
```

```
  <gmd:identificationInfo>
    <gmd:MD_DataIdentification>
     ...
      <gmd:pointOfContact>
        <gmd:CI_ResponsibleParty>
          <gmd:organisationName>
            <gco:CharacterString>I.P.F, Vienna</gco:CharacterString>
          </gmd:organisationName>
          <gmd:contactInfo>
            <gmd:CI_Contact>
              <gmd:address>
                <gmd:CI_Address>
                  <gmd:electronicMailAddress>
                    <gco:CharacterString>wd@ipf.tuwien.ac.at</gco:CharacterString>
                  </gmd:electronicMailAddress>
                </gmd:CI_Address>
              </gmd:address>
            </gmd:CI_Contact>
          </gmd:contactInfo>
          <gmd:role>
            <gmd:CI_RoleCode codeList="http://standards.iso.org/ittf/
                PubliclyAvailableStandards/ISO_19139_Schemas/resources/Codelist/
                ML_gmxCodelists.xml#CI_RoleCode" codeListValue="author">author</
                gmd:CI_RoleCode>
          </gmd:role>
        </gmd:CI_ResponsibleParty>
      </gmd:pointOfContact>
     ...
    </gmd:MD_DataIdentification>
  </gmd:identificationInfo>
...
</gmd:MD_Metadata>
```

Listing 4.20: XML code for responsible party

4.2.9 Metadata on metadata

Metadata point of contact

This attribute is given within the *contact* element of *MD_Metadata*. The name and contact information can be given like for the responsible organisation. The role of the responsible party serving as a metadata point of contact is out of scope of the INSPIRE implementing rules, but this property is mandated by ISO 19115. Its value can be defaulted to "pointOfContact". (*INSPIRE*, 2010)

```
<gmd:MD_Metadata>
...
  <gmd:contact>
    <gmd:CI_ResponsibleParty>
      <gmd:organisationName>
        <gco:CharacterString>I.P.F, Vienna</gco:CharacterString>
      </gmd:organisationName>
      <gmd:contactInfo>
        <gmd:CI_Contact>
          <gmd:address>
            <gmd:CI_Address>
              <gmd:electronicMailAddress>
                <gco:CharacterString>wd@ipf.tuwien.ac.at</gco:CharacterString>
              </gmd:electronicMailAddress>
            </gmd:CI_Address>
          </gmd:address>
        </gmd:CI_Contact>
      </gmd:contactInfo>
      <gmd:role>
        <gmd:CI_RoleCode codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/Codelist/ML_gmxCodelists.xml#CI_RoleCode" codeListValue="pointOfContact">pointOfContact</gmd:CI_RoleCode>
      </gmd:role>
    </gmd:CI_ResponsibleParty>
  </gmd:contact>
...
</gmd:MD_Metadata>
```

Listing 4.21: XML code for metadata point of contact

Metadata date

The metadata date is mandated by both ISO 19115 and INSPIRE.

```
<gmd:MD_Metadata>
...
  <gmd:dateStamp>
    <gco:Date>2012-02-09</gco:Date>
  </gmd:dateStamp>
...
</gmd:MD_Metadata>
```

Listing 4.22: XML code for metadata date

Metadata language

The language property is not mandated by ISO 19115, but is mandated for conformance to the INSPIRE metadata implementing rules (*INSPIRE*, 2010). For further encoding information please refer to section resource language.

```
<gmd:MD_Metadata>
...
  <gmd:LanguageCode codeList="http://www.loc.gov/standards/iso639-2/" codeListValue="
      eng">eng</gmd:LanguageCode>
...
</gmd:MD_Metadata>
```

Listing 4.23: XML code for metadata date

Metadata character set

For further implementation instructions please refer to section Resource Language.

```
<gmd:MD_Metadata>
...
  <gmd:characterSet>
    <gmd:MD_CharacterSetCode codeList="http://standards.iso.org/ittf/
        PubliclyAvailableStandards/ISO_19139_Schemas/resources/Codelist/gmxCodelists.
        xml#MD_CharacterSetCode" codeListValue="utf8">utf8
    </gmd:MD_CharacterSetCode>
  </gmd:characterSet>
...
</gmd:MD_Metadata>
```

Listing 4.24: XML code for metadata character set

Metadata file identifier

A metadata file identifier is optional in ISO 19115. A possible value could for example be the filename of the metadata file. For further implementation instructions please refer to section Unique Resource Identifier.

```
<gmd:MD_Metadata>
...
  <gmd:fileIdentifier>
    <gco:CharacterString>ASCAT_AMSRE_L4_SSM_V01</gco:CharacterString>
  </gmd:fileIdentifier>
```

```
...
</gmd:MD_Metadata>
```

Listing 4.25: XML code for metadata file identifier

Metadata standard name and version

ISO 19115 provides the possibility to describe which standards and standard versions have been used in the metadata file. In the following example, ISO 19115:2003 is used.

```
<gmd:MD_Metadata>
...
  <gmd:metadataStandardName>
    <gco:CharacterString>ISO 19115</gco:CharacterString>
  </gmd:metadataStandardName>
  <gmd:metadataStandardVersion>
    <gco:CharacterString>2003</gco:CharacterString>
  </gmd:metadataStandardVersion>
    <gmd:metadataStandardName>
    <gco:CharacterString>ISO 19115-2</gco:CharacterString>
  </gmd:metadataStandardName>
  <gmd:metadataStandardVersion>
    <gco:CharacterString>2009</gco:CharacterString>
  </gmd:metadataStandardVersion>
...
</gmd:MD_Metadata>
```

Listing 4.26: XML code for metadata standardname

4.3 Extended, ISO compliant implementation

The previous chapter demonstrated how to implement a XML metadata file using information mandated by INSPIRE and the ISO 19115 core. However, the information provided may not be sufficient to adequately describe remote sensing products, and should therefore be extended by additional attributes, if existent in ISO 19115 or ISO 19115-2. If a suitable element does not exist, the ISO standard can be extended under certain conditions (*ISO*, 2003, p.105). The following examples demonstrates how to implement additional information from the extension ISO 19115-2 and implementation of an extension.

4.3.1 Implementing ISO 19115-2

Chapter 3 stated that each product is derived from data obtained by a sensor onboard a satellite in a certain mode. Both INSPIRE and ISO 19115 only provide the possibility to give this information within a general lineage statement. In addition to this textual description of the lineage, ISO 19115-2 allows to give more detailed information about the data acquisition by introducing the *MI_Metadata* package. This package is a super-set of the *MD_Metadata* package that can provide important additional metadata, like descriptions of the instrumentation used to make the measurements and the processing of those measurements (*Habermann*, 2012). The class that provides this information is *MI_AcquisitionInformation*. Figure 4.2 shows how this class is linked to the ISO 19115 standard.

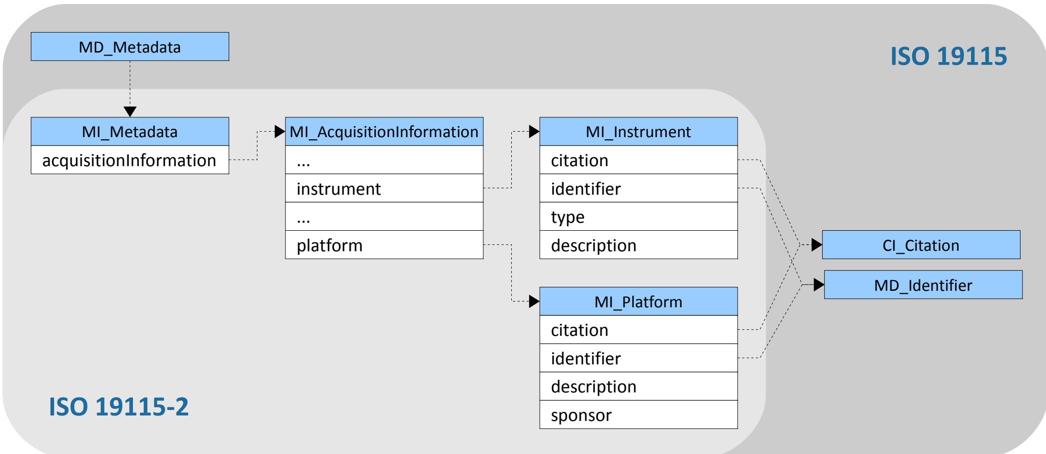


Figure 4.2: Extending the ISO 19115 standard with the ISO 19115-2 class MI_AcquisitionInfo

However, the use of this extension requires changing the name and the namespace of the object. Instead of *MD_Metadata* it is mandated to use *MI_Metadata*, the namespace changes to "gmi". As a result, the root element is now *gmi:MI_Metadata* instead of *gmd:MD_Metadata*. The use of this namespace is supported by INSPIRE (*INSPIRE* (2010), p. 49).

Listing 4.27 shows an example of how to implement this additional information. In the given example, the data was recorded by two sensors (ASCAT and ASMR-E) onboard two separate satellites. ISO 19115-2 allows the presence of multiple *acquisitionInformation* elements.

