

**DO YOU KNOW WHERE THE INFORMATION IN YOUR DATA IS?**

**William Callicott**

**5807 Delta Lane Suitland MD 20746**

**E-mail: [william.callicott@noaa.gov](mailto:william.callicott@noaa.gov)**

**Presented at the THIC Meeting at the Raytheon ITS  
Auditorium, 1616 McCormick Dr**

**Upper Marlboro MD 20774-5301**

**November 5-6, 2002**

**The Premier Advanced Recording Technology Forum**

**THIC Inc.**

The logo for THIC Inc. features the text "THIC Inc." in a bold, black, sans-serif font. The "THIC" portion is set against a light brown, textured rectangular background, while "Inc." is on a white background. A thick, dark red horizontal bar is positioned below the text.

# DO YOU KNOW WHERE THE INFORMATION IN YOUR DATA IS?

In 1934, T.S. Eliot mused in his poem, “The Rock”:

*“Where is the wisdom we have lost in knowledge?  
Where is the knowledge we have lost in information?”*

# PUTTING DATA MANAGEMENT IN PERSPECTIVE

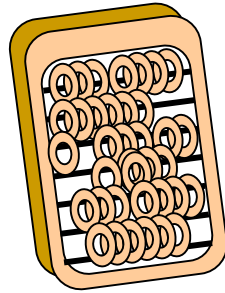
In 1985, David Byrne wrote a song called “In the Future”, in which he lampoons those who make predictions for a living:

*“In the future there will be so much going on that no one will be able to keep track of it.”*

