# THE OPEN DATA WAREHOUSE MODEL (ODWM)

Cliff Brandon
January 2009

# AN INTRODUCTION TO PROJECT DATA WAREHOUSES
By: Cliff Brandon 1/20/2009

## TABLE OF CONTENTS

## COPYRIGHTS and CAVEATS

## EXECUTIVE SUMMARY

The ODWM provides guidelines for constructing 'Open Source' like data warehouses. It is meant to be a short, concise design document directed at those who want to construct these types of data warehouses or provide related tools and services.

## INTRODUCTION

The ODWM provides guidelines for constructing 'Open Source' like data warehouses. It is meant to be a short, concise design document directed at those who want to construct these types of data warehouses or provide related tools and services.

## AUDIENCE

This document is directed at those who want to build 'Open' style data warehouses and those who want to provide related tools and services. Note that even though the data warehouses may follow the open source model, the tools and services are not required to also follow or have the 'Open Source' model.

## CURRENT STATE OF THE TECHNOLOGY

As of the initial writing of this document I am not aware of any literature concerning the design of 'Open Source' types of data warehouses though articles have been written advocating the need for it. I'm sure this will change. There is a considerable, mature body of publically available literature on data warehouses, business intelligence, data modeling, and related topics. This document does not significantly build on that body of knowledge but deals primarily with a change in delivery mechanism for data warehouses.

## THE OPEN MODEL

The ODWM does not an attempt to supplant or change the 'Open Source' model. Just the opposite, it builds, and relies heavily on its concepts.

3

## THE FOUNDATION – THE ODWM INSTANCE

An instance of the ODWM will be referred to here as a 'Foundation'.

The purpose of a Foundation is to be available to as broad an audience as possible in a specific practice area. A Foundation must have attributes that make it easy for developers, implementers, and vendors (all stakeholders) to use. It is also critical that it be useful, scalable, and maintainable by the data owner. What follows is a list of Foundation attributes followed by a short description of each attribute.

The Foundation:

- Is a collection of practice specific tables arranged for data warehouse operations,
- Has a limited number tables that are common to all Foundation instances (like ownership/licensing and information source),
- Is constructed for a specific practice area,
- Is designed by individuals with deep practice area knowledge and who are not contractually prevented from producing this type of product,
- Is platform neutral (OS and DB),
- Is source and destination neutral,
- Is scalable,
- Can be used as a stand-alone data warehouse, data mart, or in a federated model.
- Has specific boundaries,
- Has a mechanism for delivering the Foundation to its audience,
- Has a set of common components and processes (data dictionary, stated table and field naming conventions, change management process, etc.)

### Foundation Components

A Foundation is a collection of tables, specific to a practice area that can be arranged in a manner to optimize report generation and analysis. Consider mapping the most atomic data available. Consider using data modeling tools to maintain and create the foundation for various platforms.

### Foundation Common Tables

There are a small number of tables that would be common in this type of data warehouse. Examples include: ownership and data source

information. An ownership or 'copyright' table containing the ownership lineage is an example of a table that is peculiar to the ODWM model. The 'data source table' is needed to deal with the potential for multiple sources and is common to data warehouses inside and outside the ODWM.

### Foundation 'Practice Area'

It is important for data owners, developers, and implementers to understand the scope of the Foundation.

### Foundation Design

It is important that the individual(s) engaged in the design of the foundation have intimate knowledge of the practice area they are designing for. It is also important that they understand data warehouse construction and that they are not contractually prevented from constructing and delivering a Foundation.

### Foundation is Operating System and Database Neutral

Broad use is critical to a Foundation's success. The system should operate on a broad range of operating systems and databases. Consider systems that have broad use (OS and Database), are SQL 92 compliant, and have commonly available ODBC and OLE connectors. Consider avoiding development to proprietary systems.

### Foundation is Source and Destination Neutral

There may be many different sources for a Foundation and many different destinations. Design for the practice area. Let the ETL and report designers figure out the content in and out of the container.

### Foundation is Scalable

Stakeholders must be able to build on the Foundation without changing the Foundation.

### Foundation Distribution

A Foundation can be used as a stand-alone system, in a federated system, or as a mart.