```
<gmd:MI_Metadata>
...
  <gmd:identificationInfo>
    <gmd:MD_DataIdentification>
    ...
      <gmi:acquisitionInformation>
        <gmi:platform>
          <gmi:citation>
            <gmd:CI_Citation>
              <gmd:title>
                <gco:CharacterString>MetOp</gco:CharacterString>
              </gmd:title>
              <gmd:date> ... </gmd:date>
            </gmd:CI_Citation>
          <gmi:citation>
          <gmi:identifier>
            <gmd:MD_Identifier>
            </gmd:MD_Identifier>
          </gmi:identifier>
        <gmi:platform>
        <gmi:instrument>
          <gmi:citation>
            <gmd:CI_Citation>
              <gmd:title>
                <gco:CharacterString>ASCAT</gco:CharacterString>
              </gmd:title>
              <gmd:date> ... </gmd:date>
            </gmd:CI_Citation>
          <gmi:citation>
          <gmi:identifier>
            <gmd:MD_Identifier>
            </gmd:MD_Identifier>
          </gmi:identifier>
        </gmi:instrument>
      </gmi:acquisitionInformation>
    ...
    </gmd:MD_DataIdentification>
  </gmd:identificationInfo>
...
</gmd:MI_Metadata>
```

Listing 4.27: *XML code for MI_AcquisitionInformation*

4.3.2 User extension of ISO standards

Although information about the used satellite and sensor is now implemented, information about the recording mode is still missing. Both ISO 19115 and ISO 19115-2 do not provide elements for that specific information. As stated before, it is possible to create community profiles which extend the ISO standards. This can be achieved by applying the following steps:

1. A suitable metadata entity to which the new element should be added has to be found. In this case the entity *MI_AcquisitionInformation* is extended with the element *recordingMode*. This element further points to the new class *MI_RecordingMode*, which provides elements for describing the mode. Linkage between the mode and the sensor which uses it is ensured by the element *usedBy* in *MI_RecordingMode* respectively *recordingMode* in *MI_Instrument*.

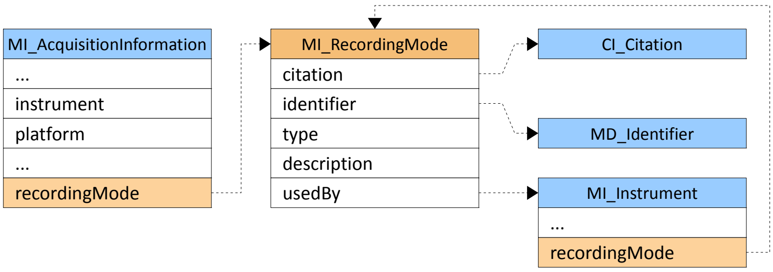


Figure 4.3: Extending the ISO 19115 standard with the new class MI_RecordingMode

2. Now the new metadata elements needs to be defined by the use of class *MD_ExtendedElementInformation* (as given in ISO 19115 B.2.11.2). Table 4.3 shows the extension information of the element *redordingMode* (implementation of the remaining elements will not be demonstrated in this work).

3. The new metadata entities and elements now have to be documented within a profile document, which has to be derived from the ISO 19115 document itself and issued as a document along with the dataset and the metadata (*ISO*, 2003, p. 117). The individual documentation steps are listed in ISO 19115 F.10 and will not be further discussed in this work.

XML implementation The new defined metadata element *recordingMode* is implemented the same way as *instrument* and *platform* (cf. 4.27).

MD_MetadataExtensionInformation		
MD_MetadataElementInformation		
name	recordingMode	
shortName	recMod	
definition	recording mode used for data acquisition	
obligation	O	
dataType	Assiciation	
maximumOccurence	N	
domainValue	MI_RecordingMode	
parentEntity	MI_AcquisitionInformation	
rule	New element	
rationale	To describe the recording mode, if used.	
source	**organisationName**	TU Wien
	role	owner

Table 4.3: Extending the ISO 19115 standard with the new element recordingMode

4.4 Conclusion

The standard ISO 19115 and its extension ISO 19115-2 offer a huge variety of elements for describing geographic metadata. However, the metadata entities and elements given in these standards are not sufficient to describe satellite based remote sensing data to their full extent. ISO provides the possibility to extend certain standards by creating a community profile, such as for example the *Marine Community Profile of ISO 19115*[8]. However, these extensions will not be known outside the community unless they are published (*ISO*, 2003, p.106).

[8] https://marinemetadata.org/references/marineprofile19115

Chapter 5

Data Management

One of the major tasks of this work is to investigate methods to handle storage and accessibility of geographic metadata. The information must be consistent and non-redundant at any time. Further, to provide that information to a wide range of users, it should be accessible from the Internet. These requirements can be met by using a relational database system (DBS). A DBS consists of a set of data and a software necessary for data processing, called database management system (DBMS). Basic functions of a DBMS are for example data storage and manipulation, provisions for data security and integrity, facilitation of multi user synchronization.

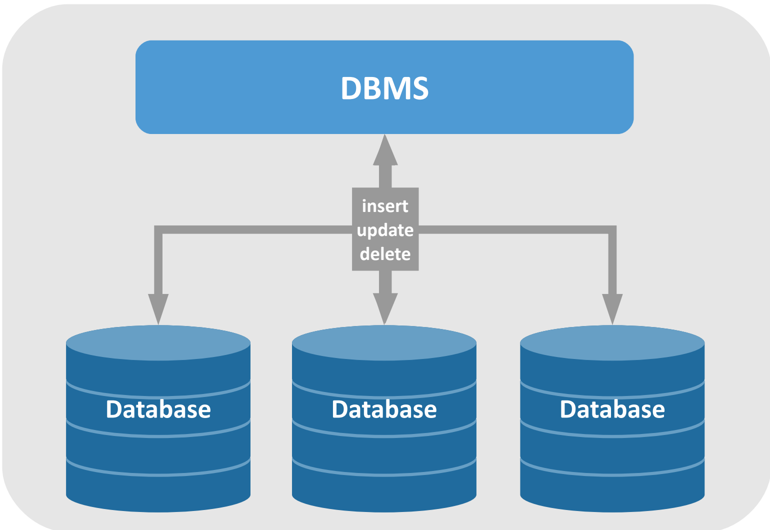


Figure 5.1: Scheme of a Database System

Each DBS provides a certain query language for data manipulation. A common query language for relational DBS is SQL (Structured Query Language). In this work PostgreSQL[1], an open source object-relational database management system was used.

[1] http://www.postgresql.org/

5.1 Database Model

The information about the products now needs to be stored in a database, following the definitions given in Chapter 3. First, a database model which describes the ontology of products and product versions should be set up. In the following chapter, an entity-relationship model is used to describe the relations between the products entities and attributes. *The entity-relationship model adopts the more natural view that the real world consists of entities and relationships. It incorporates some of the important semantic information about the real world.* *Chen* (1976)

In Chen's model, entities are represented by rectangles and Relations by diamonds (see figure 5.2). To represent the cardinalities between these relations, the Modified Chen notation (MC-Notation) was used (see table 5.1).

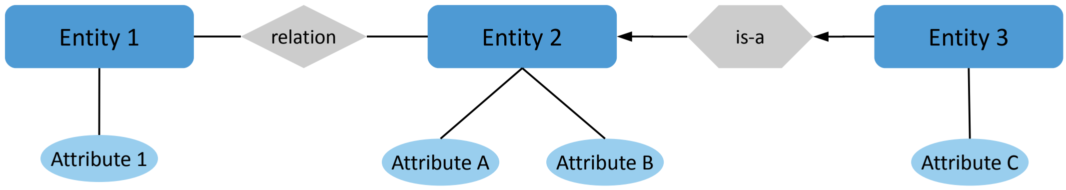


Figure 5.2: Example for ER-Model

Figure 5.2 shows the components of the ER model which are mainly used in this chapter.

- **Entity:** An entity may be defined as a thing which is recognized as being capable of an independent existence and which can be uniquely identified. An entity is an abstraction from the complexities of some domain. When we speak of an entity we normally speak of some aspect of the real world which can be distinguished from other aspects of the real world (*Beynon-Davies*, 2004).
- **Relation:** Describes the relationship between different entities.
- **Is-a:** Entity sets can have subsets with additional attributes. E.g. figure 5.2: Entity 3 is a subset of Entity 2. Both entities have Attributes A and B, but Entity 3 has additionally an Attribute C.
- **Attribute:** Both, entities and relationships can have several attributes. Each entity must have an uniquely identifying attribute, the so called primary key.

5.1.1 Product Model

The product model should contain all the attributes that uniquely define the product. This means that, if one of these attributes changes, a new product will be designed.

1:1	Each entity of the first entity set is in relation to **exactly one** entity of the second entity set.
1:C	Each entity of the first entity set is in relation to **at most one** entity of the second entity set. Each entity of the second entity set is in relation to **exactly one** entity of the first entity set.
1:M	Each entity of the first entity set is in relation to **at least one** entity of the second entity set. Each entity of the second entity set is in relation to **exactly one** entity of the first entity set.
1:MC	Each entity of the first entity set is in relation to **any number** of entities of the second entity set. Each entity of the second entity set is in relation to **exactly one** entity of the first entity set.
C:C	Each entity of the first entity set is in relation to at most one entity of the second entity set and vice versa.
C:M	Each entity of the first entity set is in relation to **at least one** entity of the second entity set. Each entity of the second entity set is in relation to **at most one** entity of the first entity set.
C:MC	Each entity of the first entity set is in relation to **any number** of entities of the second entity set. Each entity of the second entity set is in relation to **at most one** entity of the first entity set.
M:M	Each entity of the first entity set is in relation to **at least one** entity of the second entity set and vice versa.
M:MC	Each entity of the first entity set is in relation to **any number** of entities of the second entity set. Each entity of the second entity set is in relation to **at least one** entity of the first entity set.
MC:MC	Each entity of the first entity set is in relation to **any number** of entities of the second entity set and vice versa.

Table 5.1: Cardinalities in MC Notation (from Chen, 1976)

Source

Each product has at least one source, which can be satellite-, in situ or model data. Satellites have at least one sensor and each sensor has at least one recording mode. Figure 5.3 shows the ER-Model for the relation between products and source, figure 5.4 shows the corresponding database structure.

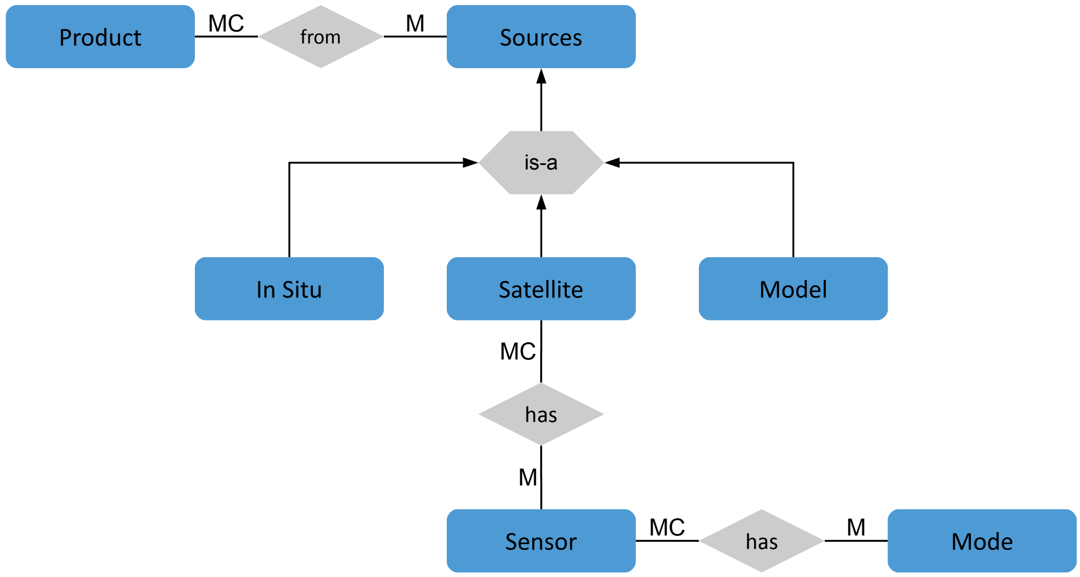


Figure 5.3: ER-Model of the product-source relation.

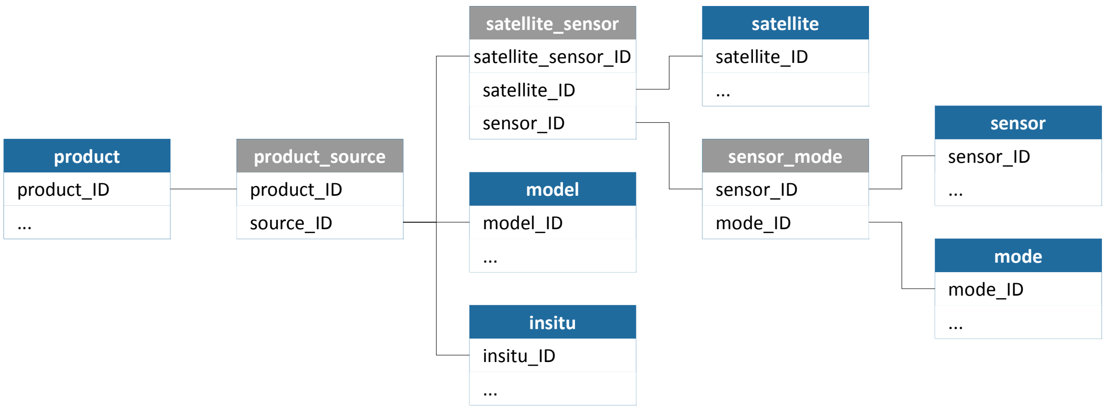


Figure 5.4: Database-Model of the product-source relation

Resolution

Each product has a spatial and temporal resolution and additionally a temporal- and spatial sampling rate. Figure 5.5 shows the ER-Model for the relation between products and resolution, figure 5.6 the corresponding database structure.

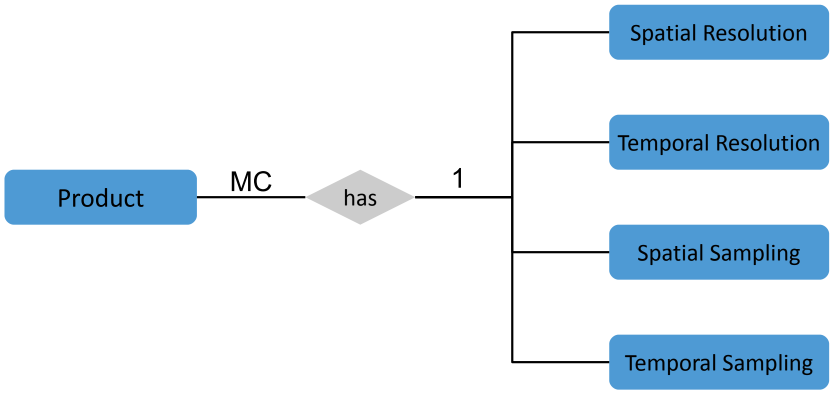


Figure 5.5: ER-Model of product-resolution.

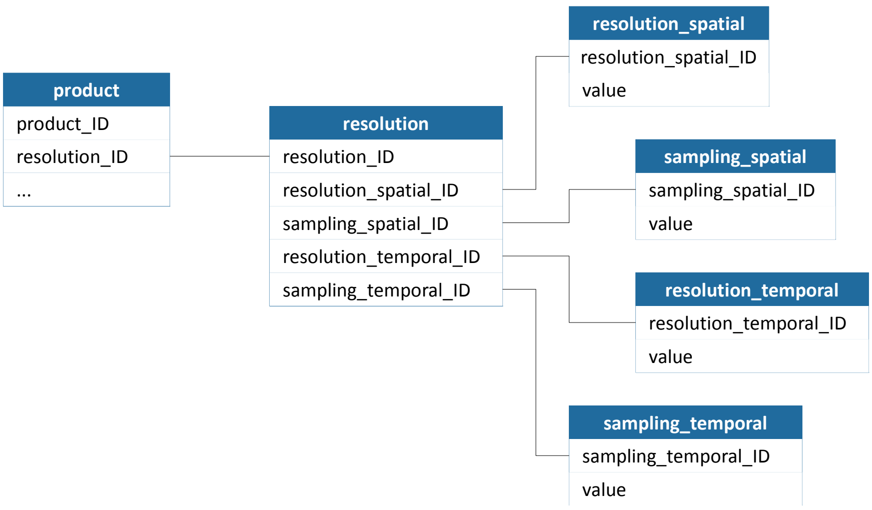


Figure 5.6: Database-Model for resolution

Other attributes

A product must have exactly one data structure, one provider and be assigned to one specific category. It further contains at least geophysical variable, and exist in a certain a product level. Additionally, according to INSPIRE, a general description (abstract) should be provided. The attribute list of the product can be extended, but to avoid redundancy, it has to be checked if this attribute would rather be connected to a product version.

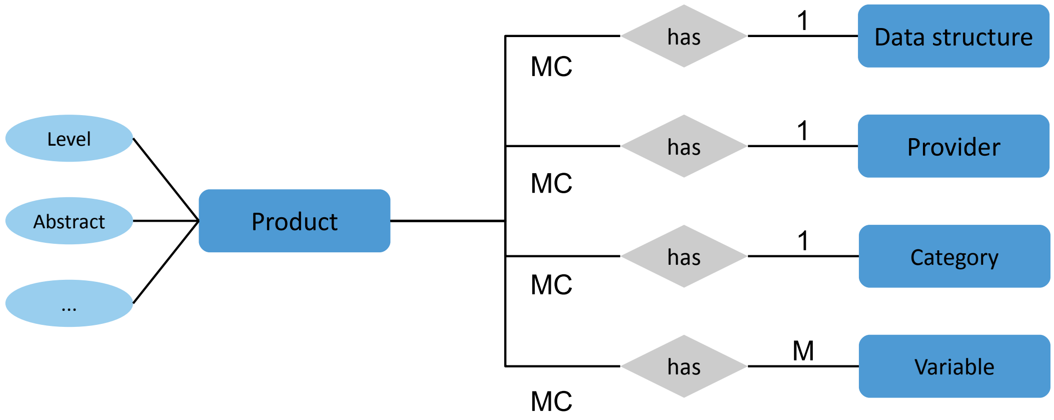


Figure 5.7: *ER-Model for product-, category-,variable-, data structure-, project- and provider relation*

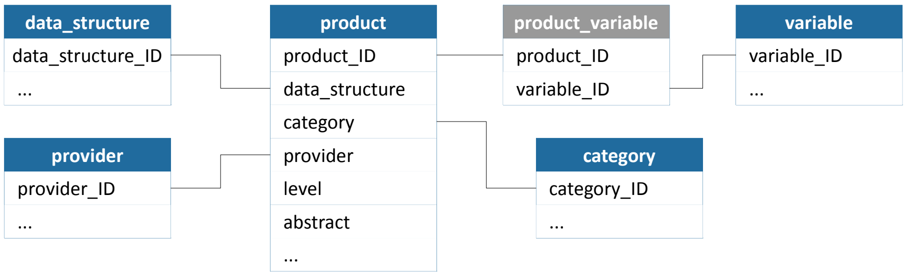


Figure 5.8: *Database-Model for product-, category-, variable-, data structure- and provider relation.*

5.1.2 Product-Version Model

The product-version model should contain all attributes whose variation does not lead to a new product.

Product Version and Parents

Each product has at least one version and can have any number of parents, where parents themselves are versions of a product. If a product is a children, it must have at least one parent. Parents can have any number of children. Each version has an internal version number, and if it additionally is a release version, it has a certain release version name (c.f. figures 5.9 and 5.10).

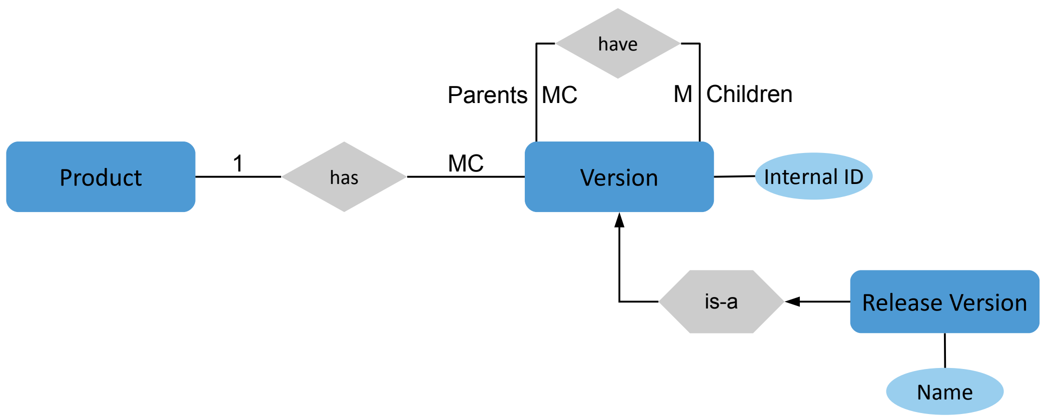


Figure 5.9: ER-Model for product-, version- and parent relation

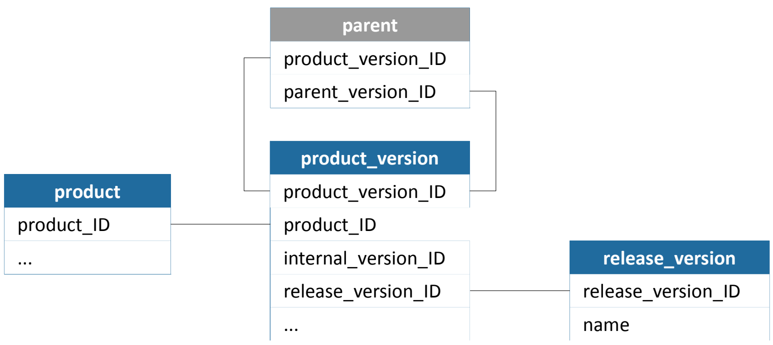


Figure 5.10: Database-Model for product-, version- and parent relation

Processing

Each product is processed by a certain software. This information can be given in different ways: at the IPF, CVS-tags are used to identify the used software version. This information can also be recovered if the processing date is known (see chapter 3.2). For external, a processor name can be given. Depending on the processing, a version has a certain file format (e.g. netCDF, GeoTiff, binary data file, ...).

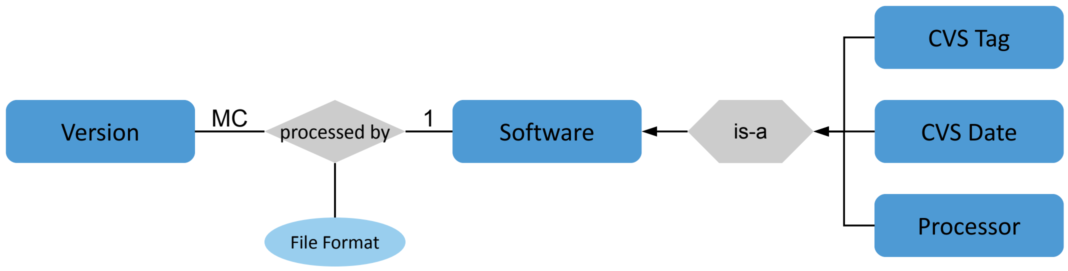


Figure 5.11: ER-Model for processing relations

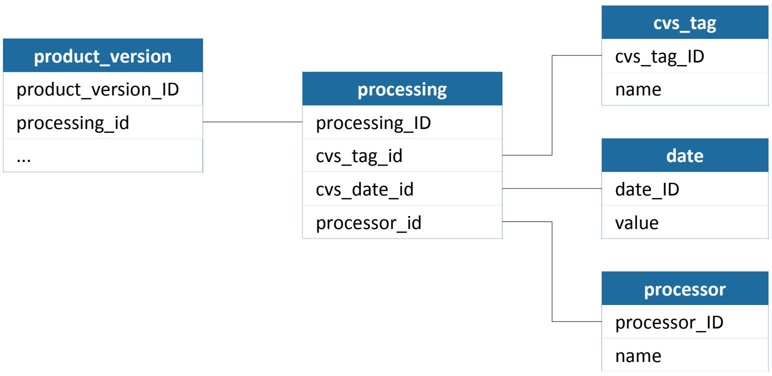


Figure 5.12: Database-Model for processing relations

Coverage

The Coverage is represented by a spatial and a temporal extent. Each product version has at least one coverage (c.f. figures 5.13 and 5.14).

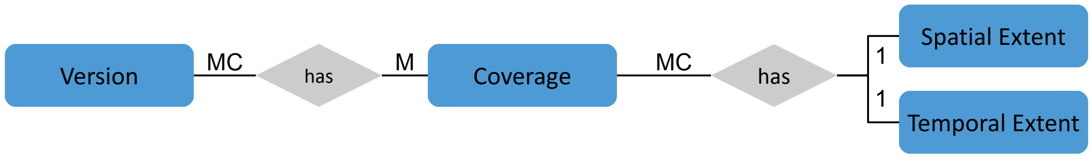


Figure 5.13: ER-Model for coverage

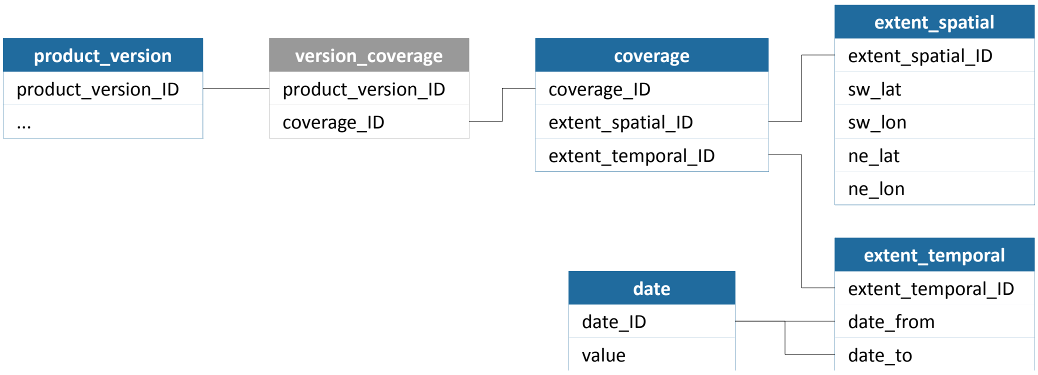


Figure 5.14: Database-Model for coverage

5.2 Metadata Information System

A metadata database as shown in the previous sections can serve as a basis for a powerful metadata information system by enabling the interplay with a web application and XML files. The following example demonstrates a possible schema of such a metadata information system.

The metadata information system shown in figure 5.15 consists of two major components, a database system and a web server. The DBS serves as metadata source and moreover provides functions to extract information from XML files into the database and further to generate XML files out of the stored data. The web server establishes the connection to the DBS and hosts a content management system (CMS). This CMS consists of two parts:

- The *Back-End* is provides functions to add, edit and delete products via a graphical user interface (**b**). It further provides the possibility to import ISO 19115 and INSPIRE compliant XML files into the database (**a**). Access to the Back-End is restricted.
- The *Front-End* serves as a product catalogue for third party users. It allows browsing through the product database (**c** and **d**) and additionally provides the possibility to export product metadata into an ISO 19115 or INSPIRE compliant XML file (**e**). Access to the Front-End can be open or restricted to a specific group.

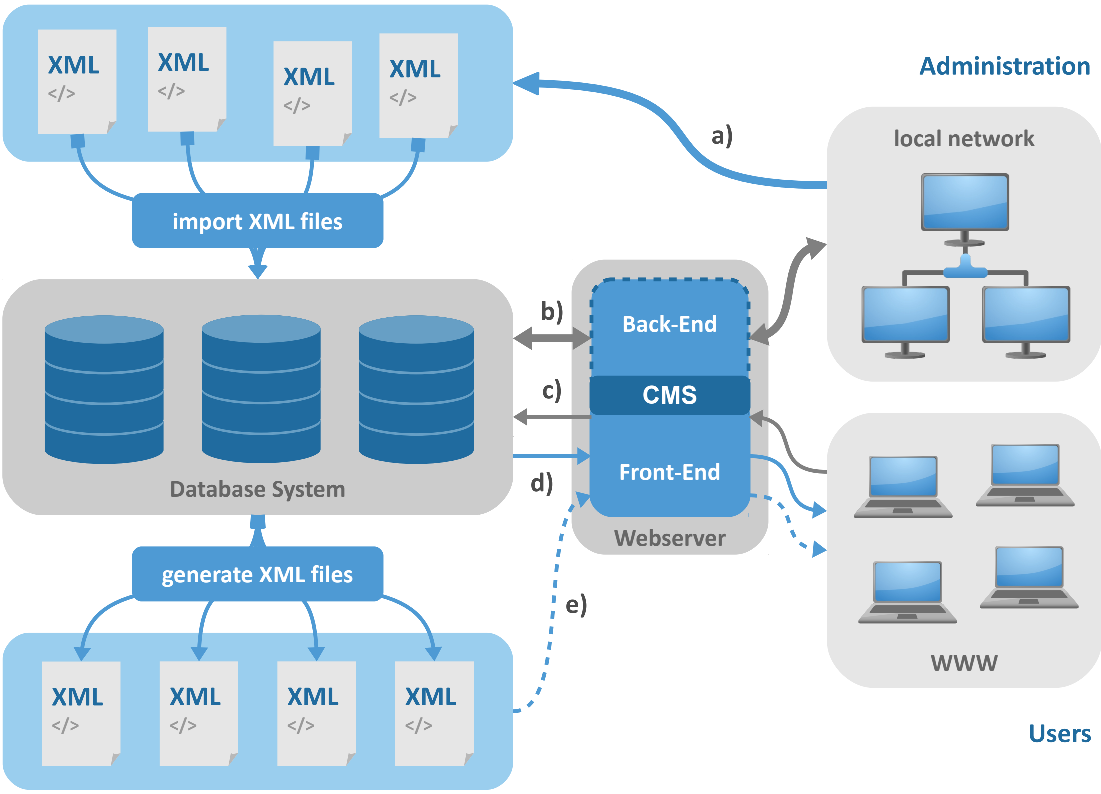


Figure 5.15: *Schema of a metadata information system*
a) *importing data via XML file* ***b)*** *inserting and browsing data through a web interface* ***c)*** *sending a request to the web server* ***d)*** *retrieving requested information* ***e)*** *exporting requested information to an XML-file (optional)*

Chapter 6

Summary and Outlook

This work investigated methods of metadata management for remote sensing products. It discussed existing standards for geographic metadata and demonstrated how to implement compliant metadata files, using XML language. Further a possible implementation of a metadata database was shown and finally a schema for a metadata information system was presented.

However, the way to a final metadata information system is long, and, apart from its technical implementation, it raises questions that are still up to discussion:

File naming (and in the following product naming) is still an open issue since the numerous existing conventions sometimes show significant differences, as comparison between the WMO file naming convention and the example in chapter 3.3 shows. The possibilities of harmonization may be a topic for further investigation.

In chapters 5.1.1 and 5.1.2 it was stated that the a product should contain all the attributes that uniquely define the product, and version model should contain all attributes whose variation does not lead to a new product. However, it may not always be clear where to draw line between product- and version attributes. This boundary may vary depending on the perception of the users or the requirements of a project.

6.1 Future Work

It has been shown that the existing standards for describing geographic metadata (ISO 19115 and ISO 19115-2) do not provide enough information to describe satellite based remote sensing products sufficiently. Therefore, a community metadata profile of ISO 19115 could be created, taking account of existing community profiles for spatial data such as, for example:

- Marine Community Profile of ISO 19115[1]

[1] https://marinemetadata.org/references/marineprofile19115

- North American Profile of ISO 19115[2]
- ANZLIC Metadata Profile[3]

This profile should include all attributes as listed in chapter 3 and moreover give clear definitions of products and product versions. The frame of this work could further include implementation of a naming convention for satellite based remote sensing products and datasets.

[2] http://www.fgdc.gov/standards/projects/incits-l1-standards-projects/NAP-Metadata
[3] http://www.anzlic.org.au/Publications/Metadata+Project/248.aspx

Bibliography

Beynon-Davies, P. (2004), *Database Systems*, Palgrave.

Bray, T. (2006), Extensible markup language (XML) 1.0, http://www.w3.org/TR/2006/REC-xml-20060816/, accessed on 2012 March 21.

Chen, P. P.-S. (1976), The entity-relationship model-toward a unified view of data, *ACM Transactions on Database Systems*, **1**(1), p. 9–36.

Clowse, B. B. (1981), *Brainpower for the Cold War: The sputnik crisis and National Defense Education Act of 1958*, vol. 3, Greenwood Press (Westport, Conn.).

ESRI (2012), GIS Dictionary, http://support.esri.com/en/knowledgebase/GISDictionary/term/radiometric%20resolution, accessed on 2012 June 3.

EUMETSAT (2012), Metop global data service, http://www.eumetsat.int/Home/Main/Satellites/Metop/Services/SP_2010034162729623?l=en, accessed on 2012 April 13.

Exceution-MiH (2012), Metadata objective and purpose - why metadata, http://www.executionmih.com/metadata/objective-purpose.php, accessed on 2012 April 2.

FileInfo.com (2009), .DAT file extension, http://www.fileinfo.com/extension/dat, accessed on 2012 May 23.

Habermann, T. (2012), ISO FAQ – NOAA environmental data management wiki, https://geo-ide.noaa.gov/wiki/index.php?title=ISO_FAQ, accessed on 2012 May 16.

Hahn, S., W. Wagner, R. Kidd, S. Hasenauer, T. Melzer, C. Paulik, and C. Reimer (2012), Improvements and challenges in the METOP ASCAT soil moisture scheme, Poster – AGU Chapman conference on Remote Sensing of the Water Cycle.

INSPIRE (2007), *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)*, The European Parliament and the Council of the European Union.

INSPIRE (2010), *INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119*, European Comission – Drafting Team Metadata and European Commission Joint Research Centre.

INSPIRE (2012), About INSPIRE, `http://inspire.jrc.ec.europa.eu/index.cfm/pageid/48`, accessed on 2012 January 24.

ISO (2002), International standard ISO 19108:2002, geographic information – temporal schema.

ISO (2003), International standard ISO 19115:2003, geographic information – metadata.

ISO (2007a), International standard ISO 19136:2007 geographic information – geography markup language (gml).

ISO (2007b), International standard ISO 19139:2007 geographic information – metadata – xml schema implementation.

ISO (2009), International standard ISO 19115-2:2009, geographic information – metadata – part 2: Extensions for imagery and gridded data.

ISO (2011), International standard ISO 19118:2011 geographic information – encoding.

ISO (2012), About ISO, `http://www.iso.org/iso/`, accessed on 2012 January 23.

Ludwig, G. H., and W. A. Whelpley (1960), Corpuscular radiation experiment of satellite 1959 Iota (Explorer VII), *Journal of Geophysical Research*, **65**(4), p. 1119–1124.

Newhall, B. (1982), *The History of Photography*, 5 ed., The Museum of Modern Art.

ON (2009a), ÖNORM EN ISO 19136, geographic information – geography markup language (GML).

ON (2009b), ONR CEN ISO/TS 19139, geoinformation – metadaten – XML-schema-implementierung.

Parkinson, C. L., A. Ward, and M. D. King (2006), *Earth Science Reference Handbook – A Guide to NASA's Earth Science Program and Earth Observing Satellite Missions*, NASA, Washington, D.C.

Rees, G. (2001), *Physical principles of remote sensing*, Topics in remote sensing, Cambridge University Press.

Refsnes-Data (2012), W3Schools – XML tutorial, `http://www.w3schools.com/xml/`, accessed on 2012 May 14.

Reichhardt, T. (2006), The first photo from space, *Air & Space magazine*, **November 2006**.

Siddigi, A. A. (2003), *Sputnik and the Soviet Space Challenge*, 1 ed., University Press of Florida.

Skolnik, M. I. (1990), *Radar Handbook*, 2 ed., McGraw-Hill Professional.

SSEC (2009), 50 year anniversary of Explorer 7 launch, `http://www.ssec.wisc.edu/media/spotlight/explorer7.html`, accessed on 2012 May 24.

WMO (Ed.) (2008), *Application of the ISO 191xx series of geographic information standards to the development of a WMO conceptual model of data representation*, World Meteorological Organization – Comission for Basic Systems OPAG on Information Systems & Servics – Expert Team on Assessment of Data Representation Systems, Washington, USA.

WMO (2009), *Manual on the Global Telecommunication System – Volume I (Annex III to WMO Technical Regulations) – WMO-No. 386*, World Meteorological Organization, 2009 ed.

Appendix A

INSPIRE compliant XML file

