Directory Interchange Format (DIF) Metadata and Handling at the German Research Center for Geosciences' Information System and Data Center

By Vivien Mende,[1] Bernd Ritschel,[1] Sebastian Freiberg,[1] Hartmut Palm,[1] and Lutz Gericke[1]

[1]Information System and Data Center, German Research Center for Geosciences, Potsdam, Germany.

The Information System and Data Center (ISDC) is managing more than 11 terabytes (TB) of geoscience data and information. Currently, these data are coming from 11 missions, including the German Research Center for Geosciences' Challenging Mini-Satellite Payload (CHAMP), the Gravity Recovery and Climate Experiment (GRACE, a joint partnership between the German Aerospace Center (DLR) and the National Aeronautics and Space Administration (NASA)), Germany's TerraSAR-X (an X-band synthetic aperture radar satellite), the International Association of Geodesy's Global Geodynamic Project (GGP), and others that have yielded nearly 300 product types and approximately 16 million products, which have been made available to more than 1,700 users. This paper gives a short overview about the development and use of metadata in the ISDC. Each product type that results from a geoscience mission or project consists of a set of products. A product is composed of a data file (or files) and a metadata document. Figure 1 shows the three product types resulting from the TerraSAR-X.

In order to describe and manage the products (data file(s) and metadata), we are using an evolution version 9.x of NASA's Directory Interchange Format (DIF) standard. The manager of NASA's Global Change Master Directory (GCMD), Lola Olsen, defines metadata as follows: "Descriptive information that characterizes a set of quantitative and/or qualitative measurements and distinguishes that set from other similar measurement sets." (See http://gcmd.nasa.gov/Aboutus/standards/.) For the management of ISDC product types, DIF is an excellent choice.

The ISDC base schema of the product type DIF XML documents is defined in the "base-dif.xsd" file. The ISDC Extensible Mark-up Language (XML) schema has been defined on the basis of the GCMD XML schema definition found online at http://gcmd.nasa.gov/Aboutus/xml/dif/dif_v9.7.1.xsd. In order to describe single products, it was necessary to extend the DIF standard and to modify the GCMD XML schema. Although the structure of the ISDC DIF XML schema is different from the GCMD schema, the ISDC product type DIF XML documents are still valid in relation to the GCMD schema. Additionally, the ISDC is using a product type-data child DIF combination. The metadata of product types are described in associated product type DIF XML files according to the "base-dif.xsd" schema. The product type DIF XML files are validated and stored in an Oracle XML database. The product- (data file-) specific metadata are documented in data-child DIF XML files. Each product type has its own schema for the data-child DIF XML files. Data-child DIF documents are used to describe the data-file-specific properties. The complex XML type <Data_Parameters> in the data-child DIF XML document provides the specific extension of the product type DIF XML structures. <Data_Parameters> includes specific metadata of the product, such as the data file name, data file size, revision, satellite identification, and other information. In order to implement this data model, we are using the redefined XML technique for the definition of complex XML types for the <Data_Parameters>. By redefining the ISDC "base-dif.xsd" schema, all data-child DIF XML documents are derived. Using the GCMD XML schema, this approach would not be possible because of the definition of XML reference structures.

The extended metadata in the data-child DIF XML documents are parsed by a Perl (an open-source software) script. If the data structure is correct, the extended metadata are stored in product-type-related tables in a relational database. The connection between the data-child DIF XML files and the product type DIF XML document is given by the equality of parts of the <Entry_ID> element in both the product type and the related product metadata documents. Additionally, the content of the <Parent_DIF> element in the data-child DIF XML document refers to the appropriate product-type DIF document. The relation between the schemata, the XML metadata files, and the storage structures is shown in figure 2.

