The Premier Advanced Recording Technology Forum

THIC Inc.

Archive – Where it Started, and the Problems of Perpetuity

William Callicott

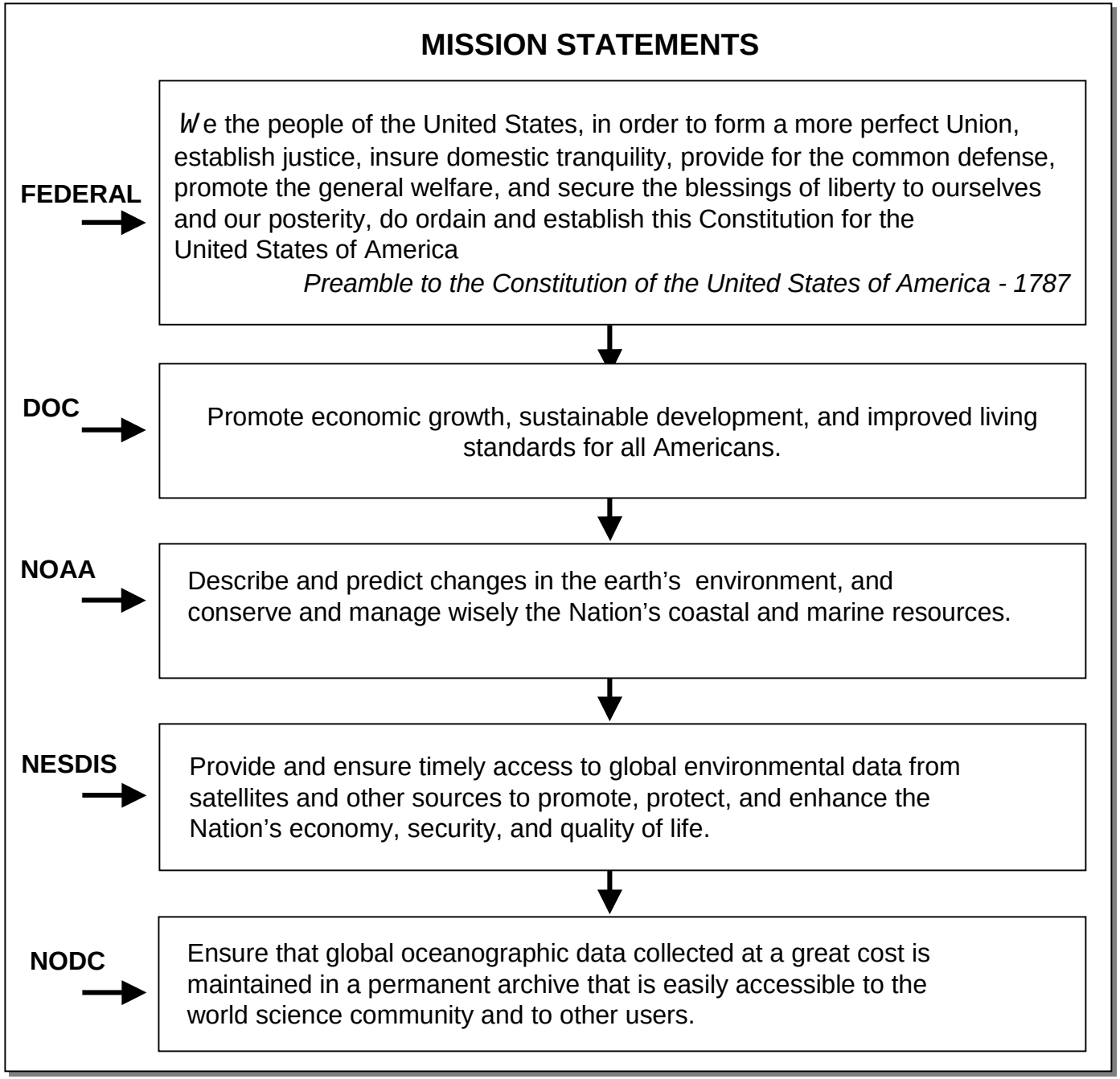
NOAA

Wcallico@erols.com

Presented at the THIC Meeting at the Naval Surface Warfare Center Carderock, 9500 MacArthur Blvd

