



## Collaborative Climate Community Data and Processing Grid (C3-Grid)

### C3Grid ISO 19115 Metadata Profile

---

Work package:	WP3 – Metadata
Authors:	C3Grid Metadata Working Group
Editors:	S.Kindermann, M.Stockhause
Version:	1.1 (07/2008)
Publishing date:	July 2008
WP-Coordination:	DKRZ
Pages:	10
Partner:	M&D, HB-Cluster, Diag.-Cluster, DLR, PIK, DWD
Contact:	S.Kindermann
Email:	kindermann@dkrz.de

---

GEFÖRDERT VOM



Bundesministerium  
für Bildung  
und Forschung

*project within the D-Grid community  
promoted by BMBF*



**Reference to project plan:**

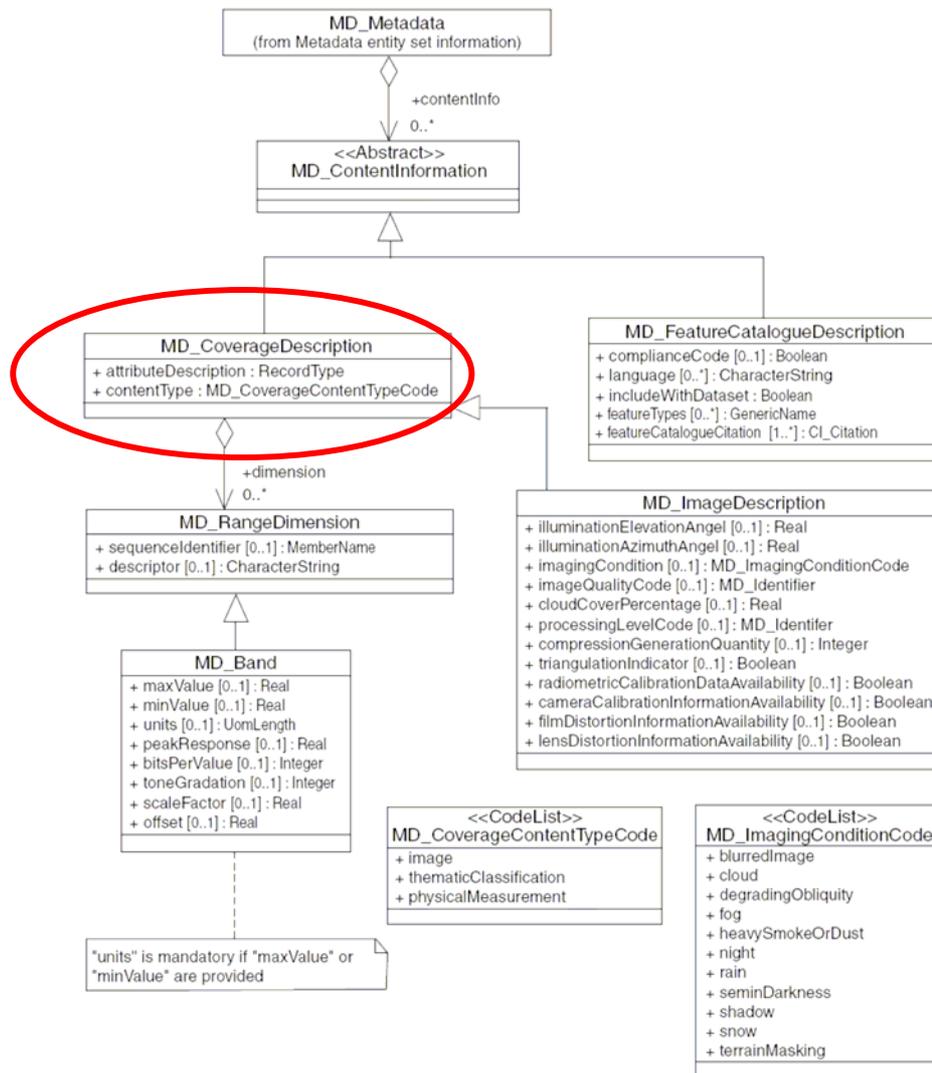
This report summarizes the work on the C3Grid adoption of the ISO 19115 standard as part of Work Package 3. The key components are summarized, based on the ISO 19115 UML definitions (chapter 1). Additionally, some key sections of the C3Grid profile are explained with XML code examples (chapter 2). Finally, the complete C3Grid ISO 19115 profile is described by a catalogue along with a C3Grid ISO 19139 XML template (Appendices A and B).