### Foundation has a Delivery Mechanism

There must be a means to deliver the Foundation (including documentation) to users, implementers, suppliers, etc. Consider using a data modeling application to deliver content for various platforms from a single source.

### Foundation has Boundaries

There must be a specific, written description of the Foundation boundaries. Developers will need to know the model boundaries to create effective tools.

### Foundation has Common Components and Processes

Naming Conventions: This document does not seek to impose any particular naming convention but indicates that there should be one and that it be conveyed to the stakeholders. Consider using naming appropriate to and easily understood by practice area users.
Data Dictionary: Table and field definitions.
Change Management Process: The foundation is a stable platform that allows developers, implementers, and users place and use content consistently. That being said, it is critical to develop a mechanism for change that allows the stakeholders to be informed of the changes and delivery schedules.

## FOUNDATION FUNCTIONAL COMPONENTS

The foundation model is designed to deliver core data to a content owner. Other components (data warehouses, data marts, OLAP cubes, metadata constructs, report repositories, etc.) can then be added to and use the content of the foundation.

A simplified model of a data warehouse appears in figure 1 below. Content from one or more sources is processed and moved into the data warehouse on a periodic basis. Reports and analytics can then be generated from the data warehouse content.

Figure 1. Simplified model of a data warehouse

A Foundation differs only in that its structure does not change over time and that it contains the most granular data available for the practice. This approach then allows organizations to extend the model to meet their organization's specific need as depicted in figure 2.
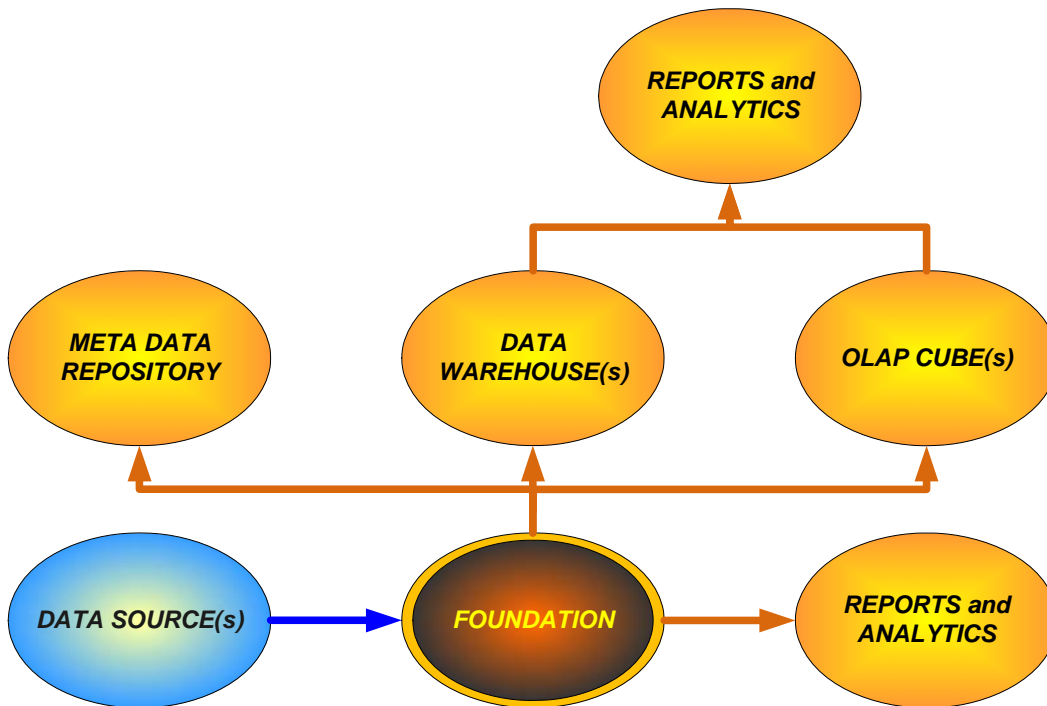


Figure 2. Simplified model of a Foundation style data warehouse

This is not a new concept. What is different is that the Foundation is consistent for a given practice area and follows an 'Open' style model. This allows developers to construct and sell ETL and reporting tools. With such a model it is possible to drop a system into an organization and be up and running with basic reporting in weeks not months or years. Organizations can then add and change other services as they see fit over any time period they desire.