Using the product-type DIF XML structures, it is possible to conduct a thematic content-based search of the different product-type documents as well as provide interoperability with other catalog systems. It is now possible to transform the XML DIF files into ISO 19115-standard documents (International Organization for Standardization, 2003) in order to use Open Geospatial Consortium-compliant Web services such as deegree's (www.deegree.org) Catalogue Service for Web 2.0. Furthermore, achieving harmony with other catalog systems

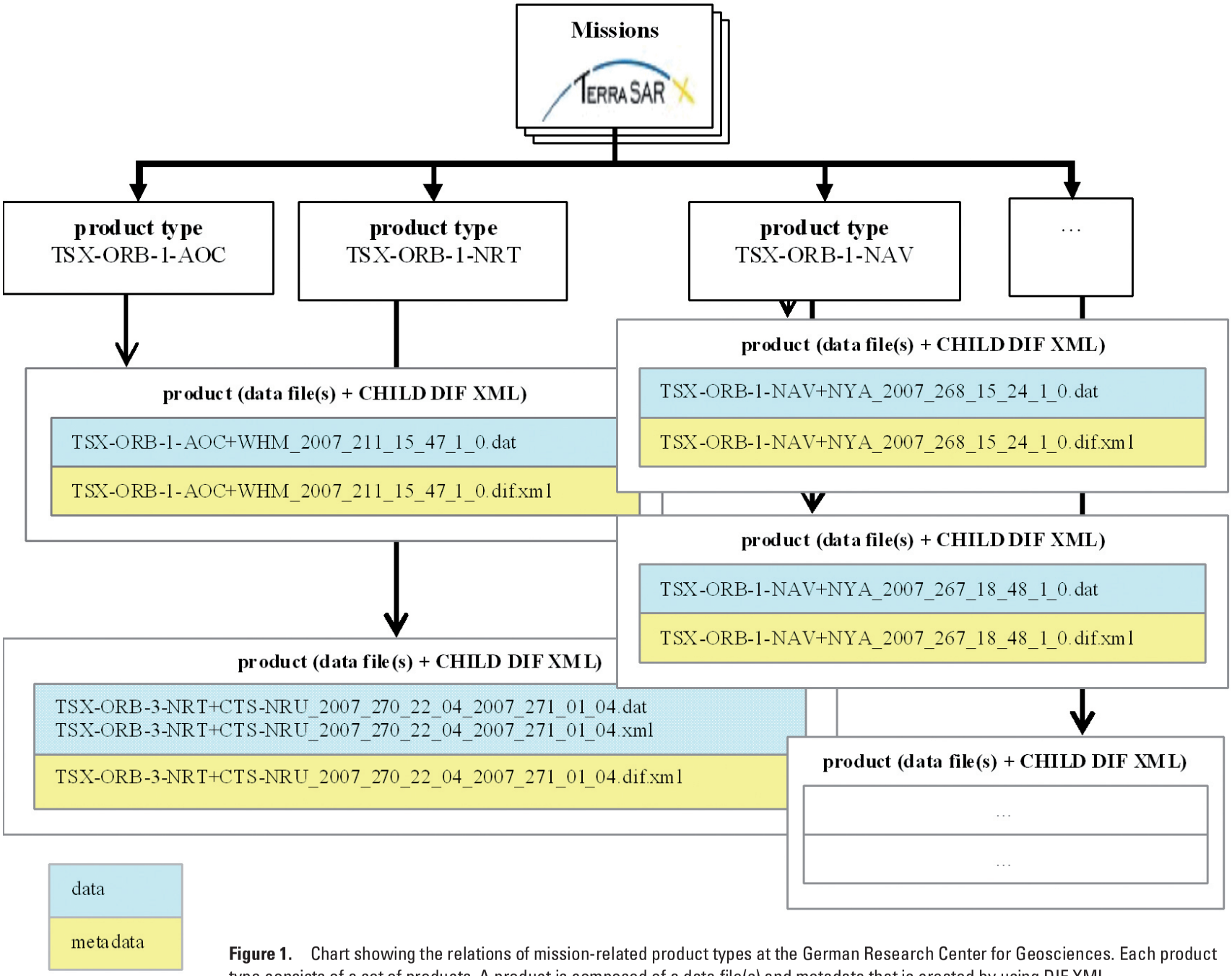

Figure 1. Chart showing the relations of mission-related product types at the German Research Center for Geosciences. Each product type consists of a set of products. A product is composed of a data file(s) and metadata that is created by using DIF XML.

