Open Data Interface
Final Report

Peter Wintoft and Lars Eliasson
Swedish Institute of Space Physics
and
Daniel Heynderickx
DH Consultancy

June 17, 2010

Swedish Institute of Space Physics

DH Consultancy

Document status sheet

Version	Date	Comment
1.0	2010-06-17	

Contents

Abstract

The Open Data Interface (ODI) is a software system to store data and metadata. The metadata syntax follows the CDF/ISTP[1] guide lines. The ODI system relies on a MySQL[2] database to store its data and metadata. There are MySQL database drivers for all major computer platforms (Linux, Mac OS X, Unix, Windows) and libraries have been implemented for all major programming languages (e.g. C, C++, IDL, Java, Mathematica, Matlab, Perl, PHP). This means that data in ODI are accessible to a wide variety of software systems.

The ODI system contains: software to ingest data and metadata for a number of specific datasets from plain text files; software to ingest data and metadata from ISTP compliant CDF files; software to export data into plain text files or CDF files; a set of command line tools to administer and explore the database; scripts to automatically download datasets; scripts to automatically ingest datasets; scripts to set up scheduled download/ingest (based on Cron).

The core ODI system has been developed in PHP which should make it platform independent. However, some datasets rely on platform dependent code and the CDF export utility relies on IDL/Dataminer[3]. As the syntax for data originating from plain text files is not standardized custom text parsing code is necessary. Several parsers are included in ODI and new parsers can be added to allow the ingestion of additional datasets. New parsers are automatically identified by the system.

A set of software systems, namely SAAPS[4], SEDAT[5], and SPENVIS[6], have been updated to make use of the ODI system.

ODI was developed by the Swedish Institute of Space Physics and DH Consultancy under the ESA/ESTEC Contract 21964/08/NL/AT. More information can be found at http://www.lund.irf.se/odi/.

1 Applicable and reference documents

ODI/SOW ODI Statement of Work

ODI/SSS ODI Software System Specification

ODI/TN/DB Technical Note on the ODI Database

ODI/TN/SSS TN on SAAPS/SEDAT/SPENVIS adaption for ODI

ODI/AG ODI Administrator Guide

ODI/UG ODI User Guide

[1] http://spdf.gsfc.nasa.gov/sp_use_of_cdf.html
[2] http://www.mysql.com/
[3] http://www.ittvis.com/ProductServices/IDL.aspx
[4] http://www.lund.irf.se/saaps/
[5] http://www.wdc.rl.ac.uk/sedat/
[6] http://www.spenvis.oma.be/

2 Terms, definitions and abbreviated terms

CDF Common Data Format

ODI Open Data interface

PRBEM Panel on Radiation Belt Environment Modeling

SAAPS Satellite Anomaly Analysis and Prediction System

SEDAT Space Environment Data Analysis Tool

SOW Statement of Work for ESTEC/Contract No. 21964/08/NL/AT

SPENVIS Space Environment Information System

SQL Structured Query Language

3 Introduction

The main purpose of the *Open Data Interface* (ODI) is to provide a common backend and database system for three existing systems at ESA, namely SAAPS, SEDAT, and SPENVIS. It shall also provide a generic interface to space environment datasets and allow for the ingestion of new datasets. The data shall be stored in a MySQL database and the ODI system shall be based on software that allows the migration to different platforms, although it is mainly targeted for a Linux platform. The requirements were given in the ODI/SOW and tailored out in the ODI/SSS.

The setup of the ODI system is described in the ODI/AG, while someone interested in only using ODI should turn to the ODI/UG.

ODI was developed by the Swedish Institute of Space Physics (Contractor) and DH Consultancy (Subcontractor) under the ESA/Estec Contract 21964/08/NL/AT.

The ODI system shall also be capable of storing both data and metadata, compliant with the CDF/ISTP/PRBEM standards. However, it shall not be restricted to only importing CDF files, but it shall offer the capability of importing any file types.

The ODI project started in October 2008 and finished in March 2010. The first phase of the project was devoted to development of the system, and for the final six months the project entered the maintenance phase. The project's Gantt chart is shown in Figure 1.