West Bethesda MD 20817-5700

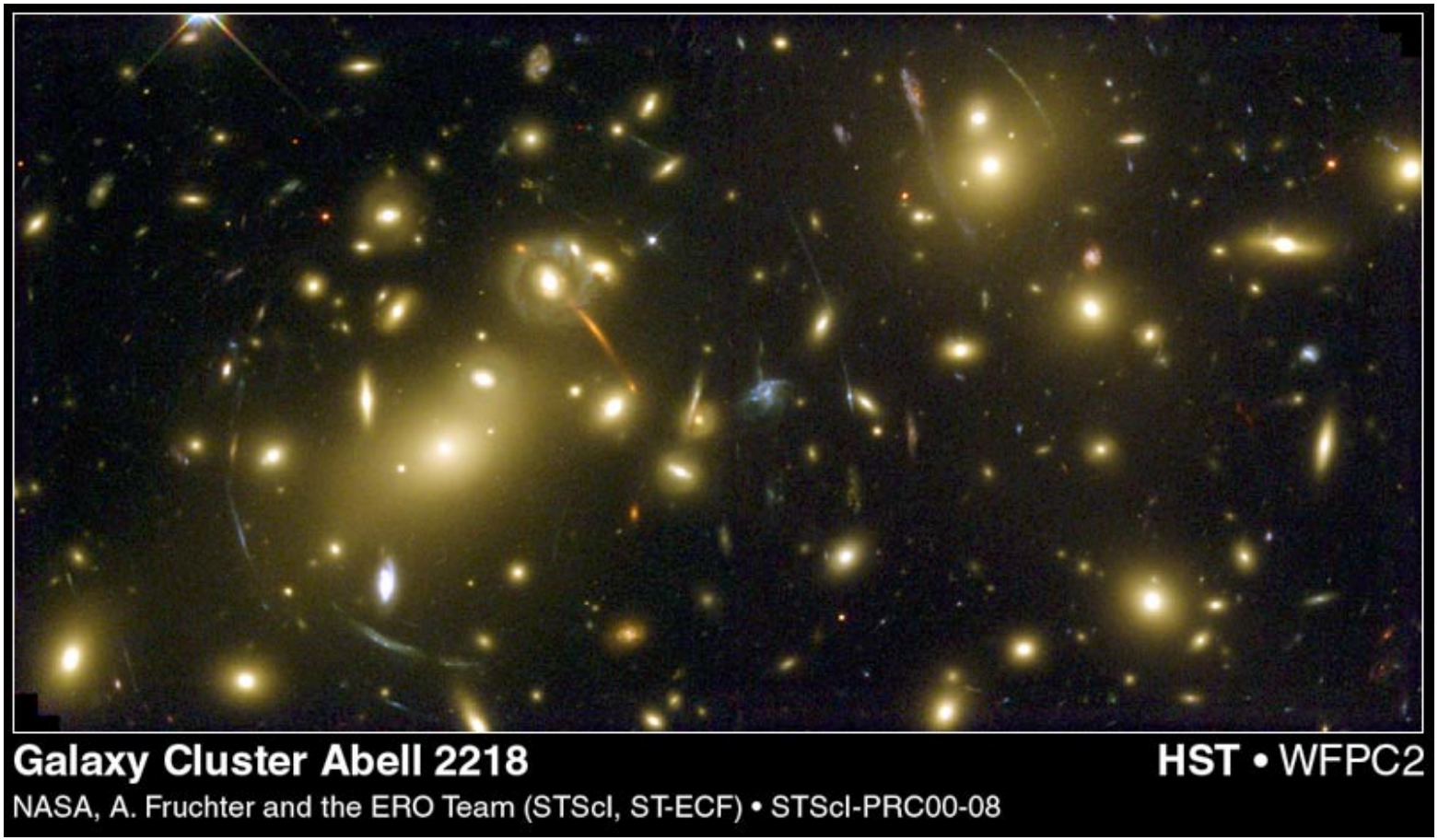
October 3, 2000


MISSION STATEMENTS
FEDERAL
We the people of the United States, in order to form a more perfect Union, establish justice, insure domestic tranquility, provide for the common defense, promote the general welfare, and secure the blessings of liberty to ourselves and our posterity, do ordain and establish this Constitution for the United States of America
Preamble to the Constitution of the United States of America - 1787
DOC
Promote economic growth, sustainable development, and improved living standards for all Americans.
NOAA
Describe and predict changes in the earth's environment, and conserve and manage wisely the Nation's coastal and marine resources.
NESDIS
Provide and ensure timely access to global environmental data from satellites and other sources to promote, protect, and enhance the Nation's economy, security, and quality of life.
NODC