```xml
<?xml version="1.0" encoding="UTF-8"?>
<gmd:MD_Metadata xmlns:gmd="http://www.isotc211.org/2005/gmd" xmlns:gco="http://www.isotc211.org/2005/gco" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:gml="http://www.opengis.net/gml" xmlns:xlink="http://www.w3.org/1999/xlink" xsi:schemaLocation="http://www.isotc211.org/2005/gmd http://schemas.opengis.net/iso/19139/20060504/gmd/gmd.xsd">
    <gmd:fileIdentifier>
        <gco:CharacterString>ASCAT_AMSRE_L4_SSM_V01</gco:CharacterString>
    </gmd:fileIdentifier>
    <gmd:language>
        <gmd:LanguageCode codeList="http://www.loc.gov/standards/iso639-2/"
            codeListValue="eng">eng</gmd:LanguageCode>
    </gmd:language>
    <gmd:characterSet>
        <gmd:MD_CharacterSetCode codeSpace="ISOTC211/19115" codeListValue="
            MD_CharacterSetCode_utf8" codeList="http://www.isotc211.org/2005/resources
            /Codelist/gmxCodelists.xml#MD_CharacterSetCode">MD_CharacterSetCode_utf8</
            gmd:MD_CharacterSetCode>
    </gmd:characterSet>
    <gmd:hierarchyLevel>
        <gmd:MD_ScopeCode codeList="http://standards.iso.org/ittf/
            PubliclyAvailableStandards/ISO_19139_Schemas/resources/Codelist/
            ML_gmxCodelists.xml#MD_ScopeCode" codeListValue="dataset">dataset</
            gmd:MD_ScopeCode>
    </gmd:hierarchyLevel>
    <gmd:contact>
        <gmd:CI_ResponsibleParty>
            <gmd:organisationName>
                <gco:CharacterString>I.P.F, Vienna</gco:CharacterString>
            </gmd:organisationName>
            <gmd:contactInfo>
                <gmd:CI_Contact>
                    <gmd:address>
                        <gmd:CI_Address>
                            <gmd:electronicMailAddress>
```