The project was divided into 7 work packages:

WP100 Core database system

WP200 Population of the database

WP300 SAAPS interface

WP400 SEDAT interface

WP500 SPENVIS interface

WP600 Maintenance and support

WP700 Administration

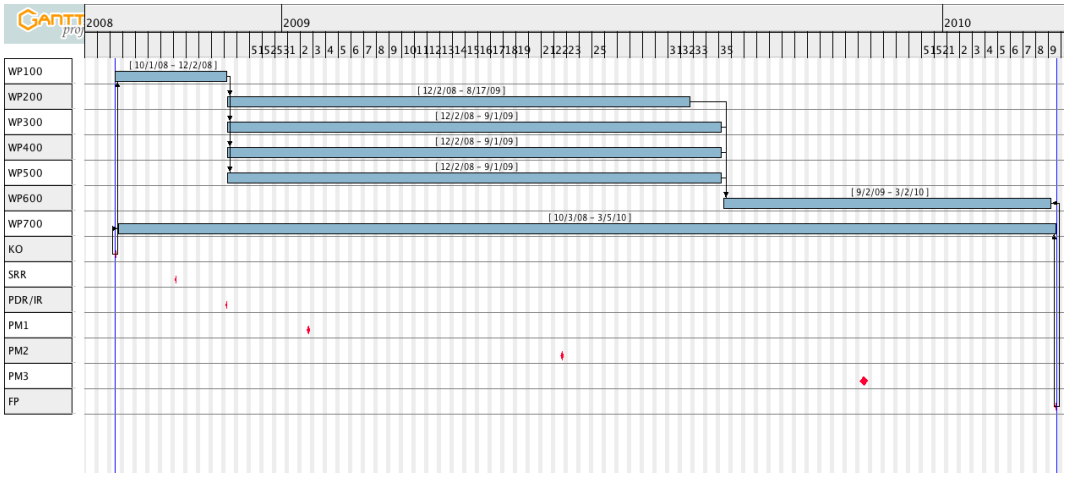

Figure 1: **The project time lines.**

4 WP100: Core database system

After a detailed analysis of the datasets that should be included in ODI it was decided to follow the metadata syntax given by CDF. CDF provides a complete syntax for the description of metadata and many files to be ingested are CDFs. However, the structure of the CDF is also very flexible which adds to the complexity of the problem. The ISTP guidelines provides a set of metadata attributes to somewhat narrow this complexity, and on top of that the PRBEM guidelines provides a set of predefined variables for radiation belt data.

The analysis led to the creation of seven MySQL tables to hold metadata (Figure 2). These tables define the core of the ODI system. When a dataset is added the metadata will be added to the core tables. The details of these tables are described in the ODI/TN/DB.

The actual data for each dataset are stored in separate tables that are created each time a new dataset is added. A dataset is defined as the data and metadata coming from one well defined entity like a spacecraft instrument, index, or equivalent. So in contrast to CDF files, where data from one instrument can be divided over multiple files, all data for one dataset goes into one MySQL table (ODI/TN/DB).

MySQL tables do not support the storage of multidimensional data. In ODI this was solved by expanding multidimensional data over multiple columns in the MySQL data

