# Table of Contents

Revision History .................................................................................................................. 3
Overview ............................................................................................................................... 4
Requirement 1: Required Columns .......................................................................................... 5
Requirement 2: Data Organization ........................................................................................... 6
Requirement 3: Must be accessible from the web ..................................................................... 6
Requirement 4: Must be updateable by suitably authorized clients .......................................... 7
Appendix 1 (Other Implementation Options) .......................................................................... 8
Appendix 2 (OBIEE Hover Text) ........................................................................................... 8
Appendix 3 (Best Practices/Guidelines for Business Definitions) ........................................... 8
Appendix 4 (Columns Considered but Not Included) ............................................................... 9
<table>
<thead>
<tr>
<th>Name</th>
<th>Date</th>
<th>Changes Made</th>
<th>Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brian Simms</td>
<td>4/6/12</td>
<td>Initial draft</td>
<td>0.1.0</td>
</tr>
<tr>
<td>Brian Simms</td>
<td>4/23/12</td>
<td>First Revision (Major)</td>
<td>0.1.1</td>
</tr>
<tr>
<td>Brian Simms</td>
<td>4/25/12</td>
<td>Second Revision</td>
<td>0.1.2</td>
</tr>
<tr>
<td>Brian Simms</td>
<td>4/27/12</td>
<td>Draft sent to Steering Team</td>
<td>1.0.0</td>
</tr>
<tr>
<td>Brian Simms</td>
<td>5/14/12</td>
<td>Added new required column ‘Derivation Logic’</td>
<td>1.0.0</td>
</tr>
<tr>
<td>Brian Simms</td>
<td>5/29/12</td>
<td>Added a ‘business definition’ for each of the columns in the required columns section and gave them a description. Removed the column ‘Derivation Logic’. This info will go into the ‘Additional Notes’ column instead of separate column. Also updated the wording in requirement 2 to provide a little more description at the beginning. Also modified the links in requirement 3 to link to IAPROD instead of IATEST.</td>
<td></td>
</tr>
</tbody>
</table>
Overview

A data dictionary, or metadata repository, as defined in the IBM Dictionary of Computing, is a "centralized repository of information about data such as meaning, relationships to other data, origin, usage, and format."

The primary goal of our data dictionary is to allow users to better understand the data in the data warehouse.

The Data Dictionary should be:

- Accessible – For both reading and updating.
- Reliable – The information in the data dictionary must be kept up to date and managed by the appropriate data stewards.
- Robust – The richer the information, the more useful. Must also be searchable.
- Central – One and only one data dictionary.

The main requirements in the remainder of this document intend to describe:

- What information we will store about the data elements in the data dictionary.
- How the information will be stored.
- How it will be accessed.
- How it will be updated.
- Other considerations.

The process of building out information in the data dictionary will last for as long as the data warehouse is in existence. It is expected that as clients use the system, especially Subject Matter Experts/Data Stewards, they will update the Data Dictionary as needed with new and useful information.
## Requirement 1: Required Columns

