*Astronomical Data Analysis Software and Systems IV*
*ASP Conference Series, Vol. 77, 1995*
*R. A. Shaw, H. E. Payne, and J. J. E. Hayes, eds.*

# Development of an ADS Data Dictionary Standard

C. S. Grant, A. Accomazzi, G. Eichhorn, M. J. Kurtz, and S. S. Murray

*Smithsonian Astrophysical Observatory, 60 Garden St., Cambridge, MA 02138*

**Abstract.** We present a proposed standard for data dictionaries associated with catalogs being accessed through the Astrophysics Data System (ADS) Mosaic Catalog Service. Each catalog made available for searching must have a data dictionary, which consists of a set of specified keywords describing format, content and location of data. The data dictionary provides the ability to describe catalogued data accurately, and can be used to facilitate examining data from different sources in a common environment.

## 1. Introduction

The ADS Mosaic Catalog Service[1] currently provides access to about 150 data catalogs located at various nodes across the country. This service was developed to provide a simple means of making on-line catalogs available for searching. Sites with catalogs already in a relational database need only install a server which we provide and create a data dictionary with tools we will provide.

In the current implementation of the service, users may search on any field contained within the catalog, either by typing in a Standard Query Language (SQL) request, or by filling in a table template (QBT). In order to provide the capability of searching catalogs by position, our software needs to be able to identify what kinds of positions are contained in which columns. Because we do not maintain the data, we have no control over what coordinate types and units are in the catalogs. We therefore propose a standard method of identifying column contents with an associated data dictionary file. The data dictionary can also be extended beyond positional column identification and used to key on any astronomically-interesting field (such as magnitude or class).

## 2. Data Dictionaries

We propose having one file per database which lists information about each table in the database (**Database Data Dictionary**). In addition, we propose having one file per table which lists information about each column in the table (**Table Data Dictionary**). Files will be in FITS-like format, with *keyword=value*

[1] http://adscat.harvard.edu/catalog_service.html

Table 1. Database Data Dictionary

| Keyword | Description |
|---|---|
| **DB_NODE** | Node name |
| **NAMEnnn** | Database name |
| **TBLEnnn** | Table name |
| **DESCnnn** | Description of table |
| **COLSnnn** | Number of columns in table |
| **ROWSnnn** | Number of rows in table |
| **SUBJnnn** | Colon separated list of subjects (for lists) |
| **KWDnnn** | Colon separated list of keywords (for searching) |
| **URLnnn** | URL for description file |

pairs. We will provide software which creates these files from the catalog documentation file. Nodes will maintain the files so that updates to the databases can be handled without requiring ADS intervention.

Both the descriptions of the catalogs and the keywords assigned to the catalogs are WAIS-indexed. This allows users to search the catalog content for information about the catalogs (such as which catalogs contain redshift).

### 2.1. Database Data Dictionary

There is one **Database Data Dictionary** file per database at a node. This file is used by the software to construct lists of tables for selection ("by node" or "by subject"). The field specifications are detailed in Table 1. The subjects are used for grouping similar catalogs together. The keywords are used for WAIS searches to enable users to find a catalog about a particular subject. Both of these are to be taken from word sets already in use by the ADS project. The software provided to help generate the data dictionary files will facilitate the selection of appropriate subjects and keywords.

### 2.2. Table Data Dictionary

There is one **Table Data Dictionary** per table at a node. The counter, nnn, specifies the column number in the table. Additional keywords (such as NAXIS1 and NAXIS2) are added when a FITS table is written out containing results of a query. The field specifications are detailed in Table 2.

The keywords which begin with "DB_" correspond to keywords in the associated Database Data Dictionary. These are essentially a header to the data dictionary, giving information about the table (as opposed to information about the columns). TTYPE and TFORM are the only two required keywords for each column. Unused keywords may be omitted.