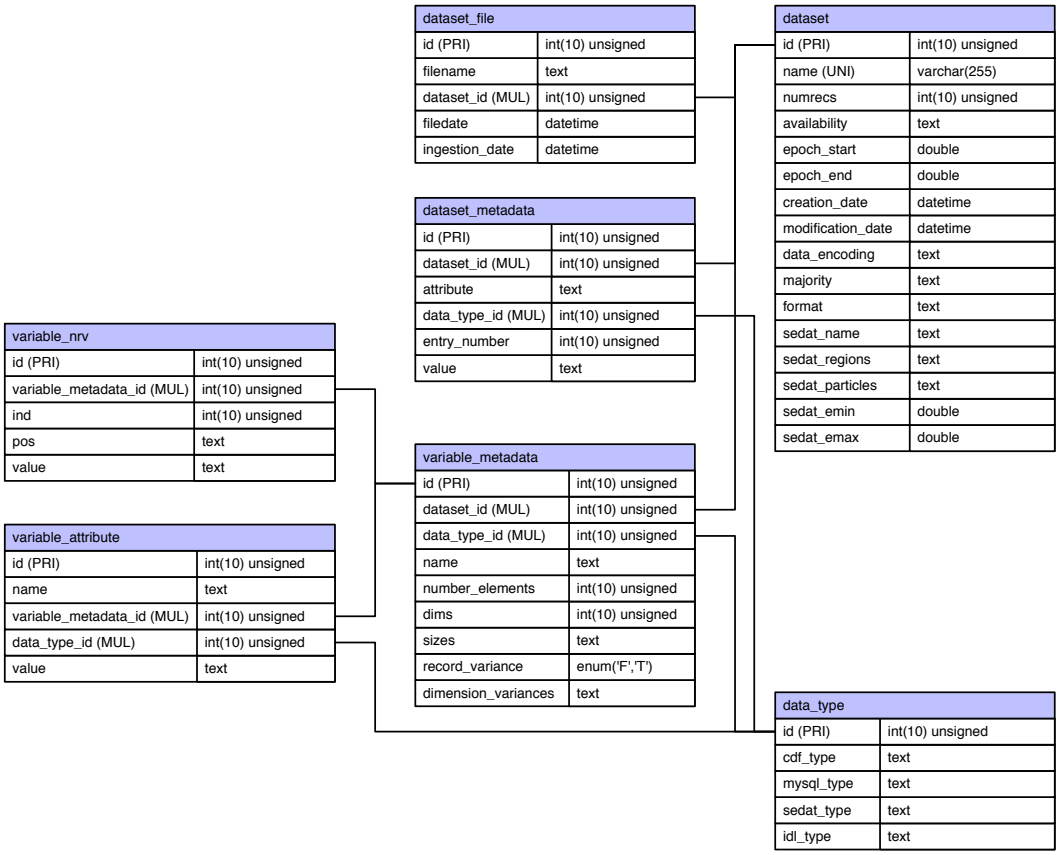

Figure 2: **The ODI core database tables.**

table. The ODI system identifies multidimensional data from the metadata and performs this expansion automatically when a new dataset table is created (ODI/TN/DB).

A set of scripts (or programs) written in PHP and SQL provides the tools to setup the ODI database, downloading data, ingesting data, exporting data, and to perform automatic download and ingestion. In total there are about 3000 lines of PHP code and 200 lines of SQL code.

5 WP200: Population of the database

The original data comes from many different sources and are provided as CDF, plain text, and HTML files. The ODI system parses the files and stores them into the ODI database.

The metadata for each dataset is defined in a text file that follows the syntax for CDF skeleton files. To create a new dataset in the ODI database this skeleton file must be provided. This can be accomplished by creating the skeleton file from an existing CDF file or creating a skeleton file from scratch.

For CDF files the format is well defined and the ODI system contains generic routines to parse CDFs and store the metadata and data into the ODI system. However, many of the original CDF files did not follow the ISTP standard and the skeleton files with the metadata needed considerable editing, something that was not anticipated at the project start.

For plain text files and HTML files there is no standard syntax, so for each dataset a dedicated parser had to be written. However, the ODI system has been developed so that the custom parser function follows a standard interface and only one file needs to be added and edited for each new parser. The custom parser is automatically loaded by the system when it is needed. This means that it is straightforward to add new datasets based on non-standard formats.

The ODI database currently contains about 100 datasets totalling more than 40 GB. The parser functions consist of about 2000 lines of PHP code while the skeleton files for the datasets contain in total about 100 000 lines of text.

6 WP300: SAAPS interface

SAAPS is written in Java and has been modified to connect to the ODI system for the data retrieval. A few classes have been modified that handles the connection to ODI, while the rest of the SAAPS functionality is unmodified. SAAPS now provides the same functionality but accesses data from the ODI system. An example is given in Figure 3.

7 WP400: SEDAT interface

SEDAT uses the STPDF as an engine to manage and access its datasets. STPDF keeps track of the available datasets through a series of control and descriptive files (tin and chp