# INFORMATION TECHNOLOGY PAST, PRESENT, AND FUTURE



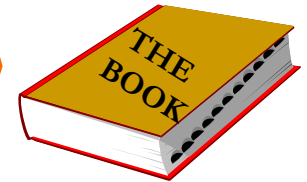
Images in Stone  
20,000+ Years Ago



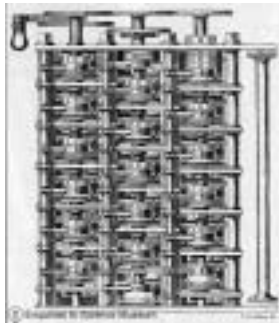
Abacus  
5,000 Years Ago



Scribed Text  
2,500 Years Ago



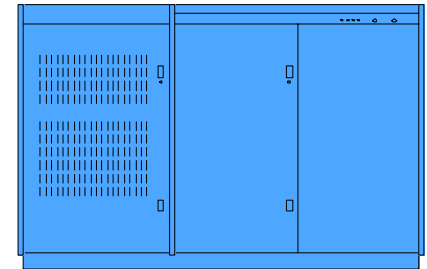
Printed Text  
550 Years Ago



Babbage Analytical Calculator  
Late 1800s



Illiac I  
Late 1940s



Mainframe Computers  
1950s to Present



Micro Computers  
1980s to Present

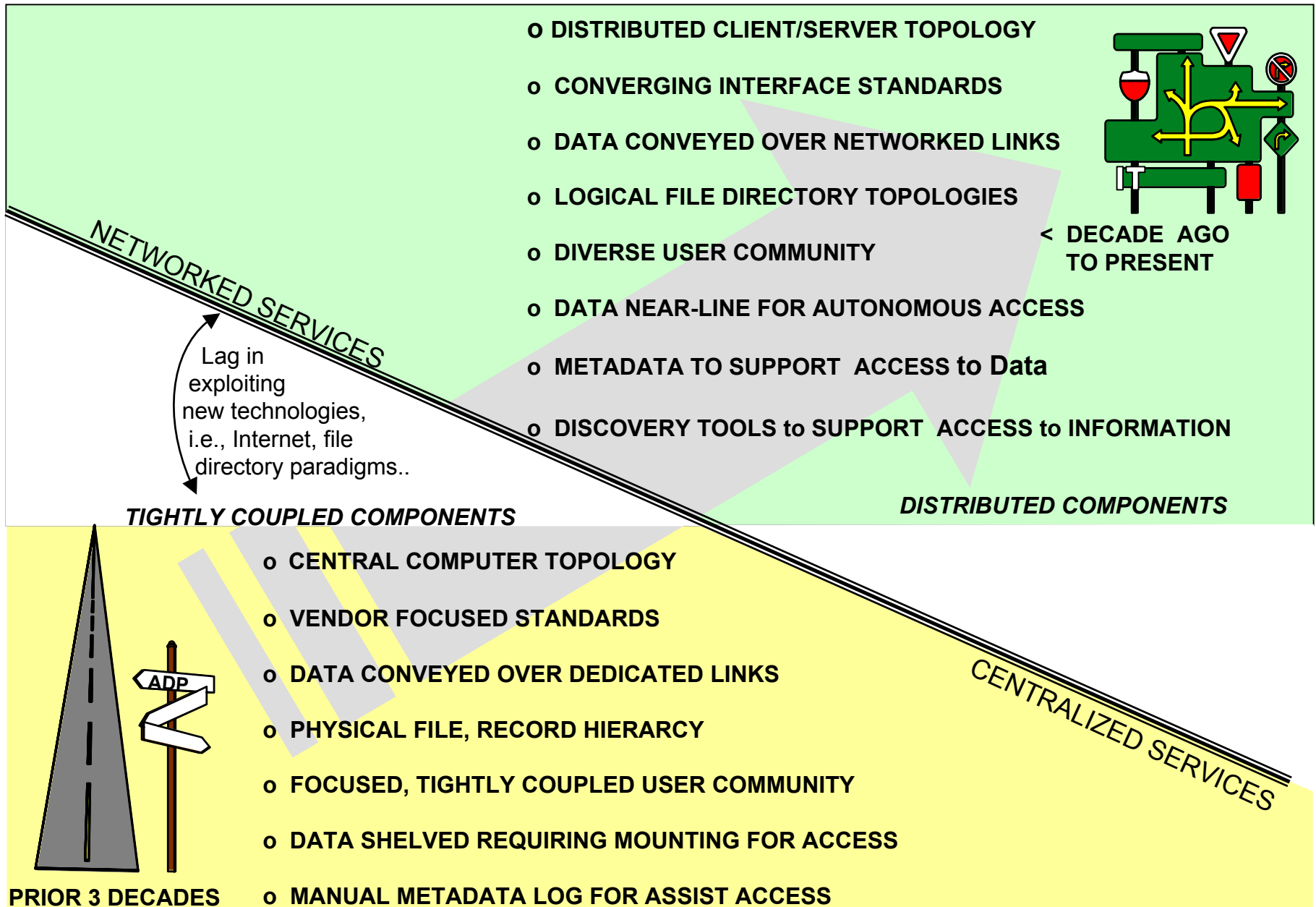


The Future

# THE WORLD OF COMPUTING HAS CHANGED THE WAY WE ACQUIRE, MANAGE, AND PROVIDE DATA AND INFORMATION

- NOW, ABOUT EVERYTHING IS DIGITAL
- ELECTRONIC NETWORKS ARE PERVASIVE
- SUPER COMPUTING IS AN AFFORDABLE COMMODITY
- ELECTRONIC ACCESS AND EXCHANGE OF DATA IS A WAY OF LIFE
- KNOWLEDGE IS THE INDUSTRY OF THE FUTURE

# TRANSITION IN DIGITAL PROCESSING



# WHY DATA CAN ACCUMULATE OUT OF CONTROL

- It's coming at a rate beyond which we can manage (more preoccupied with collecting it than managing it)
- No decision on its disposition attributed to:
  - (1) fear of being blamed for loss of an important data set (not on my watch), or,
  - (2) beyond a typical budget cycle (not obvious until it's necessary), and
  - (3) not a driver in the planning cycle (normally not a mission asset once used)
- It's value as a future asset not understood (combination of (1) and (3) above)
- As long as there is a place to shelve it (Fibber McGee's closet)

# WHY SAVE DATA AND INFORMATION?

- **An immediate asset** - used to support operational mission (in the case of NOAA, for monitoring, and supporting forecast and warning services)
- **A continuous asset** - used during a prolonged period of principal investigation (NASA EOSDIS program for example)
- **A future asset** - once used data and information known to have future value (environmental change detection, demographics, economy, astronomy, etc.)
- **Historical artifact** - Museum collections
- **Not sure** - not fully understood or appreciated, lack of decision, apprehensive of risks and blame associated with disposal of data and information, procrastination

