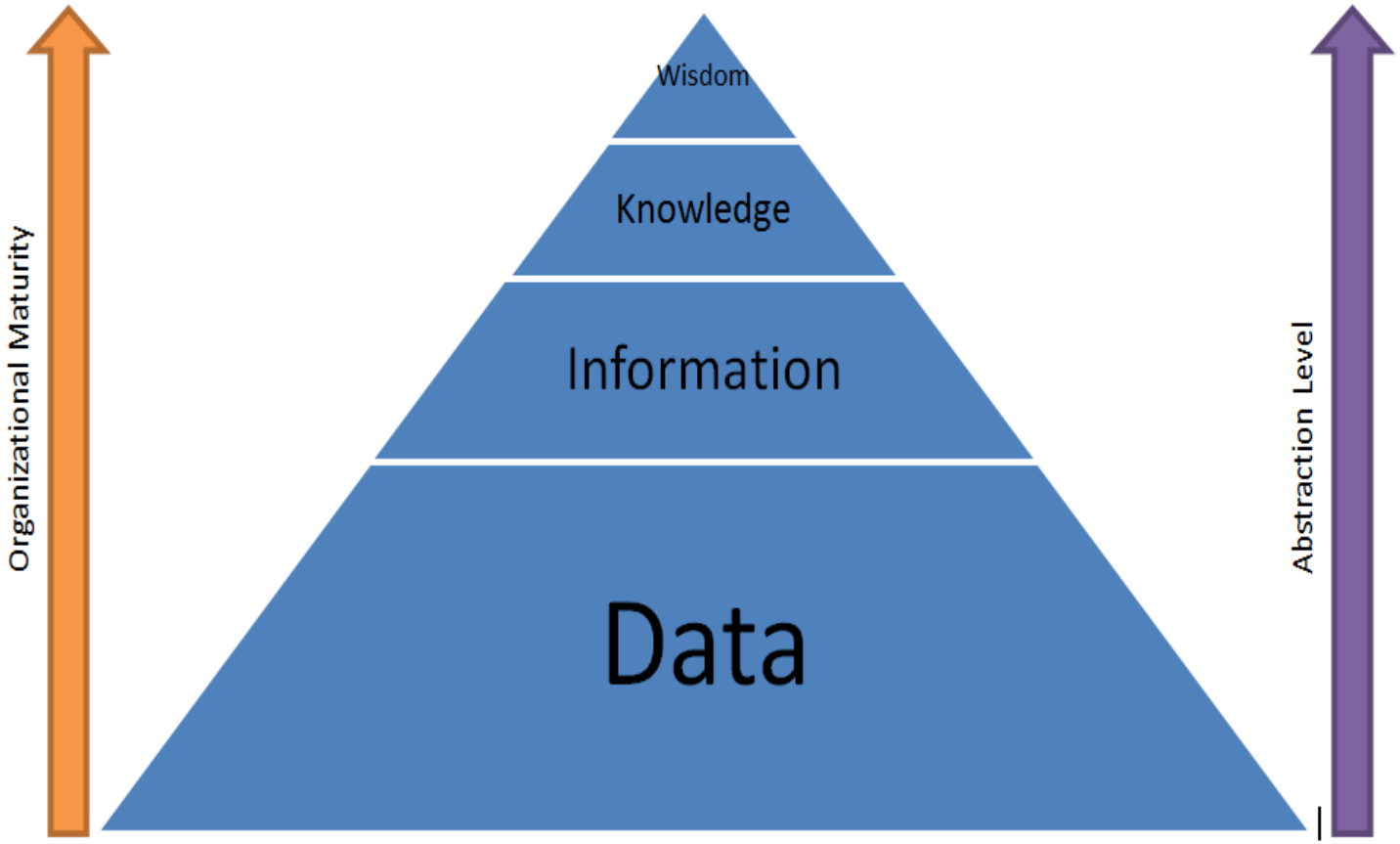
ECOCHEM
Data Quality

# Data Management: A 30 Year Retrospective


**LINDA BOHANNON**
**PRINCIPAL / QA SPECIALIST**

# The Never-ending Quest…




Andrew Kaminski, Information Management Specialist
www.rplead.com/blog/ecm/transition-data-information-knowledge-wisdom/

# Our Comfort Zone…

| Concept | Definition | Abstraction Level |
|---|---|---|
| **Data** | Simplest representation of numbers, characters, graphics, images, sound and video. Initially in 'raw' format, needs to be further processed to gain meaning. | Low |
| **Information** | Processed data, with associated metadata describing the context. Data in a context. | Medium |