The **MD\_ContentInformation** part of ISO19115 (see Figure 4) is used in C3Grid to characterize the physical properties of the described data sets by means of the widely accepted CF (Climate and Forecast [5]) standard names catalogue. This provides a unique description of climate parameters used in climate data sets.

C3Grid decided to define no additional feature catalogues, but to use the **MD\_CoverageDescription** part to specify this information. Therein CF standard names are specified as **attributeDescriptions** and its unit as **dimension**. In the **dimension** section the vertical coordinate reference system (CRS) on which it is defined (**verticalElement** section) is referenced via the xlink:href attribute (see example in section 2.1).



**Figure 4:** ISO 19115 content part with coverage description (holding CF variables in C3Grid)

Within C3Grid metadata, information how and where to access the data at the data providers, their sizes and available formats have to be described as well. This information is given in the **MD\_DigitalTransferOptions** and **MD\_Format** sections (see Figure 5).

The **MD\_DigitalTransferOptions** part references the two data access interfaces, which were developed in C3Grid: The web service interface reference, where data can be ordered, as well as the C3Grid MDS reference. In the MDS technical metadata for the data providers are specified, e.g. the local part of the C3Grid workspace, where the data are accessible by the C3Grid data management service (see example in section 2.2). Other non-C3Grid data accesses are describable in this section as well (e.g. GridFTP or openDAP accesses).