# CHARACTERISTICS OF AGING DATA

*(AS COMPARED IN HUMAN TERMS)*

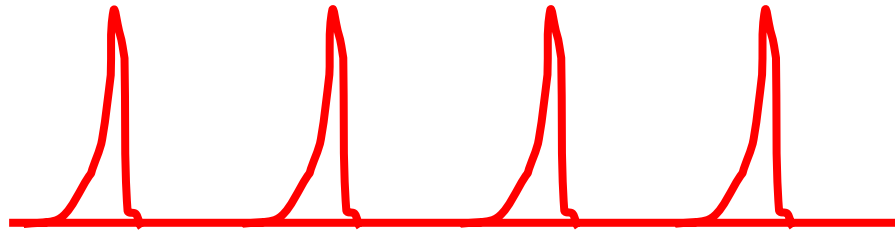
**For data and information retained beyond its principal use:**

- **contents and structures are soon forgotten**  
*(position quickly filled with one having new approaches and different methods)*
- **interest soon wanes in its absence**  
*(remembered as "someone in the past")*
- **continued care and maintenance becomes becomes a burden**  
*(health begins to deteriorate accompanied with increased care costs)*

**In order to maintain data beyond its period of principal use:**

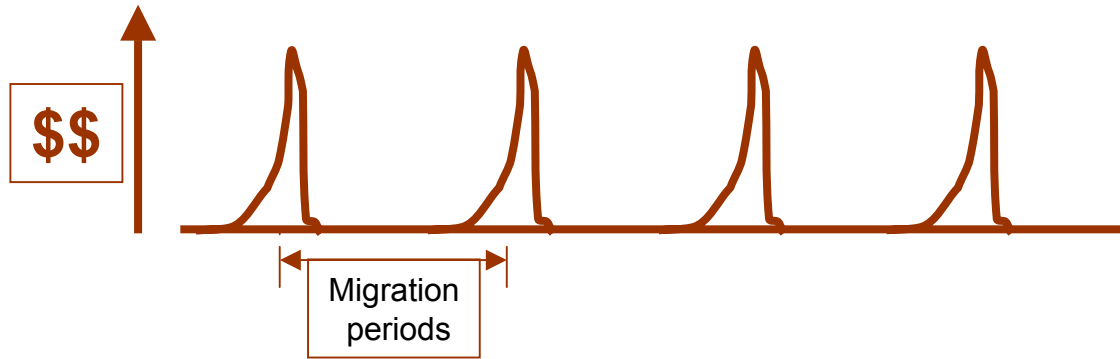
- **structured metadata must be developed to perpetuate memories**  
*(got to label all the old family pictures for future generations)*
- **access mechanisms must be enriched to perpetuate interest**  
*(family pictures displayed in albums to convey family knowledge to others)*
- **planned care necessary to extend longevity**  
*(preventative medicine, active life style, diet, and financial planning)*