```xml
                <gco:CharacterString>wd@ipf.tuwien.ac.at</gco:CharacterString>
              </gmd:electronicMailAddress>
            </gmd:CI_Address>
          </gmd:address>
        </gmd:CI_Contact>
      </gmd:contactInfo>
      <gmd:role>
        <gmd:CI_RoleCode codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/Codelist/ML_gmxCodelists.xml#CI_RoleCode" codeListValue="pointOfContact">pointOfContact</gmd:CI_RoleCode>
      </gmd:role>
    </gmd:CI_ResponsibleParty>
  </gmd:contact>
  <gmd:dateStamp>
    <gco:Date>2012-02-09</gco:Date>
  </gmd:dateStamp>
  <gmd:metadataStandardName>
    <gco:CharacterString>ISO19115</gco:CharacterString>
  </gmd:metadataStandardName>
  <gmd:metadataStandardVersion>
    <gco:CharacterString>2003/Cor.1:2006</gco:CharacterString>
  </gmd:metadataStandardVersion>
  <gmd:identificationInfo>
    <gmd:MD_DataIdentification>
      <gmd:citation>
        <gmd:CI_Citation>
          <gmd:title>
            <gco:CharacterString>ASCAT AMSRE L4 SSM</gco:CharacterString>
          </gmd:title>
          <gmd:date>
            <gmd:CI_Date>
              <gmd:date>
                <gco:Date>2010-08-03</gco:Date>
              </gmd:date>
              <gmd:dateType>
                <gmd:CI_DateTypeCode codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/Codelist/ML_gmxCodelists.xml#CI_DateTypeCode" codeListValue="creation">creation</gmd:CI_DateTypeCode>
              </gmd:dateType>
            </gmd:CI_Date>
          </gmd:date>
          <gmd:identifier>
```