<table>
<thead>
<tr>
<th>#</th>
<th>Column Name</th>
<th>Business Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Subject Area</td>
<td>The OBIEE subject area</td>
</tr>
<tr>
<td>1.2</td>
<td>Presentation Folder Name</td>
<td>The top level folder in a subject area within OBIEE</td>
</tr>
<tr>
<td>1.3</td>
<td>Presentation Subfolder Name</td>
<td>The subfolder (if any) within a presentation folder</td>
</tr>
<tr>
<td>1.4</td>
<td>Presentation Column Name</td>
<td>The data element name</td>
</tr>
<tr>
<td>1.5</td>
<td>Business Definition</td>
<td>The definition of the data element supplied by the subject matter experts from various offices around campus.</td>
</tr>
<tr>
<td>1.6</td>
<td>Additional Notes</td>
<td>Field to allow for entry of a variety of useful notes about the data element. This field can and should be used for:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1. <strong>Usage Notes.</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. <strong>Things to watch out for.</strong> Example: ‘Do not confuse Student Campus Group with Course Campus Group.’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. <strong>Operational Definition.</strong> Example: ‘Total Instructional fee = Instructional Fee + OOS Surcharge + Comprehensive Fee.’</td>
</tr>
<tr>
<td>1.7</td>
<td>DW Table Name</td>
<td>Name of the Data Warehouse table that stores the data.</td>
</tr>
<tr>
<td>1.8</td>
<td>DW Column Name</td>
<td>Name of the Data Warehouse column that stores the data.</td>
</tr>
<tr>
<td>1.9</td>
<td>Data Type</td>
<td>The oracle data type. Number, Varchar2, Date, etc...</td>
</tr>
<tr>
<td>1.10</td>
<td>Source System</td>
<td>The system from which the data originated. Most often, this will be Banner. But other sources of data are possible.</td>
</tr>
<tr>
<td>1.11</td>
<td>Source Table</td>
<td>The table that held the data in the source system</td>
</tr>
<tr>
<td>1.12</td>
<td>Source Column</td>
<td>The column that held the data in the source system</td>
</tr>
<tr>
<td>1.13</td>
<td>Cube Section</td>
<td>For conversion purposes only. If the subject area existed in Microsoft Analysis Services Cubes then this information could be useful to determine the folder structure within an OBIEE Subject Area.</td>
</tr>
<tr>
<td>1.14</td>
<td>Cube Section Sub Group</td>
<td>For conversion purposes only. If the subject area existed in Microsoft Analysis Services Cubes then this information could be useful to determine the folder structure within an OBIEE Subject Area.</td>
</tr>
<tr>
<td>1.15</td>
<td>Cube Field</td>
<td>For conversion purposes only. If the subject area existed in Microsoft Analysis Services Cubes then this information could be useful to determine the folder structure within an OBIEE Subject Area.</td>
</tr>
<tr>
<td>1.16</td>
<td>Last Modified Date</td>
<td>System Maintained field. The Date the information was last updated.</td>
</tr>
<tr>
<td>1.17</td>
<td>Last Modified By</td>
<td>System Maintained field. Unique id of the person that last updated the information.</td>
</tr>
</tbody>
</table>

While this is the list of columns that will be available to store data, all columns will not necessarily be displayed on all reports.

Note: some of the columns listed do not currently contain any data. The ‘Additional Notes’ field is a good example. This field is intended to be updated by clients over time as the system is used.
Requirement 2: Data Organization
As the IA project expands, there will be instances where data elements will share a common name but will be defined differently based on their role in the particular Subject Area. As such, we need to allow for a different definition for the same data element in different Subject Areas.

To accomplish this, one row will be stored per data element per subject area. This will allow for a different definition per subject area. However, initially, there is no requirement for the definition to be different. This allows data elements and their definitions to be entered once and then duplicated for each subject area.

Consider the following example: Account Code

Base definition: The account code will represent an asset, liability, control account, fund balance, revenue, or expenditure transaction.

Bursar Course Revenue Definition: Debit Account Code – Account debited by the bursar transaction.

Payroll Definition: Account Code from which the payroll was paid.

Human Capital Job Definition: Account Code from which the job is budgeted.

Human Capital Position Definition: Account Code from which the position is budgeted.

Having one row per data element per subject area allows for the client to search the data dictionary for ‘Account Code’ and see all the subject areas where it is being used.

One significant drawback of this method is the potential for needing to make a change in multiple places if the data definition changes. One way to get around this drawback would be to put a base definition in the ‘Data Definition’ field and keep that the same for each subject area. Then in the ‘Additional Notes’ field specify how it is used specific to the subject area.

Requirement 3: Must be accessible from the web
Primary access to the data dictionary will be from within OBIEE itself. Several reports will be created to access the data dictionary:

1) Data Dictionary Dashboard Report
   This report will contain only the data elements on each dashboard. A link to this report should be included somewhere on the dashboard.
One of these types of reports should be created for each set of dashboards created. This will allow the user to get a list of data definitions for relevant data elements.

Link to this report in IAPROD for the VP Dashboards: Dashboard Data Dictionary

2) Data Dictionary Report
   This report will be the main report used for general purpose querying of the data dictionary. It will include all data elements in the data dictionary.

   A prompt will be added to the top of the report to allow for searching.