# MIGRATION IS THE HEARTBEAT OF PERPETUITY!



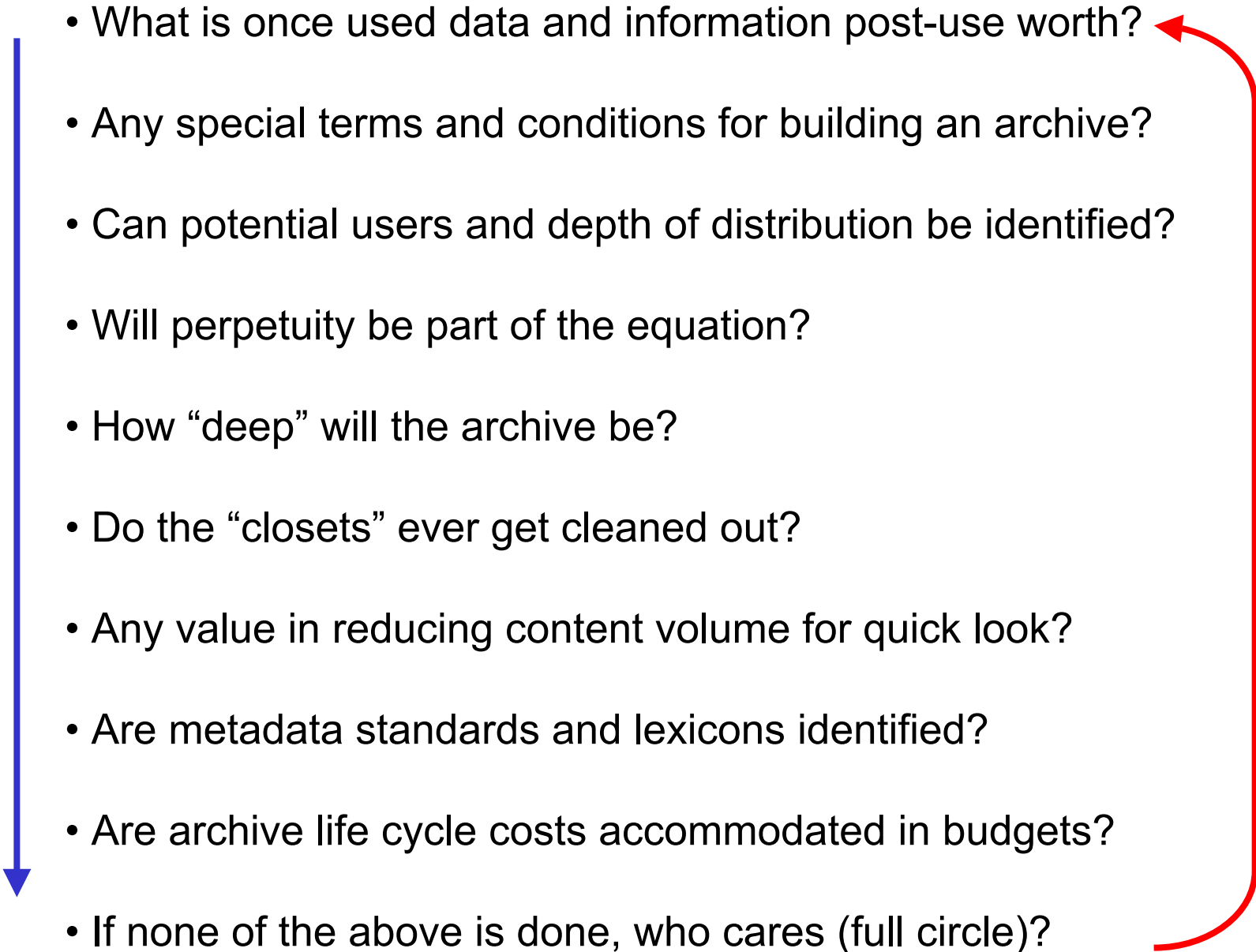
- Migration offers a “**rebirth**” cycle as data transition from one systems environment to another (*activities of perpetuity*)
- The utility of data stored long-term is directly proportional to the **understanding** of data and its characteristics (*discovery tools necessary as effective agents of understanding*)
- Each time data moves from one systems environment to another, an opportunity is presented to **re-describe its characteristics** in consonance with new user requirements and technology paradigms (*couples credibility and integrity*)
- Essential to maintain data within a current technology environment in terms of form-factor and in-use system interfaces for **perpetual utility** (*activities of usability*)

# COST OF DATA MIGRATION IS PERIODIC



- Migration costs peak in cycles, as data are normally stored as a whole within a single long-term systems environment (paradigm of most storage technology systems)
- With extended media life, the requirement to migrate now occurs when an interface service or form-factor technology becomes untenable
- Migration cycles extend beyond normal government budget cycles, abrogating any realism in amortizing migration costs
- More often than not, it is the budget bureaucracy that places data at risk

# PERPETUAL DATA MANAGEMENT CONSIDERATIONS

- What is once used data and information post-use worth?
  - Any special terms and conditions for building an archive?
  - Can potential users and depth of distribution be identified?
  - Will perpetuity be part of the equation?
  - How “deep” will the archive be?
  - Do the “closets” ever get cleaned out?
  - Any value in reducing content volume for quick look?
  - Are metadata standards and lexicons identified?
  - Are archive life cycle costs accommodated in budgets?
  - If none of the above is done, who cares (full circle)?
- 

## **IF METADATA IS?**

- An inventory of holdings.
- A description of data context and content.
- A resource for launching relational searches.

## **THEN, WHAT IS DISCOVERY?**

- Happening upon unknown relationships extracted from metadata search results
- Anything beyond discovery is luck!