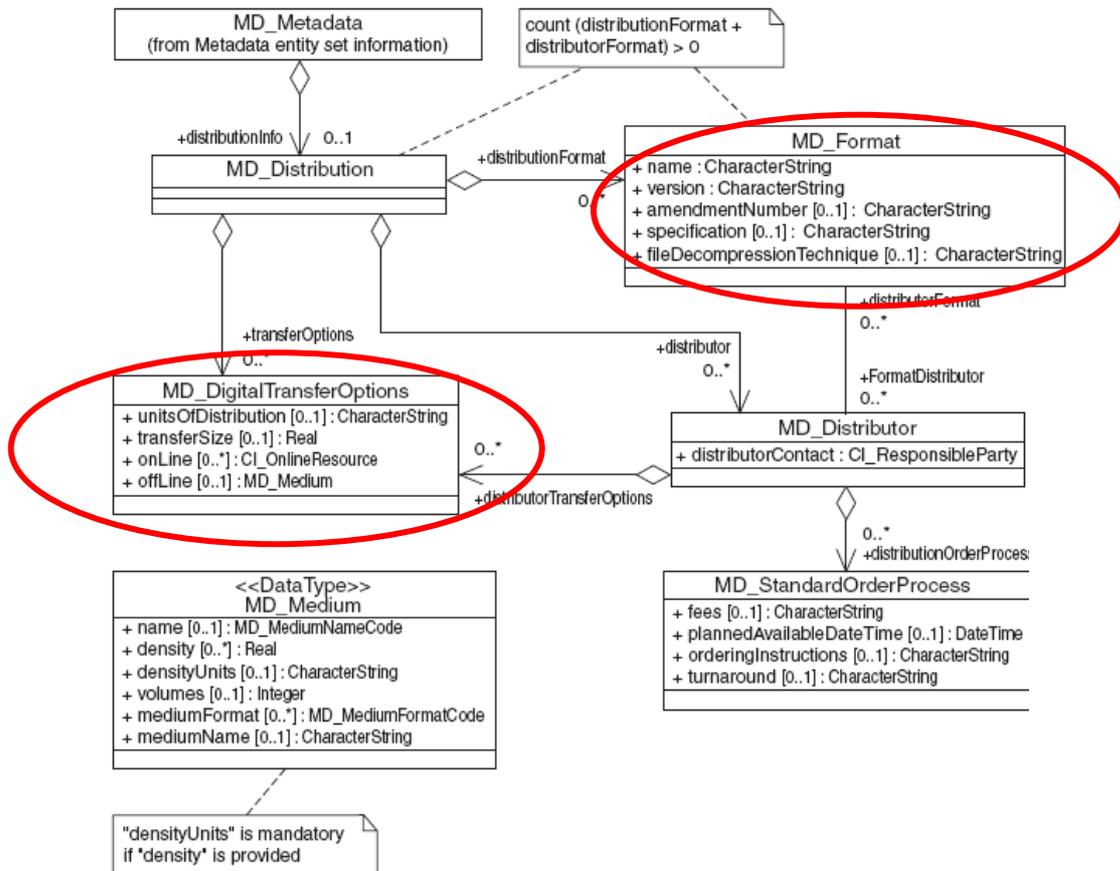


Figure 5: ISO parts holding C3Grid data access information

One important aspect in C3Grid is the reconstructability of newly generated data and thus clear descriptions of data processing history as well as data origin are important parts of a concise data quality description. This information is collected in the ISO 19115 **LI\_Lineage** part of **DQ\_DataQuality** (see Figure 6 and example in section 2.3). The processing history in C3Grid is recorded in individual process step descriptions (**LI\_ProcessStep**), where the original data is referenced in the **source/LI\_Source** section.

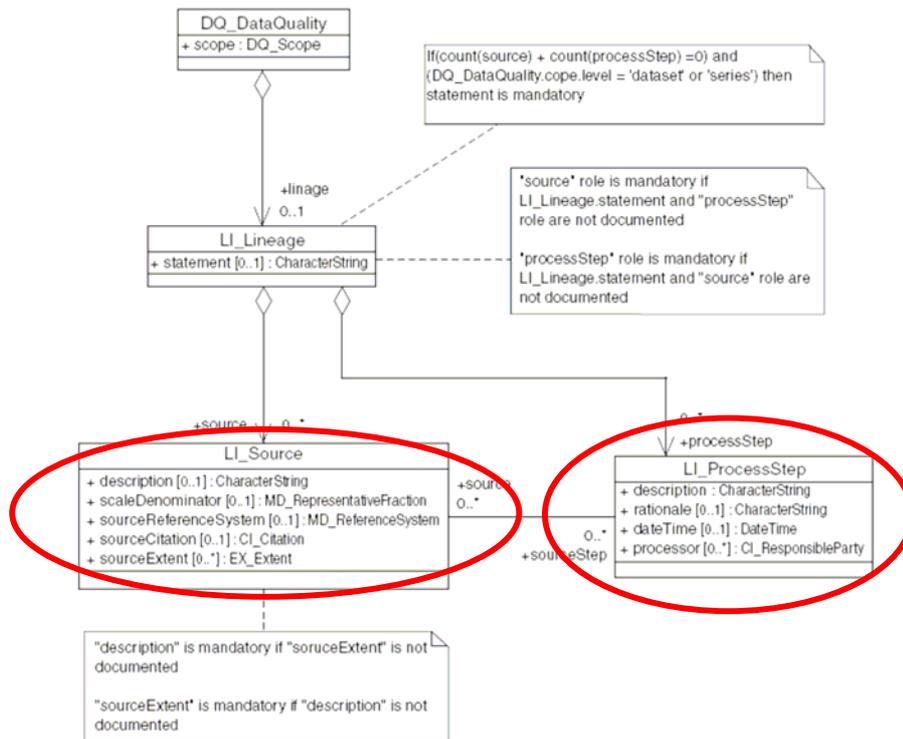


Figure 6: ISO 19115 parts holding C3Grid data provenance information

## 2 C3Grid Metadata Profile Examples

The above described metadata profile does not describe all aspects of the data in C3Grid. It covers the discovery metadata plus some additional information needed for an intelligent data access with data amount reduction functionality.

