

# **Infographic: Data Catalog vs Data Dictionary**

**Hang Yang**

**SYS 5013: System Engineering Analysis**

**Prof. Hitam Zidan**

**Sep 20, 2024**

# Definitions



## Data Catalog

A data catalog is a centralized inventory for all collected data information from the organization, aiming to explore and share common understanding of the data across different platforms or departments.

VS



## Data Dictionary

A data dictionary is a collection of data elements, or metadata to describe the data information. It usually include the data element name, data type, length/size, description, etc.

# Key Features



## Data Catalog

- **Dynamic and Automation:** automatically integrates with designed sources to pull the latest metadata.
- **Support Access Control:** provides role-based or user-based access control over the data assets
- **Efficient/Rapid Retrieval Approach:** provides more search options, like filters and orderby to allow users to get more enriched and desired data result than simple keyword search.
- **Metadata and lineage tracking:** it mitigates data breaches by enabling data lifecycle tracking and data tracing

VS



## Data Dictionary

- **Static:** it stays unchanged as its early building stage until manually updates
- **Specific:** it can provide very specific and detailed information about each single data element, like data type, size, default values etc.
- **Enhance communication with standardization terms:** beyond the specifications on the data , we would also provide a riched descriptive explanation or definition for the data itself.

# Pros/Cons



## Data Catalog

### Pros

- Improve productivity and data quality
- Enhance data governance and data security
- Support data lineage tracking for better data insights
- Ensure data regulatory compliance

### Cons

- Learning curve is high to manage the data catalog
- Initial integration might be complicated

VS



## Data Dictionary

### Pros

- Efficient anomalies detection
- Enables reference relationships with physical data stores
- Improve data security and compliance
- An ideal documentation for new hires

### Cons

- Inefficient search capabilities
- Time-consuming, costly and laboring when maintaining data dictionaries manually

# Use Cases



## Data Catalog

- Snowflake data cloud use data catalog to boosting data search, discovery, tracing and governance
- Nasdaq utilizes data catalog in data discovery and govenance to build a more scalable, user-oriented platform
- Online learning platform Brainly use data lineage and data tracing from data catalog to empower data integrity and data accoutability

VS



## Data Dictionary

- Oak Ridge National Library(ORNL) maintains data dictionary to illustrate basic information on its entries
- NASA's Planetary Data System (PDS) maintains data dictionary on a webpage to allow users to learn its science concepts

# Conclusion

In most modern scenarios, we could see the data catalog become more and more popular in the data management due to its automation, efficient data retrieval and other enriched features. But it does not mean the data dictionary is exclusive from data catalog. Within a scrum team, a data dictionary maybe a good documentation for team knowledge sharing and common sense.

As more and more data teams are feeling the need for and adopting data catalogs, the difference between a catalog and a dictionary is fast vanishing—and becoming more complementary—because catalog tools now crawl and inventory data dictionaries for metadata. Data dictionaries are now an integral part of a data catalog. (Atlan, 2024)

# References

Salml, C. (2024, July 29). What is the Difference Between A Business Glossary, A Data Dictionary, and A Data Catalog, and How Do They Play A Role In Modern Data Management? Analytics8. <https://www.analytics8.com/blog/data-dictionary-vs-data-catalog-vs-business-glossary/>

Atlan. (2024, January 23). Data Catalog vs. Data Dictionary: Differences, Benefits & Use Cases. Atlan. <https://atlan.com/data-catalog-vs-data-dictionary/>