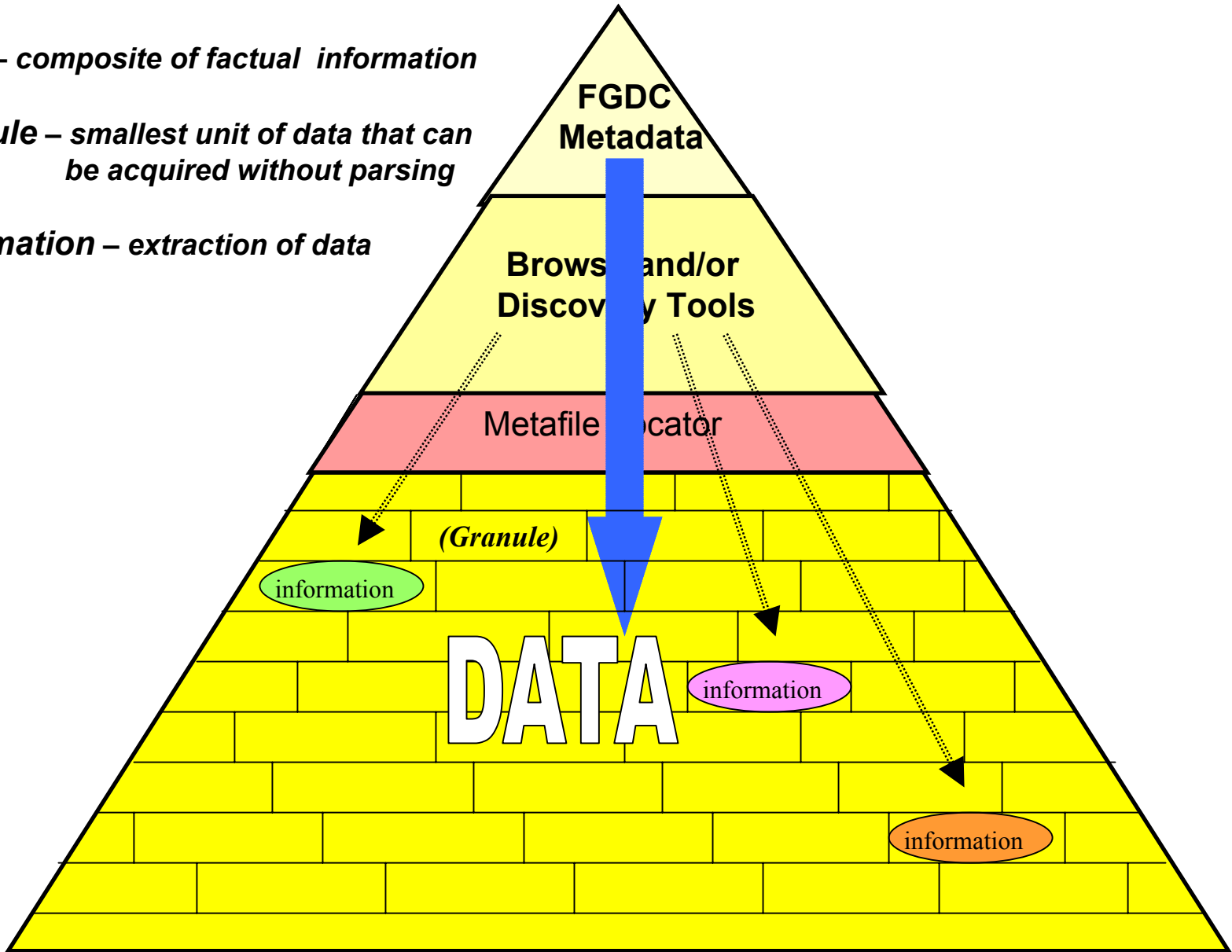
## **HOW COULD ONE ASSIST IN DISCOVERY?**

- Mine metadata to form a metadata hierarchy.
- Enriching metadata through dynamic feedback mechanism
- Maintain metadata as a living document expanding context over time