Figure 7 gives an overview over the roles of the main parts of the C3Grid metadata profile in the C3Grid implementation of generation 2:

- Information of the user
- Users' search
- Users' selection of parts of data
- Users' access of selected parts of the data

The parts of the C3Grid profile printed in blue are explained as selected central parts in the following sections.

### 2.1 Content & Vertical Coordinate Reference System

C3Grid defined a vertical coordinate reference system (CRS) catalogue as references for the different used systems in the available data groups. Since C3Grid uses aggregated data for data discovery and access, several different vertical CRS' can be present in a single metadata description. For raw data of model applications usually all content parameters are present in each dataset and thus also all used reference systems for the spatial representation of the variable.

Therefore an ID for each vertical CRS is introduced, which is referenced in the content section. An example is shown in Figure 8.

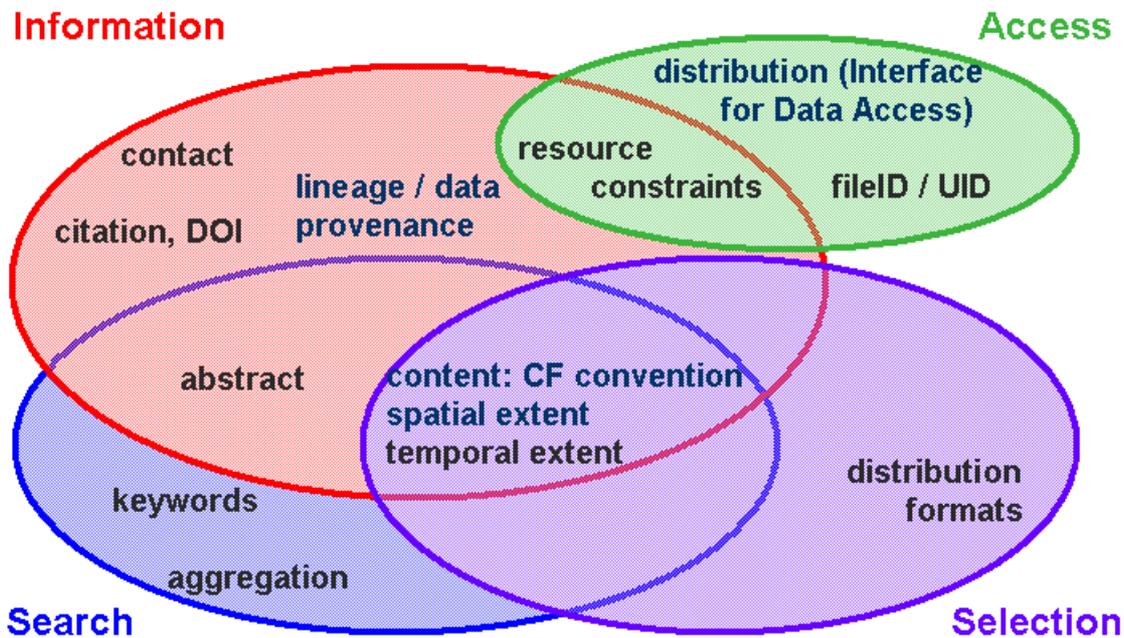


Figure 7: Roles of the main parts of the C3Grid metadata profile in the generation 2 implementation; blue sections are given as examples in subsections of chapter 2.