Andrew Kaminski, Information Management Specialist
www.rplead.com/blog/ecm/transition-data-information-knowledge-wisdom/

# In 30 Years…

## Have We Moved Out of Our Comfort Zone???

# Comparison of Trajectory…
# Data Management and Analytical Quality

| Era | Environmental Data Management | Mindset | Analytical Quality | Mindset |
|---|---|---|---|---|
| **Early to 1979** | Focus on control of data, documents and records | Data as Data (and Records and Documents) | Focus on Analyst Quality Control (QC) at the bench | Methods do not include QC |
| **1980's 1990's** | Focus on data quality and integrity through standardized electronic formats for transmission and QC | Data as Information | Focus on Method and Laboratory QA/QC Project Quality Assurance (QA) Plans | Promulgated methods with QC elements; method compliance, metrics, records and documentation. |
| **2000 to present** | Focus on data management system planning which includes access, sharing, and re-use | Data as Knowledge | Focus on Quality System Planning and Quality Management Plan (QMP) | Desire for National QA Standard for Laboratories (NELAP) |

# Challenges to Data Quality Management (year 2000)

- Proprietary tools and reports developed by consultants and sub-consultants
- Lack of focus for standardized metadata
- Lack of focus for development of standardized data dictionaries and criteria libraries
- National historical and legacy databases (using different data structures)

*Environmental Testing & Analysis, September/October 2000*
Paul Mills & Phillip Edwards

# Challenges to Data Quality Management (year 2014)

- Multiple researchers/chemists/scientists who are singularly focused on answering specific questions or theories regarding the site or data collection
- Multiple agency oversight, usually with already existing environmental information systems
- Multiple consulting firms (concurrent or across time) with preferences usually based on business and financial goals
- Multiple laboratories (concurrent or across time) who have varying levels of LIMS and who have been tasked with producing multiple formats for EDD to support all of the above stakeholder desires

# The Final Steps Are The Hardest…

| Concept | Definition | Abstraction Level |
|---|---|---|
| **Knowledge** | Awareness, understanding, and recognition of situational patterns and trends, based on synthesis of collected information.<br>Information in a perspective. | Medium High |
| **Wisdom** | Making the best use of knowledge, acting with appropriate judgment in complex and dynamic environments, that actually achieves the stated purpose. Directly related to maturity but not related to how long the organization is in business.<br>Wisdom is applied knowledge. | High |

Andrew Kaminski, Information Management Specialist
www.rplead.com/blog/ecm/transition-data-information-knowledge-wisdom/

# Data ➡ Information ➡ Knowledge

## Why is Data Management Important?

1. Ensures that data are in consistent and correct formats, organized and annotated
2. Data are verified and validated using consistent standards and criteria
3. Avoid unnecessary duplication
4. Ability to find, retrieve and understand data when needed.
5. Enable data sharing and reuse through data archiving and preservation

# Data Management System

## Goals of a Data Management System

- System should be scalable
- All metadata should be included in system
- Access, sharing and re-use of data is key

The Data Management System is the compilation of all policy, transactions and procedures and should be documented in the Data Management Plan.

# Components of Data Management Plan

- Description of data types, which may include:
  - Analytical
  - Experimental
  - Observational
  - Images
  - Software
- Acquisition of data
- Processing e.g., software used, algorithms, scientific workflows
- Data formats e.g., hardcopy notebook, instrument outputs, ASCII, html, jpeg, etc.
- Naming conventions
- QA/QC measures for collection, analysis and processing
- Origins of existing data, if used.  Relationship to new data collected.
- Security and protection
- Responsibility for implementation and management of plan components

# Components of Data Management Plan

- Metadata Content and Format

  Metadata is sometimes referred to as the *'data about the data*'. Contextual details, including information important for using data. Includes descriptions of temporal and spatial details, instruments, parameters, units, files, etc.

- Policies for Access, Sharing and Re-Use

  *'Access to data'* refers to data made accessible without explicit request from interested party e.g., those posted on website or made available to a public database.

  *'Data sharing'* refers to the release of data in response to a specific request from a interested party.

- Obligations for Sharing Data Collected
  - Address ethical, confidentiality and intellectual property issues
  - Describe intended future uses and users for the data
- Long-Term Storage, Archive and Future Data Management
- Budget

# A Fond Farewell

Thank you for attending my presentation…Questions??

My attendance at this conference is, in a way, my  personal retrospective of 30+ years in the Environmental Quality industry. Thanks to all of my colleagues, friends and competitors for a interesting and challenging career.

Next up…a career in wine and food.