Ensure that global oceanographic data collected at a great cost is maintained in a permanent archive that is easily accessible to the world science community and to other users.

NUMBEROLOGY (invented word)

The 1996 CocaCola Company Annual Report reads:

"A billion hours ago, human life appeared on earth.
A billion minutes ago, Christianity emerged.
A billion CocaColas ago was yesterday morning."

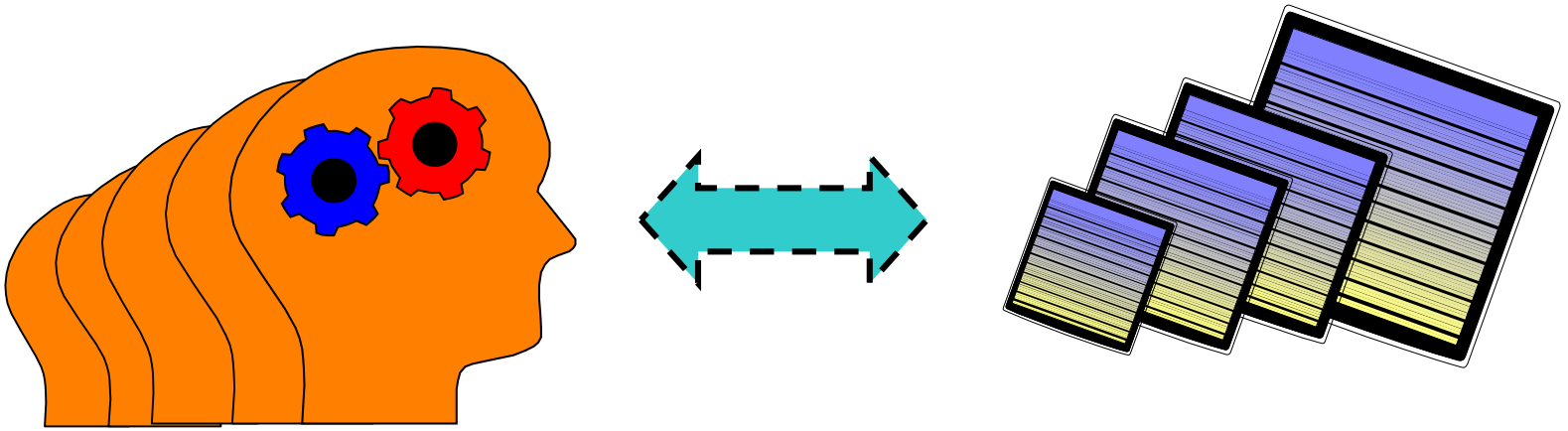


Since our Milky way galaxy is estimated to contain upwards of 200 billion stars, let's assume, on an average, the galaxies in the universe contain 100 billion stars each.

The estimate for the number of galaxies in our universe ranges from 50 billion to 100 billion (not sure that what we see back in time has not compressed itself to nothingness by now)

At a minimum, the number of stars could be 5 sextillion (10^{21}), so there soon will be more bytes of data (a yottabyte- 10^{24} is on the horizon) than stars in the universe we know.

