Data Management – Overview

Claire Osgood

November 2017

CEHI
Children's Environmental Health Initiative

What is "Data Management"?

Collecting and organizing data so that it is accurate, complete, secure, and easy to interpret and use appropriately.
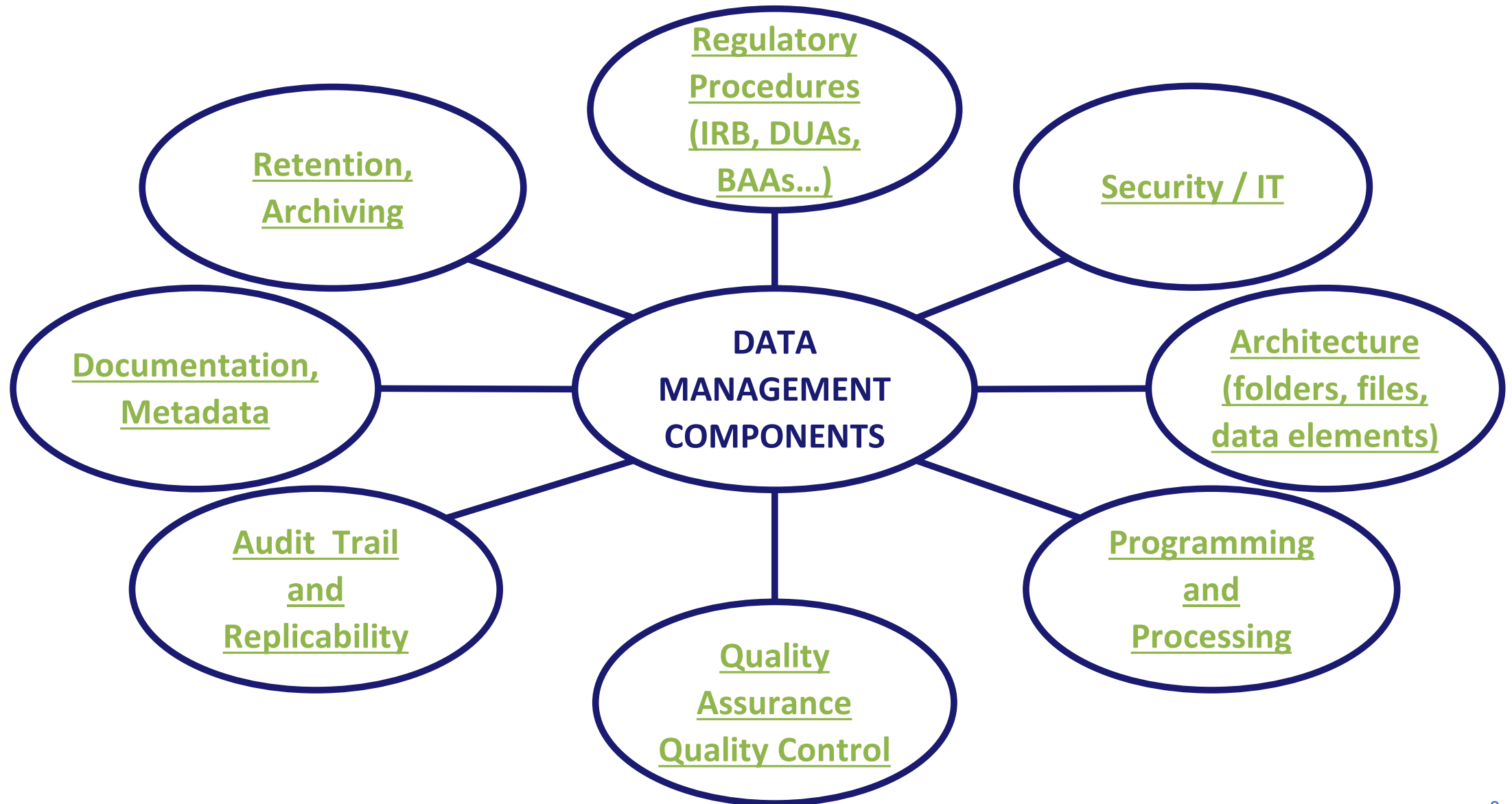
DM Components

Regulatory Procedures (IRB, DUAs, BAAs…)

IRB protocols, Data Use Agreements (DUAs), Business Associate Agreements (BAAs), compliance with both university and data supplier policies. Working with the following departments:

- Industrial Contracts
- Sponsored Projects & Research Compliance
- Office of Technology Transfer

DM Components

Security / IT

The infrastructure for a system to store and share files, mechanism for transferring files on/off the system, managing access to the system, managing permissions on the system, training users on security policies and procedures. Working with the following departments:

- IT Security
- Systems Engineering
- Networking Telecom DCO
- Campus Services (local IT)

The Human Element – You need to do your part to help!

DM Components

Architecture (folders, files, data elements)

Managing folder structure, standards for folder and file names, standard structures for common variables, standards for documentation.

