CHAPTER V

# Data Storage

“Metadata is key to ensuring that resources will survive and continue to be

accessible into the future.”[[1]](#footnote-1)

## Metadata

*What is Metadata?*

Often referred to as “information about information,” metadata is a description of data. Generally it includes answers to the most basic questions about data: how was this data collected (the means of collection)? When was this data collected? What was the purpose of collection? Who was the creator of this data? What can be said about the quality of the data? Are there any peculiar characteristics to this data?

**Types of Metadata**

*Structural Metadata:* describes how the objects that compose data are put together. For example, how pages are ordered to form chapters in a book.

*Descriptive Metadata:* “describes a resource for purposes such as discovery and identification. It can include elements such as title, abstract, author, and keywords.”[[2]](#footnote-2)

*Administrative Metadata:* it refers to the aspects of data that concern data management, such as when was it created, the types of copyrights and licensing and how to archive it and process it.

## 

## Data Warehousing

*What is Data Warehousing?*

It is a system used for storing data for reporting and analysis. They are compilations of one or more data sources in which both historical and current data is stored. It provides an opportunity to aggregate data from different sources, improving data quality. Data warehousing originated from various types of enterprises compiling data from their different divisions, but is now used by any organization, company, or individual who has large amounts of data to organize and store. Data warehousing was invented by William H. Inmon, “who first described a data warehouse as being a subject-oriented, integrated, time-variant and nonvolatile collection of data that supports management's decision-making process.”[[3]](#footnote-3)

**CKAN**

*What is CKAN?*

CKAN is a data management tool developed and managed by the Open Knowledge Foundation to make data accessible. It provides a platform for sharing, publishing, finding and using data. CKAN offers a possibility to warehouse data and make it open to the public according to the standards of open data.

*Who uses CKAN?*

CKAN is open sourced, which means it is used by both governments and non-profit organizations who would like to make their data open, public, accessible and shareable. CKAN makes it easier for data communities to share and collaborate with each other.

*Datahub*

Datahub is a data management platform based on CKAN that provides “free access to many of CKAN's core features, letting you search for data, register published datasets, create and manage groups of datasets, and get updates from datasets and groups you're interested in.”[[4]](#footnote-4)

1. Niso. Understanding Metadata. http://www.niso.org/publications/press/UnderstandingMetadata.pdf [↑](#footnote-ref-1)
2. Niso. Understanding Metadata. http://www.niso.org/publications/press/UnderstandingMetadata.pdf [↑](#footnote-ref-2)
3. Data warehouse definition. <http://searchsqlserver.techtarget.com/definition/data-warehouse> [↑](#footnote-ref-3)
4. About datahub. http://datahub.io/about [↑](#footnote-ref-4)