## 3. Data and Auxiliary Flags

There are two keywords in the Table Data Dictionary which indicate to the server software that the corresponding column contains special attributes. The data flag, **TDFLGnnn**, describes the type of attribute (such as position or image); the auxiliary flag, **TAFLGnnn**, describes additional information as-

Table 2. Table Data Dictionary

| Keyword | Description |
|---|---|
| **DB_NODE** | Node name |
| **DB_NAME** | Database name |
| **DB_TBLE** | Table name |
| **DB_DESC** | Description of table |
| **DB_COLS** | Total number of columns in table |
| **DB_ROWS** | Total number of rows in table |
| **DB_SUBJ** | Colon separated list of subjects (for lists) |
| **DB_KWD** | Colon separated list of keywords (for searching ) |
| **DB_URL** | URL for description file |
| **TTYPEnnn** | Column name |
| **TFORMnnn** | Format of data in database |
| **TUNITnnn** | IAU/NSSDC units code |
| **TDESCnnn** | Description of column |
| **TEXMPnnn** | Example of data |
| **TFMATnnn** | Desired output format |
| **TWDTHnnn** | Width of output column |
| **TJUSTnnn** | Justification of output column |
| **TDFLGnnn** | Data flag |
| **TAFLGnnn** | Auxiliary data flag |
| **TACCYnnn** | Accuracy of data in database |
| **TLMINnnn** | Lower limit of valid data values |
| **TLMAXnnn** | Upper limit of valid data values |
| **TDMINnnn** | Minimum value of column in database |
| **TDMAXnnn** | Maximum value of column in database |
| **TSIZEnnn** | Size in bytes of data referenced in columns |

sociated with the data flag (such as coordinate specifications or a directory pathname). Table 3 lists examples of data flags.

## 3.1. Positional Flags

Positional data flags indicate to the server software that the corresponding column contains coordinates. The primary set of positions to be used in searching (determined by the data providers) is flagged with a data flag “p”. There can be only one set of primary positions, but there can be multiple secondary position data flags. The positional auxiliary flags (see Table 4 for examples) describe what kind of coordinates are represented in a colon-separated list, including: which coordinate ($x$ or $y$), the coordinate units (DD, RAD, sex1, sex2, etc.), the coordinate name where appropriate (some sexagesimal representations), the coordinate system (EQ, GAL, or ECL), and the coordinate epoch where appropriate (B1950 or J2000).

## 3.2. Non-positional Flags

The remaining data flags give tremendous flexibility in the search capabilities. For example, by flagging which column contains the name of the associated image, data archives can be built to retrieve lists of images available in a given catalog. Likewise, by flagging which column is a magnitude column, plot routines can be automated to size the plot symbols based on magnitude. Other data flags can be added by the nodes as desired.

Table 3. Data Flags

| TDFLG | Description |
|---|---|
| **p** | positional primary search field |
| **s** | positional secondary search field (s1, s2, etc.) |
| **t** | path name of ASCII text file |
| **i** | path name of binary data file |
| **m** | astronomical magnitude |
| **e** | ellipticity |
| **z** | size in arcmin |
| **o** | orientation in degrees |
| **c** | object classification |

Table 4. Positional Auxiliary Flags

| TAFLG | Example |
|---|---|
| x:sex1:EQ:B1950 | 122430.0 |
| y:sex1:EQ:B1950 | +261630.0 |
| x:sex2:RAH:EQ:J2000 | 12 |
| x:sex2:RAM:EQ:J2000 | 27 |
| x:sex2:RAS:EQ:J2000 | 00.1 |
| y:sex2:DECD:EQ:J2000 | +25 |
| y:sex2:DECM:EQ:J2000 | 59 |
| y:sex2:DECS:EQ:J2000 | 54.2 |
| x:DD::GA: | 223.237 |
| y:DD::GA: | +84.421 |

## 4. Summary

The ADS Mosaic Catalog Service offers data sites the opportunity to make their data available from a centralized location. We provide the software to search catalogs individually, and once the data dictionaries are in place, we will implement positional searches as well as multiple catalog searches. Output from the Catalog Service may be returned in a variety of formats (including ASCII, FITS, and ADS tables) giving users the necessary flexibility for manipulation of their results.

Putting a catalog on-line requires only four steps: (1) running an httpd, (2) installing SQL software[2] (we will assist), (3) writing catalog documentation, and (4) creating data dictionary (we will assist). The ADS provides WWW access to catalogs while still allowing the nodes to maintain ownership of all associated files for updating and maintenance. We offer a homogeneous Mosaic-based query mechanism which is already in place and ready to be expanded. For more information, please contact *ads@cfa.harvard.edu.*

**Acknowledgments.** This project is funded by the NASA Astrophysics program under grant NCCW–0024.

[2]http://adsdoc.harvard.edu/install.html