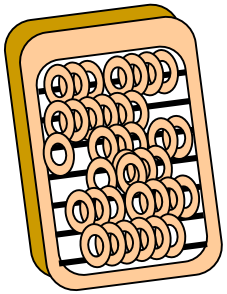
PROGRESSION OF KNOWLEDGE IN TERMS OF CAPABILITY



TIME

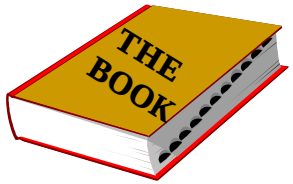
INFORMATION TECHNOLOGY
PAST, PRESENT, AND FUTURE

Images in Stone
20,000+ Years Ago

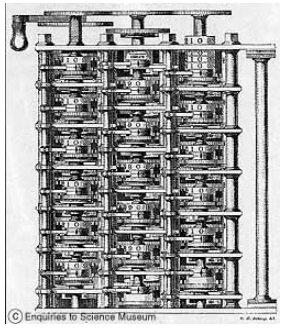


Abacus
5,000 Years Ago

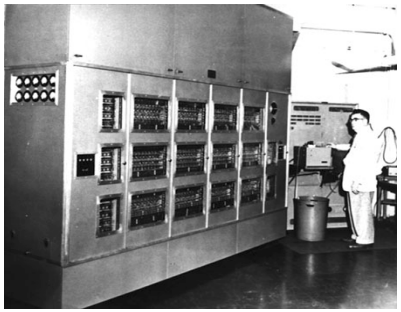
Scribed Text
2,500 Years Ago



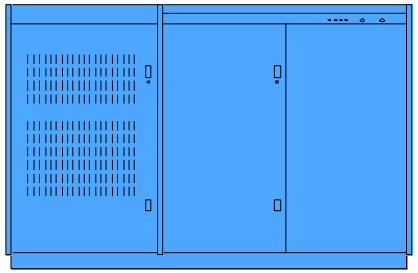
Printed Text
550 Years Ago



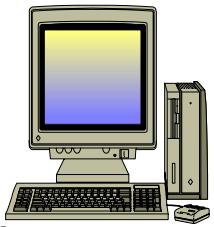
Babbage Analytical Calculator
Late 1800s