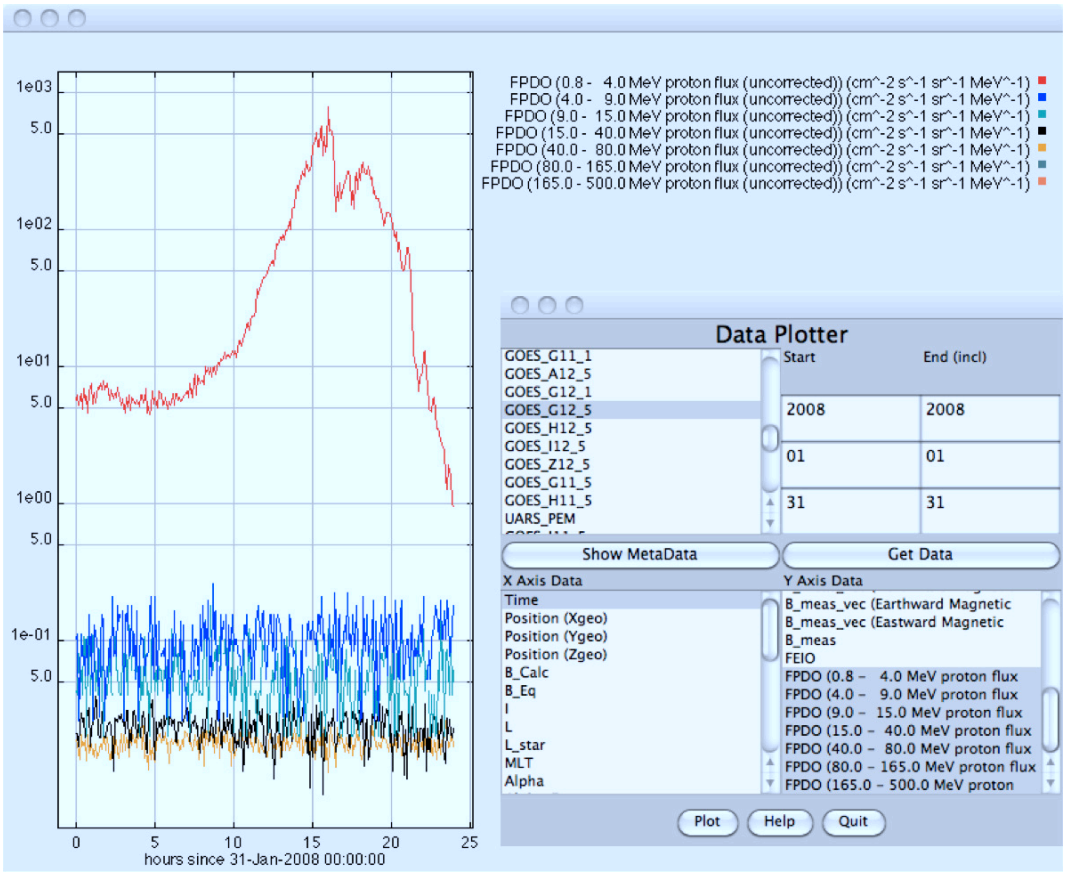

Figure 3: **Example of one SAAPS tool.**

files). SEDAT distinguished between system datasets and user datasets: the former are ingested and controlled by the system administrator, the latter are created and managed by individual users.

In the ODI project, the STPDF backend to the system datasets has been replaced by a MySQL backend. Handling of the user datasets has not been modified, which means that the upgraded version of SEDAT uses ODI as well as STPDF.

The modifications made to the SEDAT system only affect the backend for system datasets, and are transparent to users using the SEDAT GUI. Thus, users will not see any changes in the operation of SEDAT. An example is given in Figure 4.

