Data Management & Strategy Office

Describing *the* Data Catalog

TALKING POINTS FOR LEADERS

HAPPY DATA™

Help your team understand the concept of a data catalog and what it means for them. Discuss the data catalog in conjunction with data virtualization and data governance so your team has a clear picture of how everything works together.

The Opportunity

If you have ever searched the firm for a data set, it may have gone something like this: you ask a person you know in the department you think would have the data you're looking for. This person in the department you ask has part of the data, but not everything you need. But they point you in the direction of someone else in a different department that might have the rest of what you need. This time, you're lucky and you get what you need from these two departments. Unfortunately for you, each department stored and labeled the data in different ways. So you spend some time matching the formats of each data set. You generate your reports and are ready to share your analysis. But, before you can do that, a person from a third department tells you that the data you used is private, and you shouldn't be using it at all. The data catalog and data virtualization platforms upend this process, giving users seamless access to the data that is available to them.

How to Describe the Data Catalog

Consider an electronic card catalog at a library. At your local library, you sit down at a computer with the card catalog open, and you type into the search bar the name of a book, an author, or even a genre that interests you. You press enter, and the card catalog returns a list of relevant results. You click on one of those results, say, a book title, and the catalog takes you to a page with data about that book: the author, publication date, a summary, where in the library it is located, and links to related items.

The data catalog is the card catalog for our data library. You can search our data sets and pick a data set or point that is most relevant to your search. Upon selecting, you will get a page with information about the data you searched for: the steward for that data, publication date, definitions for the data points and any synonyms, links to related items, and instructions for how to find the data in the data virtualization software.

Quick Facts

- Anyone can use the data catalog, though we expect to have around 60 "power users."
- Data sets will be divided by "domain," not by business unit (e.g., we will have one "People" domain with a master data set for our people, rather than an "HR" data set).
- Each domain will have a data steward, or a person who verifies the accuracy of the data and answers questions about its use.
- Initial and ongoing training will be provided.
- Any questions should be directed to the Service Desk.