Illiac I
Late 1940s



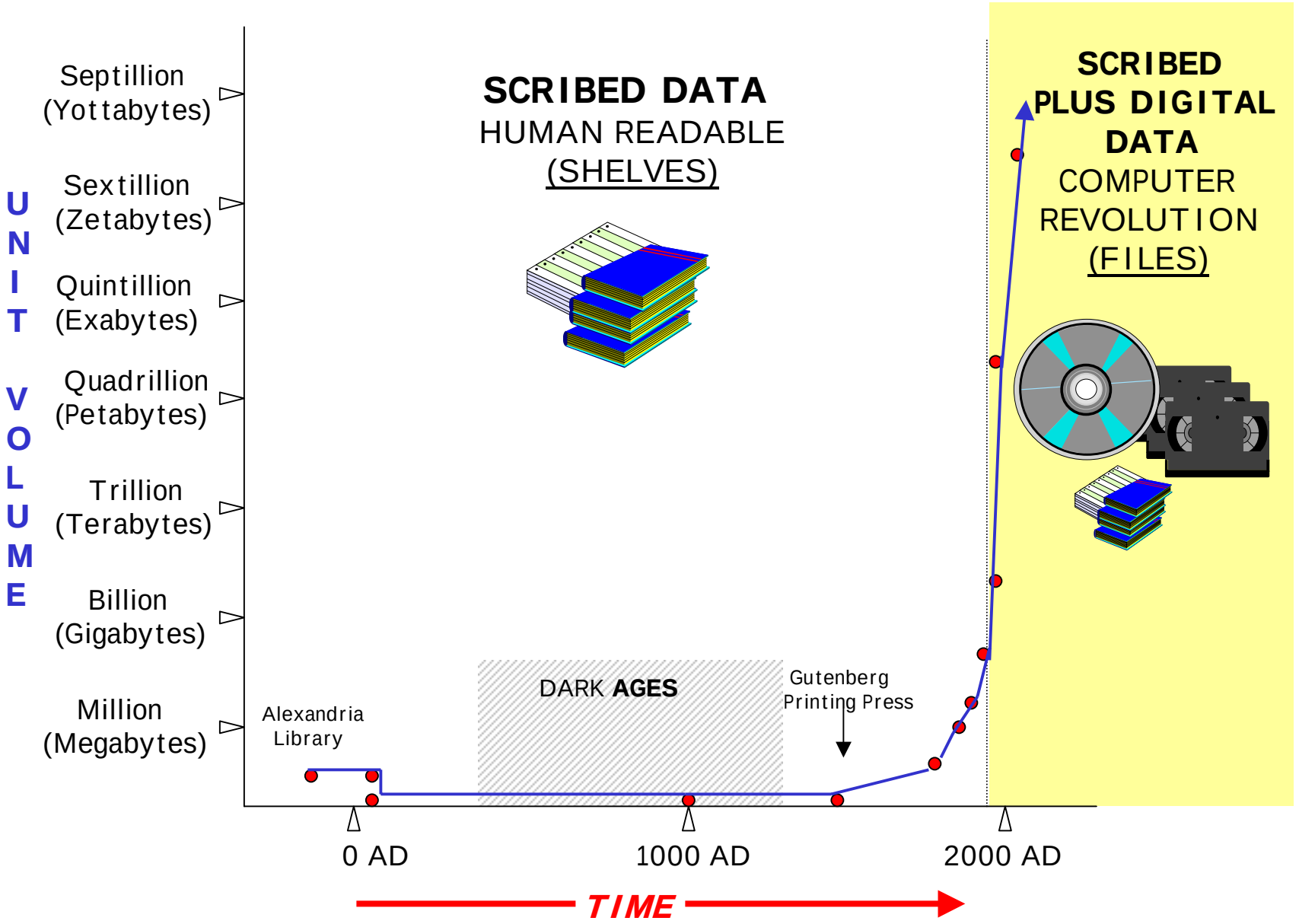
Mainframe Computers
1950s to Present



Micro Computers
1980s to Present

The Future

THE DATA AND INFORMATION EXPLOSION



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.”

Source: Mike Mills
Washington Post Bookworld
December 21, 1997

CONNOTATIONS

ARCHIVE: *noun* - '*Repository for stored memories or information.*'

Passive Data center connotation - A storage of data and information (at a minimum - one copy stored off-line in a deep archive)

PRESERVE: *verb* - '*Safety from injury, peril; protect, keep in a perfect, or unaltered condition; maintain unchanged.*
From Latin 'servare' to guard.'

Active Data center connotation - To maintain data and information in an access environment for perpetuity. Requires recurring maintenance to insure integrity, to secure from alteration, and to provide for utility in the future.

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

IF DATA IS A "PRODUCT OF INVESTIGATION", YOU DON'T WANT TO THROW THE EVIDENCE AWAY!