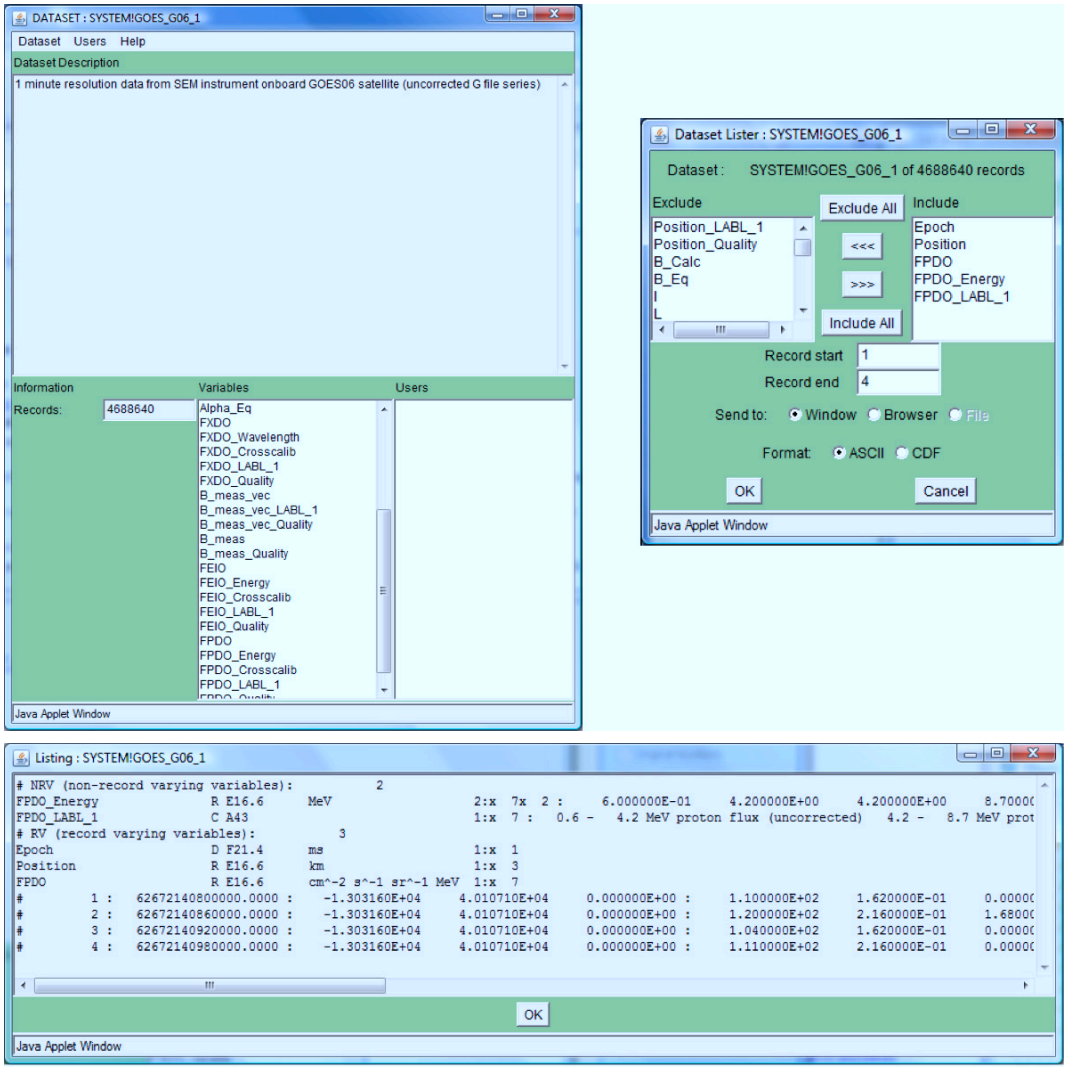

Figure 4: **Example of the SEDAT interface.**

8 WP500: SPENVIS interface

SPENVIS uses a custom built set of IDL tools to:

- create cdf files from downloaded datasets;
- produce HTML templates for each dataset;
- query the cdf files;
- plot the results from the queries.

In the ODI project, this data backend has been replaced by a MySQL backend. To this effect, new HTML templates and php scripts have been written that query ODI and display dataset information without the need for dataset specific templates. An example is given in Figure 5.

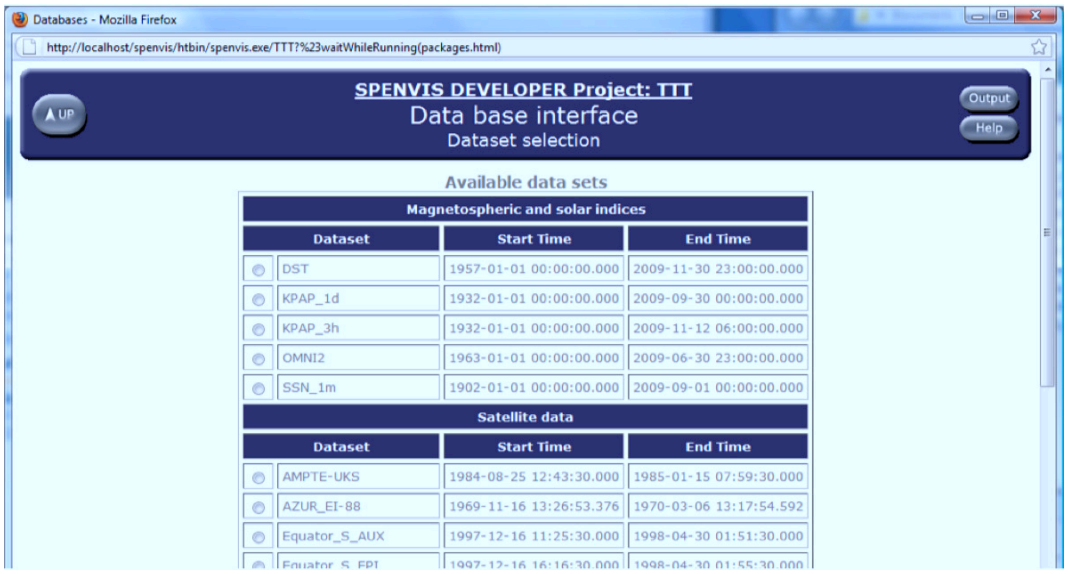

Figure 5: **Example of the SPENVIS interface.**

9 WP600: Maintenance and support

During a six month period the ODI system has been regularly monitored and problems have been identified by Estec, IRF, and DHC. These problems have been dealt with and solved.

10 Summary

The ODI system has been developed to store space physics data into a MySQL database. Almost 100 datasets have been ingested with a total size of more than 40 GB. The SAAPS, SEDAT, and SPENVIS systems have been updated so that they connect to the ODI system to retrieve data. The system is currently operational on a server at Estec and data is continuously added.