```xml
                    <gmd:RS_Identifier>
                        <gmd:code>
                            <gco:CharacterString>ASCAT_AMSRE_L4_SSM_V01</gco:CharacterString>
                        </gmd:code>
                    </gmd:RS_Identifier>
                </gmd:identifier>
            </gmd:CI_Citation>
        </gmd:citation>
        <gmd:abstract>
            <gco:CharacterString>Combining information derived from satellite-based passive and active microwave sensors has the potential to offer improved estimates of surface soil moisture at global scale. We develop and evaluate a methodology that takes advantage of the retrieval characteristics of passive (AMSR-E) and active (ASCAT) microwave satellite estimates to produce an improved soil moisture product in flat_earth projection.</gco:CharacterString>
        </gmd:abstract>
        <gmd:pointOfContact>
            <gmd:CI_ResponsibleParty>
                <gmd:organisationName>
                    <gco:CharacterString>TU Wien, Vienna</gco:CharacterString>
                </gmd:organisationName>
                <gmd:contactInfo>
                    <gmd:CI_Contact>
                        <gmd:address>
                            <gmd:CI_Address>
                                <gmd:electronicMailAddress>
                                    <gco:CharacterString>wd@ipf.tuwien.ac.at</gco:CharacterString>
                                </gmd:electronicMailAddress>
                            </gmd:CI_Address>
                        </gmd:address>
                    </gmd:CI_Contact>
                </gmd:contactInfo>
                <gmd:role>
                    <gmd:CI_RoleCode codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/Codelist/ML_gmxCodelists.xml#CI_RoleCode" codeListValue="author">author</gmd:CI_RoleCode>
                </gmd:role>
            </gmd:CI_ResponsibleParty>
        </gmd:pointOfContact>
        <gmd:pointOfContact>
            <gmd:CI_ResponsibleParty>
                <gmd:organisationName>
```