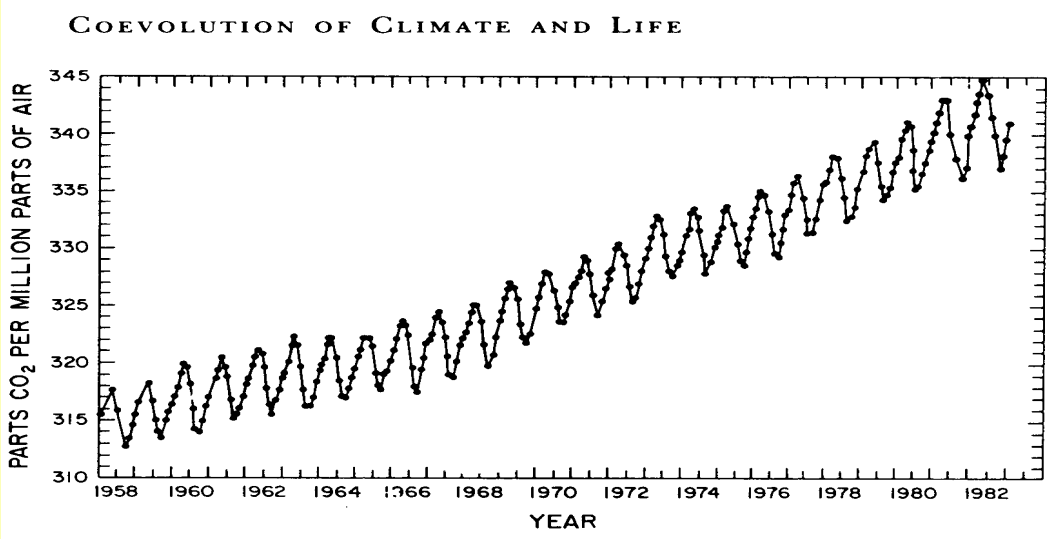
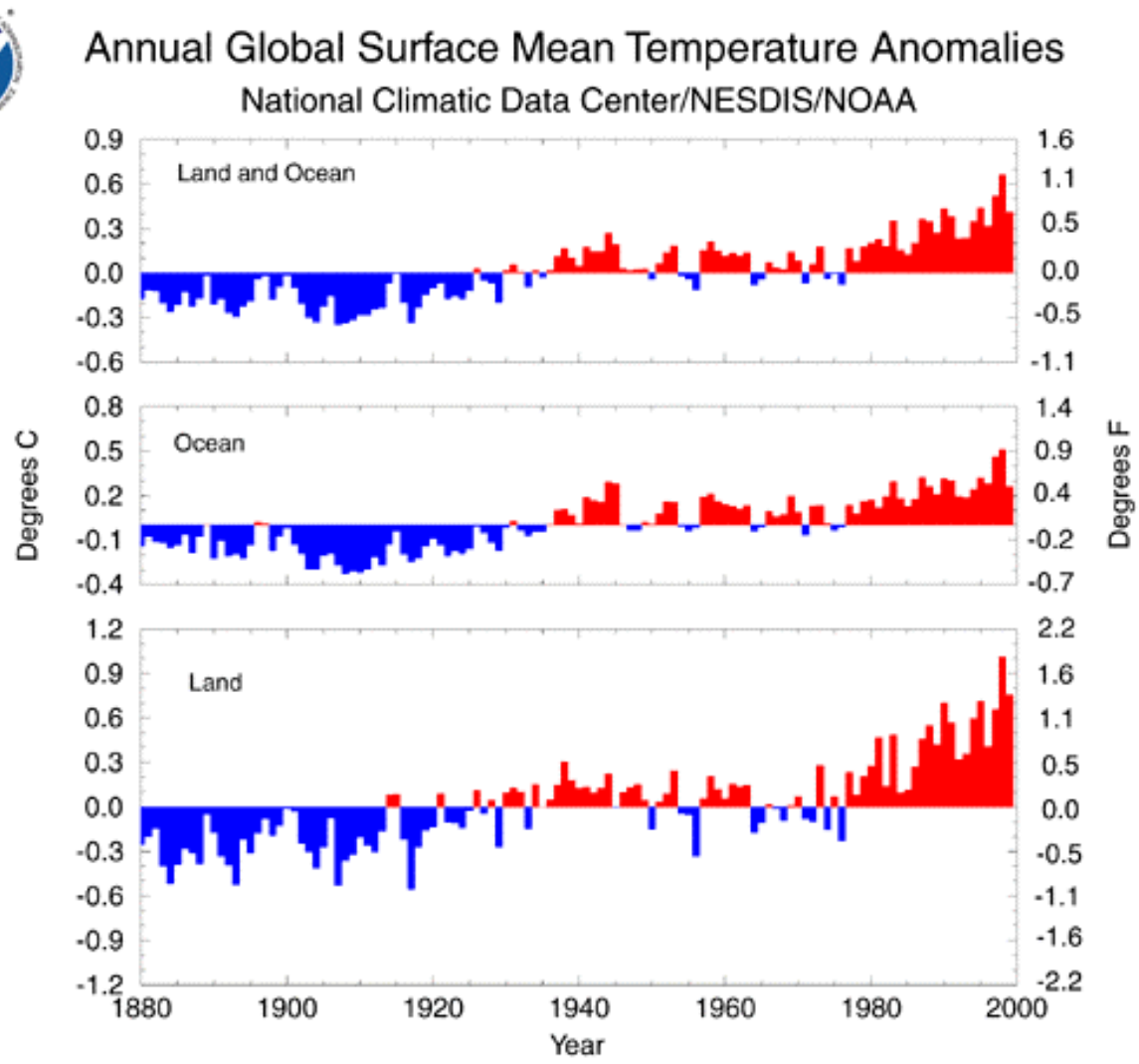