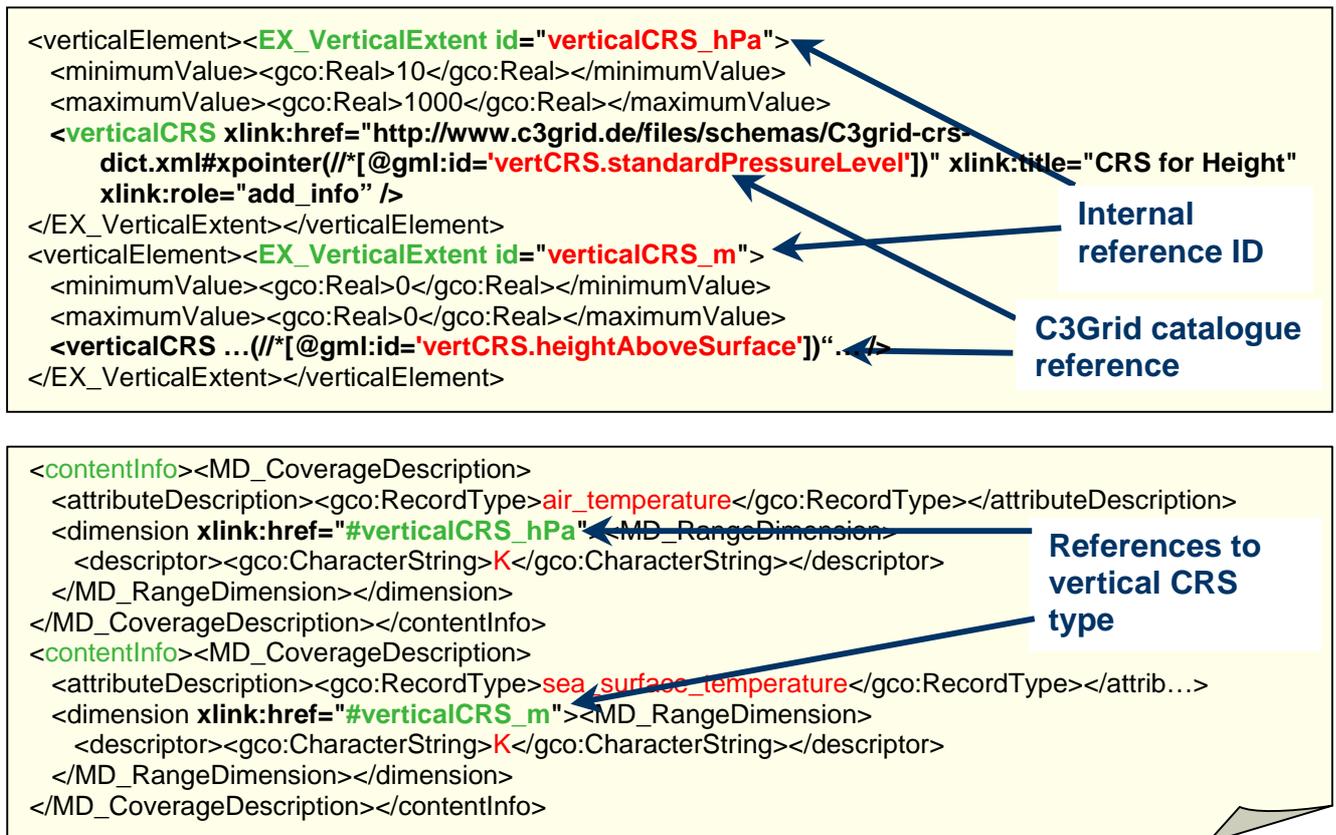


Figure 8: C3Grid vertical CRS catalogue and internal reference from content section.

## 2.2 Distribution / Interface for Data Access

As a second example the distribution section is presented. In the C3Grid metadata two entries are specified:

- C3Grid WebMDS: information about configuration at the data provider, e.g. location of local C3Grid workspace (upper part of Figure 9)
- WebService interface for data access (lower part of Figure 9)

```
<distributor><MD_Distributor>
  <distributorTransferOptions><MD_DigitalTransferOptions>
    <onLine><CI_OnlineResource>
      <linkage><URL>
        http://c3grid-gt.e-technik.uni-dortmund.de:8080/webmds/webmds?info=indexinfo
      </URL></linkage>
      <protocol><gco:CharacterString>http</gco:CharacterString></protocol>
      <name>
        <gco:CharacterString>C3Grid_ResourceInformationService</gco:CharacterString>
      </name>
      <description><gco:CharacterString>
        Production System: DKRZ C3Grid Data Request Interface Implementation
      </gco:CharacterString></description>
      <function><CI_OnlineFunctionCode .....>order</CI_OnlineFunctionCode></function>
    </CI_OnlineResource></onLine>
  </MD_DigitalTransferOptions></distributorTransferOptions>
</MD_Distributor></distributor>
```

```
...
  <onLine><CI_OnlineResource>
    <linkage><URL>
      https://anticyclone.dkrz.de:9443/axis/services/C3ProviderSOAP
    </URL></linkage>
    <protocol><gco:CharacterString>SOAP</gco:CharacterString></protocol>
    <name>
      <gco:CharacterString>C3Grid_DataProvider_Webservice</gco:CharacterString>
    </name>
    <description><gco:CharacterString>
      C3Grid Data Provider Webservice at DKRZ
    </gco:CharacterString></description>
    <function><CI_OnlineFunctionCode .....>order</CI_OnlineFunctionCode></function>
  </CI_OnlineResource></onLine>
</MD_DigitalTransferOptions></distributorTransferOptions>
</MD_Distributor></distributor>
```