   Link to this report in IATEST Data Dictionary
   (The link above will be updated in the near future after the 4.2 release is promoted to IATest and then IAProd.)

3) APEX Data Dictionary Application – This application was created primarily to satisfy the requirement for making updates, however, it can be used to view the data dictionary as well. This application allows for searching.

   The link to the APEX application in IAPROD is: APEX Data Dictionary App

**Requirement 4: Must be updateable by suitably authorized clients.**
In order to facilitate an accurate and up to date Data Dictionary, it is important that it can be updated quickly and easily by the subject matter experts.

   a. Data stewards (AKA Subject Matter Experts) should be able to update the areas where they are the experts.
   b. In addition to the Data Stewards, Authors and/or Publishers should be able to update the ‘Additional Notes’ field for immediately providing information about using the data element.
   c. Changes made by people other than the Data Stewards should result in the creation of an ‘Edit’ report that is sent to the Data Steward.
   d. Needs to include an archiving ability to save who made the last update and on what date.
   e. We need ‘Best Practices’ for data definition entry.

The DBA’s have developed an Oracle APEX application to satisfy this requirement. The application allows the appropriately authorized clients (data stewards) to make updates and is searchable. The updates will be made to a ‘staging’ table and copied over to the main data dictionary table. The copy process will run nightly with on request updates as requested. The solution does not currently satisfy item b. or c. from the above list. Only the data stewards currently have the ability to update the information in the data dictionary.

The link to the APEX application is: APEX Data Dictionary App
Appendix 1 (Other Implementation Options)
There are a couple of options for implementing a data dictionary. Most of the systems I have seen have all been built in house. This is currently our plan as well. However I wanted to include a couple of other options for considerations:

A) Purchase a system. There is at least one system out there specifically built for data dictionary management. It is from iData inc. and it is called the ‘Data Cookbook’. It is a cloud application. It can be found here: http://www.datacookbook.com/

B) Wiki. Thinking of the Wikipedia web site, it does seem well suited to handle the requirements. But we are building a dictionary here, not an encyclopedia.

Appendix 2 (OBIEE Hover Text)
The Revenue Metadata Team is proposing to not require the hover text until we have a way of reading directly from the Oracle table.

In order to get the data definitions in the hover text, IT would have to manually update each data element in the RPD. While this hover text is convenient, both the BI team and the Metadata team have indicated that there does not seem to be all that much value. The maintenance required to keep the RPD up to date is a serious dilemma. Changes for data definitions would have to be coordinated with other RPD changes and the promotion schedule, since other changes are going to be happening to the RPD. As a result, it could be weeks before updated data definitions make it into the hover text.

Additionally, only 1 field – the data definition – is available from the hover text.

Finally, the only place to view the hover text is when creating a new report using the ‘Answers’ tool. The people using the ‘Answers’ tool is likely to be the same people who entered the data definitions to begin with.

Since little value is added, it is proposed that we don’t spend the time attempting to keep the hover text updated until we can find a way to read directly from the data base tables.

The alternative to the ‘Hover Text’ is to use the Data Dictionary report(s). This report displays all the information that would normally be in the Hover text plus some additional fields from the Data Dictionary.

Appendix 3 (Best Practices/Guidelines for Business Definitions)
- The ‘Business Definition’ column should not include any Banner specific table names or other type of unfamiliar Banner terms. Definitions need to be understandable by non-Banner users. (See the ‘Source System’, ‘Source Table’ and ‘Source Column’ columns to find Banner or other system specific table/column names.)
- The definition should use only well-known abbreviations.
Appendix 4 (Columns Considered but Not Included)
The following list of columns contains data elements suggested for possible inclusion in the Data Dictionary. Some of these were mentioned at 2012 HEDW Conference. They are not being implemented at this time for various reasons, but they could be considered for possible future enhancements.

1.1 Data Steward/Data Steward contact info.
1.2 Status - Y/N field indicating that the information has been verified by the data steward.
1.3 Related objects/See also – The ability to include links in the Data Dictionary would be nice.
1.4 Range of Values – Keeping this information up to date is a concern.
1.5 Default Values – ‘No Data Found’ and ‘None’ are often default values and an explanation of what each means could be useful.