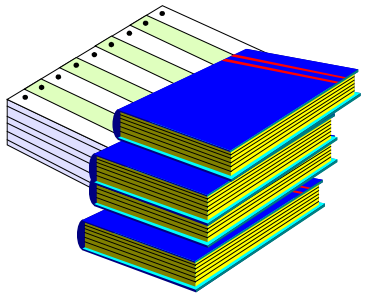
Figure 8.1

The observed trend of the concentration of atmospheric carbon dioxide (CO_2) as measured at Mauna Loa Observatory on the island of Hawaii. Each year CO_2 undergoes a cycle caused by the growth and decay of seasonal plants. Superimposed on this annual cycle is a long-term upward trend of some 9 percent over the 24-year period of record shown. The trend is widely believed to be the result of human activities, and it could cause significant global climatic warming if it continues. [Source: U.S. National Oceanic and Atmospheric Administration data, based initially on the work of C. D. Keeling at the Scripps Institution of Oceanography.]


Annual Global Surface Mean Temperature Anomalies
National Climatic Data Center/NESDIS/NOAA
Land and Ocean
Ocean
Land
Degrees C
Degrees F
Year
1880
1900
1920
1940
1960
1980
2000

ARCHIVE ATTRIBUTES

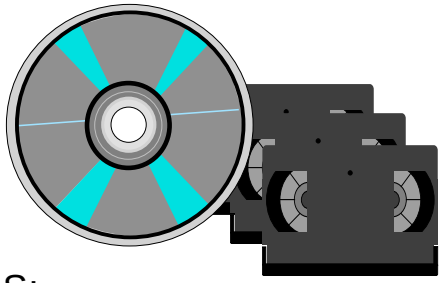
SCRIBED



PROS:
- Human Readable
- Standards optional
- Simple to catalog

CONS:
- Bulk
- Narrow breadth of access
- Manual search
- Replication limited
- Backup copy limited
- Time consuming

DIGITIZED

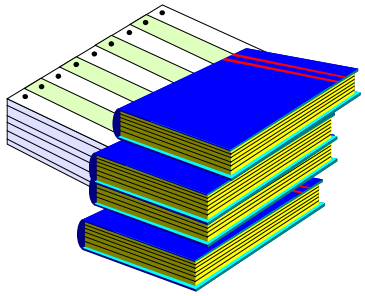


PROS:
- Compact
- Easily replicated
- Amenable to search queries
- Broad breadth of access
- Amenable to backup protection
- Info can be manipulated

CONS:
- System dependent
- Standards adherence
- Longevity dependent on many variables (i.e., media, systems, environment, preventative maintenance)
- Extraordinary steps required to avoid corruption (handling, controlled access)
- Long-term credibility directly proportional to extent of metadata
- Info can be manipulated