Figure 9: Distribution metadata in C3Grid: upper part the WebMDS entry and lower part the description of the data access webservice interface

## 2.3 Provenance / Lineage

As a last example, the prototypical usage of the lineage section of ISO 19115 is presented, in which coarse provenance information is stored. In the C3Grid implementation of generation 2, the basic information about the alteration steps is described, which the present dataset has undergone. At the moment, these information is free text and unchecked. The source dataset for each processing step is referenced via its universal C3Grid metadata identifier (fileIdentifier). The extent of the information about a processing step is to be approved and improved in the operational C3Grid (Figure 10).

```

<lineage><LI_Lineage>
  <statement>
    <gco:CharacterString>Model Run, C3Grid preprocessed data (DKRZ, MPI-M)
  </gco:CharacterString>
</statement>
  <processStep><LI_ProcessStep id="Step_0">
    <description>
      <gco:CharacterString>Model Run</gco:CharacterString>
    </description>
    <dateTime><gco:DateTime>2004-07-12T14:19:19</gco:DateTime></dateTime>
  </LI_ProcessStep></processStep>
  <processStep><LI_ProcessStep id="Step_1">
    <description>
      <gco:CharacterString>C3Grid Data Preprocessing at MPI-M/IFM-GEOMAR/DKRZ:
cera_dkrz.py, version P0.1, 2008-01-15 (martina.stockhause AT zmaw.de) with Experiment:
'de.dkrz.wdcc.iso2093450'</gco:CharacterString>
    </description>
    <dateTime><gco:DateTime>2008-03-20T13:21:51</gco:DateTime></dateTime>
    <source><LI_Source><description>
      <gco:CharacterString>de.dkrz.wdcc.iso2093450</gco:CharacterString>
    </description></LI_Source></source>
  </LI_ProcessStep> </processStep>
</LI_Lineage></lineage>

```

Figure 10: ISO 19115 lineage section for provenance information in the initial implementation of C3Grid generation 2.

#### References:

- [1] ISO 19115:2003 "Geographic information – Metadata"; ISO standard
- [2] ISO/PRF TS 19139 "Geographic information – Metadata – XML schema implementation"; ISO technical standard, <http://eden.ign.fr/xsd/isotc211/iso19115/20060504>
- [3] Version 0.3 of WMO Core Metadata Standard, July 2006, <http://wis.wmo.int/2006/metadata/WMOCoreTextVer0.3.doc>
- [4] INSPIRE metadata implementing rules based on ISO 19115 and ISO 19119, draft, 25<sup>th</sup> April 2008, MD\_IR\_and\_IOS\_20080425.pdf, [http://inspire.jrc.it/reports/ImplementingRules/metadata/Draft\\_Guidelines%20INSPIRE\\_metadata\\_implementing\\_rules.pdf](http://inspire.jrc.it/reports/ImplementingRules/metadata/Draft_Guidelines%20INSPIRE_metadata_implementing_rules.pdf)
- [5] CF Standard Names, <http://cf-pcmdi.llnl.gov/documents/cf-standard-names>

## Appendix A) C3Grid Metadata Catalogue

## Appendix B) C3Grid XML template