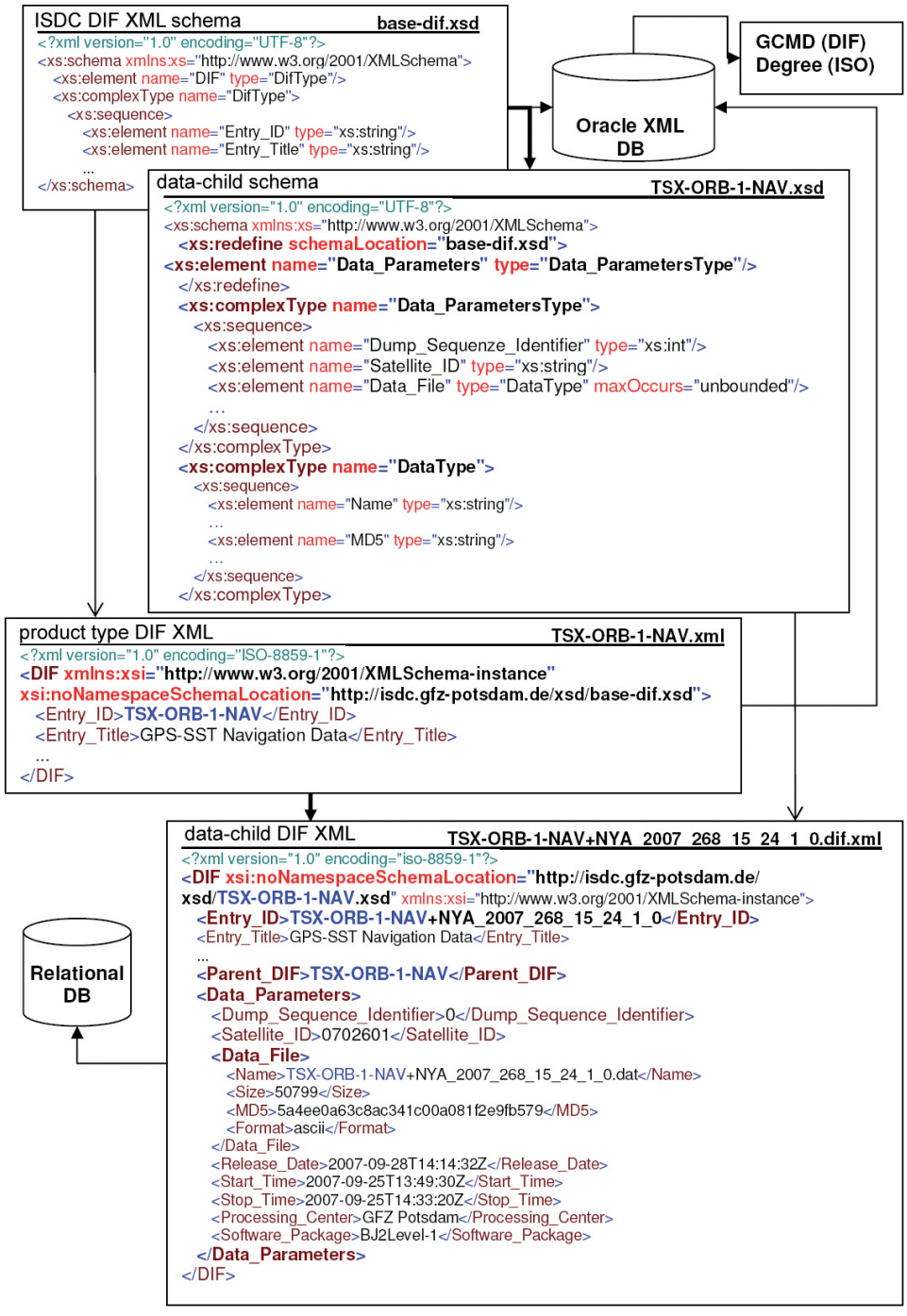

Figure 2. Chart showing the relation between schemata (base and the data-child schema) and the metadata (product-type metadata and child-metadata). The storage mechanisms, such as relational and XML-based databases, also are shown.

is possible by using international standards. The structure of XML easily allows extending the DIF standard in future. Using the parent-child DIF concept, only a small amount of mandatory metadata must be included in both the product type and data-child DIF XML documents.

Reference Cited

International Organization for Standardization, 2003, Geographic information—Metadata: Geneva, Switzerland, International Organization for Standardization, 140 p.