```
      <gco:CharacterString>I.P.F, Vienna</gco:CharacterString>
    </gmd:organisationName>
    <gmd:contactInfo>
      <gmd:CI_Contact>
        <gmd:address>
          <gmd:CI_Address>
            <gmd:electronicMailAddress>
              <gco:CharacterString>dc@ipf.tuwien.ac.at</gco:CharacterString>
            </gmd:electronicMailAddress>
          </gmd:CI_Address>
        </gmd:address>
      </gmd:CI_Contact>
    </gmd:contactInfo>
  </gmd:CI_ResponsibleParty>
</gmd:pointOfContact>
<gmd:descriptiveKeywords>
  <gmd:MD_Keywords>
    <gmd:keyword>
      <gco:CharacterString>Soil Moisture</gco:CharacterString>
    </gmd:keyword>
    <gmd:thesaurusName>
      <gmd:CI_Citation>
        <gmd:title>
          <gco:CharacterString>GEOSS - Earth Observation Vocabulary, version 1.0</gco:CharacterString>
        </gmd:title>
        <gmd:date>
          <gmd:CI_Date>
            <gmd:date>
              <gco:Date>2011-05-01</gco:Date>
            </gmd:date>
            <gmd:dateType>
              <gmd:CI_DateTypeCode codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/Codelist/ML_gmxCodelists.xml#CI_DateTypeCode" codeListValue="publication">publication</gmd:CI_DateTypeCode>
            </gmd:dateType>
          </gmd:CI_Date>
        </gmd:date>
      </gmd:CI_Citation>
    </gmd:thesaurusName>
  </gmd:MD_Keywords>
</gmd:descriptiveKeywords>
```