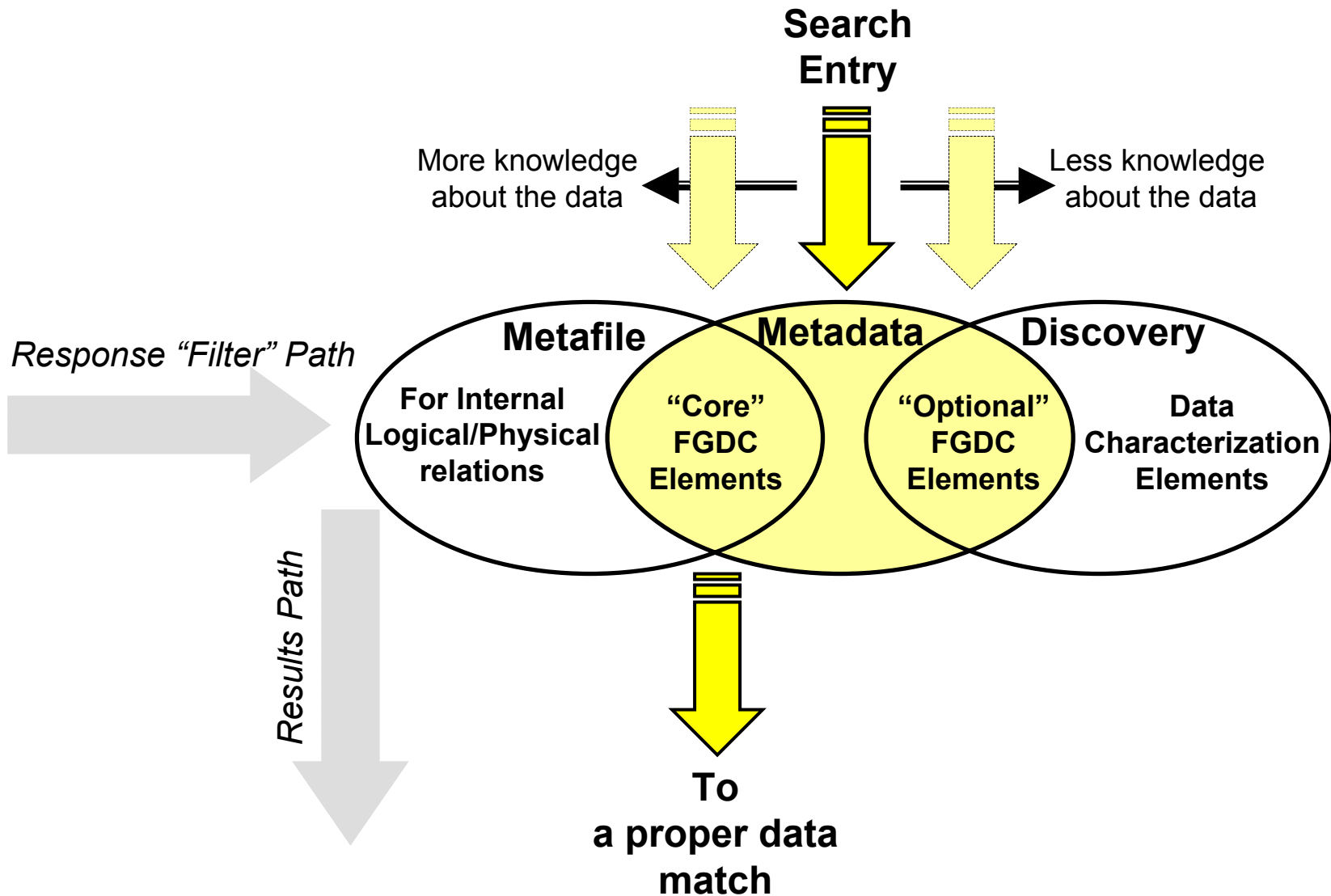
**Data** – composite of factual information

**Granule** – smallest unit of data that can be acquired without parsing






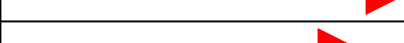














**Information** – extraction of data



# HIERARCHY OF METADATA SUPPORT



## DATA STORAGE (Sources: The Web pages)

Fossils		Millions of years
Clay tablets		Thousands of years
Scribed vellum		1,500 years
Rossetta ion beam		1,000 years
Bound alkaline paper		1,000 years
Silver halide Microfilm		300 years
B&W film & prints		250 years
Phthalocyanine CD		150 years
Kodachrome		100 years
Cyanine CD		50 years
Human productivity		50 years
Acidic paper		30 years
Color film		30 years
Digital Tape		30 years
Metadata proficiency		30 years
Laser/Inkjet printout		30 years
Frozen DNA		20 years
Digital disks		20 years
Flash memory		10 years
 IT Obsolescence		10 years

15. \*\* Estimates based on optimum dark storage conditions