DM Components

Programming and Processing

Converting raw data files received into analysis-ready datasets. Investigating, cleaning, standardizing, and linking data to other sources.

DM Components

Quality Assurance Quality Control

Checking and verifying all manual and programming processes.

DM Components

Audit Trail and Replicability

Every finalized dataset must have a clean audit trail and must be replicable. Date/Time stamps in chronological order, raw and intermediary datasets exist and are frozen, programs are checked and frozen, documentation is in order.

DM Components

Documentation/Metadata

Document every step in the process, standard documents as well as additional documentation needed, metadata for GIS files
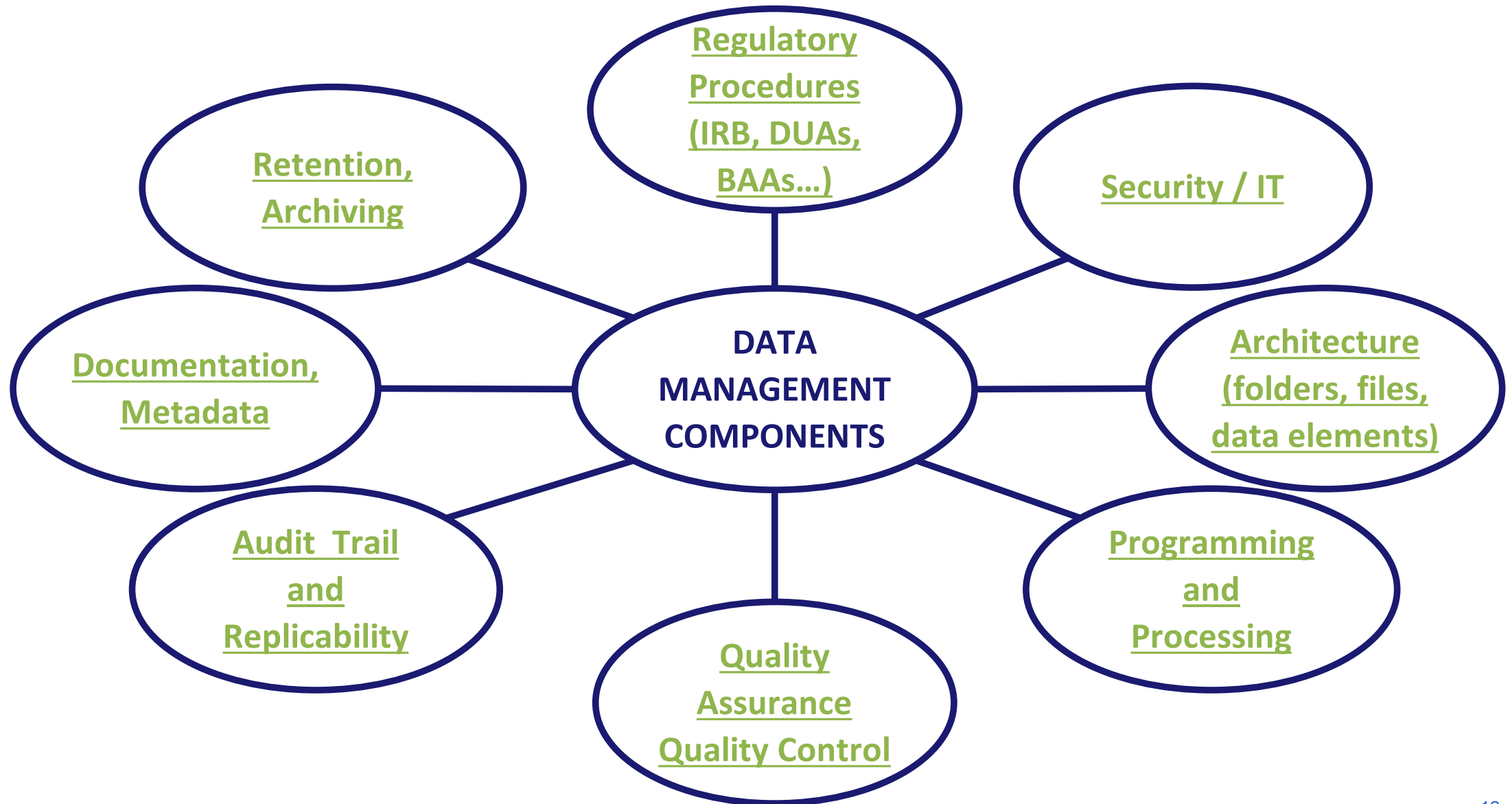
DM Components

Retention/Archiving

This varies depending on the project, IRB protocol, DUAs, and other agreements. Typically files are retained for 3-5 years, then archived for up to 7 years.

Data Life Cycle

Collect/ Acquire

Process (clean, standardize, construct variables)

Stats/ Analysis

Verify/ Check

Verify/ Check

Document

Document

Document

Store un-processed data

Processed/Analysis data (cleaned, value-added, analysis ready)

Publish results

Archive