```xml
            <gmd:descriptiveKeywords>
                <gmd:MD_Keywords>
                    <gmd:keyword>
                        <gco:CharacterString>Soil</gco:CharacterString>
                    </gmd:keyword>
                    <gmd:thesaurusName>
                        <gmd:CI_Citation>
                            <gmd:title>
                                <gco:CharacterString>GEMET - INSPIRE themes, version 1.0</gco:CharacterString>
                            </gmd:title>
                            <gmd:date>
                                <gmd:CI_Date>
                                    <gmd:date>
                                        <gco:Date>2008-06-01</gco:Date>
                                    </gmd:date>
                                    <gmd:dateType>
                                        <gmd:CI_DateTypeCode codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/Codelist/ML_gmxCodelists.xml#CI_DateTypeCode" codeListValue="publication">publication</gmd:CI_DateTypeCode>
                                    </gmd:dateType>
                                </gmd:CI_Date>
                            </gmd:date>
                        </gmd:CI_Citation>
                    </gmd:thesaurusName>
                </gmd:MD_Keywords>
            </gmd:descriptiveKeywords>
            <gmd:resourceConstraints>
                <gmd:MD_Constraints>
                    <gmd:useLimitation>
                        <gco:CharacterString>No conditions apply</gco:CharacterString>
                    </gmd:useLimitation>
                </gmd:MD_Constraints>
            </gmd:resourceConstraints>
            <gmd:resourceConstraints>
                <gmd:MD_LegalConstraints>
                    <gmd:accessConstraints>
                        <gmd:MD_RestrictionCode codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/Codelist/gmxCodelists.xml#MD_RestrictionCode" codeListValue="otherRestrictions">otherRestrictions</gmd:MD_RestrictionCode>
                    </gmd:accessConstraints>
```

```
                <gmd:otherConstraints>
                    <gco:CharacterString>no limitations</gco:CharacterString>
                </gmd:otherConstraints>
            </gmd:MD_LegalConstraints>
        </gmd:resourceConstraints>
        <gmd:spatialResolution>
            <gmd:MD_Resolution>
                <gmd:distance>
                    <gco:Distance uom="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/uom/ML_gmxUom.xml#m">25000</gco:Distance>
                </gmd:distance>
            </gmd:MD_Resolution>
        </gmd:spatialResolution>
        <gmd:language>
            <gmd:LanguageCode codeList="http://www.loc.gov/standards/iso639-2/" codeListValue="eng">eng</gmd:LanguageCode>
        </gmd:language>
        <gmd:topicCategory>
            <gmd:MD_TopicCategoryCode>geoscientificInformation</gmd:MD_TopicCategoryCode>
        </gmd:topicCategory>
        <gmd:topicCategory>
            <gmd:MD_TopicCategoryCode>climatologyMeteorologyAtmosphere</gmd:MD_TopicCategoryCode>
        </gmd:topicCategory>
        <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:temporalElement>
                    <gmd:EX_TemporalExtent>
```

```xml
                        <gmd:extent>
                            <gml:TimePeriod gml:id="extent">
                                <gml:beginPosition>1978-01-01</gml:beginPosition>
                                <gml:endPosition>2010-12-31</gml:endPosition>
                            </gml:TimePeriod>
                        </gmd:extent>
                    </gmd:EX_TemporalExtent>
                </gmd:temporalElement>
            </gmd:EX_Extent>
        </gmd:extent>
    </gmd:MD_DataIdentification>
</gmd:identificationInfo>
<gmd:distributionInfo>
    <gmd:MD_Distribution>
        <gmd:distributionFormat>
            <gmd:MD_Format>
                <gmd:name>
                    <gco:CharacterString>unknown</gco:CharacterString>
                </gmd:name>
                <gmd:version>
                    <gco:CharacterString>unknown</gco:CharacterString>
                </gmd:version>
            </gmd:MD_Format>
        </gmd:distributionFormat>
        <gmd:transferOptions>
            <gmd:MD_DigitalTransferOptions>
                <gmd:onLine>
                    <gmd:CI_OnlineResource>
                        <gmd:linkage>
                            <gmd:URL>http://wacmos.itc.nl/?q=node/5</gmd:URL>
                        </gmd:linkage>
                    </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 codeListValue="dataset" codeList="http://
                        standards.iso.org/ittf/PubliclyAvailableStandards/
                        ISO_19139_Schemas/resources/Codelist/ML_gmxCodelists.xml#
                        MD_ScopeCode">dataset</gmd:MD_ScopeCode>
```

```xml
                </gmd:level>
            </gmd:DQ_Scope>
        </gmd:scope>
        <gmd:report>
            <gmd:DQ_DomainConsistency xsi:type="gmd:DQ_DomainConsistency_Type">
                <gmd:result>
                    <gmd:DQ_ConformanceResult xsi:type="
                        gmd:DQ_ConformanceResult_Type">
                        <gmd:specification>
                            <gmd:CI_Citation>
                                <gmd:title>
                                    <gco:CharacterString>COMMISSION REGULATION (EC)
                                         No 1205/2008 of 3 December 2008
                                        implementing Directive 2007/2/EC of the
                                        European Parliament and of the Council as
                                        regards metadata</gco:CharacterString>
                                </gmd:title>
                                <gmd:date>
                                    <gmd:CI_Date>
                                        <gmd:date>
                                            <gco:Date>2008-12-04</gco:Date>
                                        </gmd:date>
                                        <gmd:dateType>
                                            <gmd:CI_DateTypeCode codeList="http://
                                                standards.iso.org/ittf/
                                                PubliclyAvailableStandards/
                                                ISO_19139_Schemas/resources/
                                                Codelist/ML_gmxCodelists.xml#
                                                CI_DateTypeCode" codeListValue="
                                                publication">publication</
                                                gmd:CI_DateTypeCode>
                                        </gmd:dateType>
                                    </gmd:CI_Date>
                                </gmd:date>
                            </gmd:CI_Citation>
                        </gmd:specification>
                        <gmd:explanation>
                            <gco:CharacterString>See the referenced specification</
                                gco:CharacterString>
                        </gmd:explanation>
                        <gmd:pass>
                            <gco:Boolean>true</gco:Boolean>
                        </gmd:pass>
                    </gmd:DQ_ConformanceResult>
                </gmd:result>
            </gmd:DQ_DomainConsistency>
```

```xml
                </gmd:report>
                <gmd:lineage>
                    <gmd:LI_Lineage>
                        <gmd:statement>
                            <gco:CharacterString>Volumetric soil water content (m\^3 m\^-3)
                                from AMSR-E and degree of saturation (\%) from ASCAT are
                                rescaled against a reference land surface model data set
                                using a cumulative distribution function matching approach
                                . While this imposes any bias of the reference on the
                                rescaled satellite products, it adjusts them to the same
                                range and preserves the dynamics of original satellite-
                                based products. Comparison with in situ measurements
                                demonstrates that where the correlation coefficient
                                between rescaled AMSR-E and ASCAT is greater than 0.6 ("
                                transitional region"), merging the different satellite
                                products increases the number of observations while
                                minimally changing the accuracy of soil moisture
                                retrievals. The error characterisation is represented by
                                the variable "sm\_noise" in the header of the NetCDF file.
                                For full description of the errors see Dorigo et al.
                                (2010; doi:10.5194/hess-14-2605-2010)
                            </gco:CharacterString>
                        </gmd:statement>
                    </gmd:LI_Lineage>
                </gmd:lineage>
            </gmd:DQ_DataQuality>
        </gmd:dataQualityInfo>
</gmd:MD_Metadata>
```