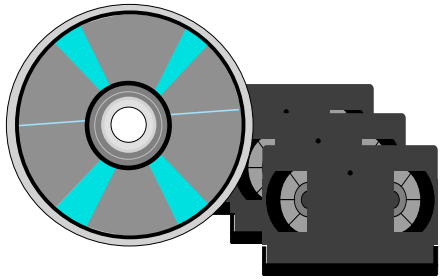
ARCHIVE ISSUES

SCRIBED



- Storage Space
- Logical/Physical management
- Life cycle management
- Query servicing
- Total or partial replication
- Access services supported
- Loss avoidance

DIGITIZED



- Sufficient metadata for long-term servicing
- Systems obsolescence
- Media deterioration
- Form factor changes
- Mandatory migration to extend data life
- Intrusion protection
- Copy credibility
- Standards utility
- Disposal
- Access services supported
- Maintenance costs
- Backup

Even in 1934, T.S.Eliot foresaw a fundamental problem with what seemed to be a data explosion then when he wrote the dramatic poem, ***"The Rock",*** where he lamented:

Where is the wisdom we have lost in knowledge?
Where is the knowledge we have lost in information?

Imagine how he would have felt today……

ACCESS

Access is more than a value added tool as it:

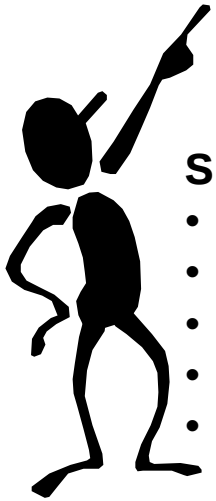
- Enhances the value of data and information
- Increases the likelihood of its longevity through use

Real value of data and information is directly proportional to the degree of access to it.

Enables the purpose of data and information to expand the breadth and wealth of knowledge

BROWSE IS GOLDEN, BROWSE IS SURE,
SO ACCESS IS PURE - FROM WHAT
YOU'RE HOLD'N!

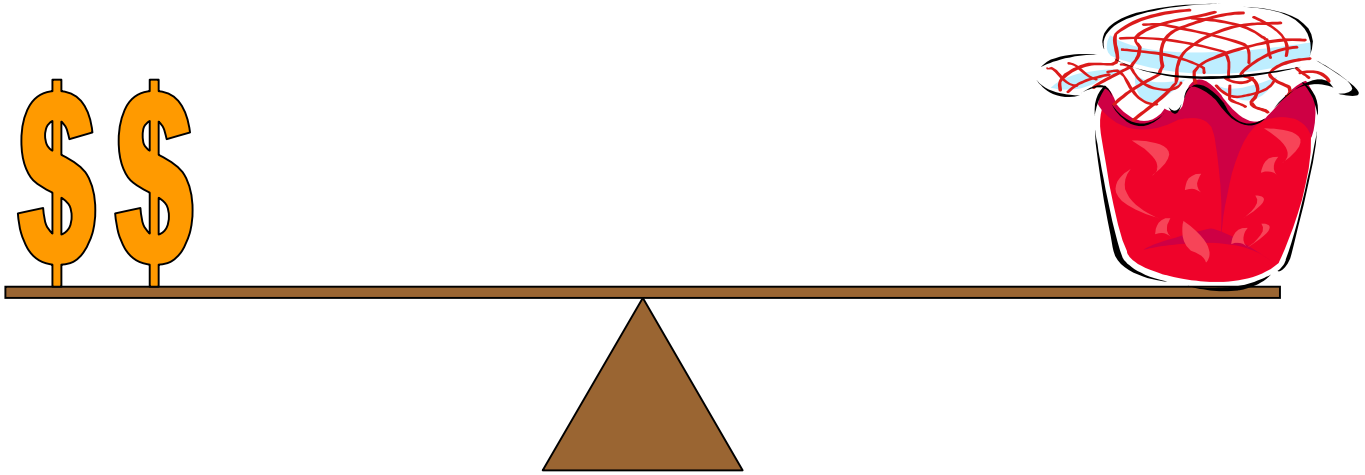
ALSO, BROWSE SHOULD BE FREE,
TO ENCOURAGE ONE TO SEE……..



SUGGESTIONS:

- Large volume data sets should incorporate static browse as an additional data set.
- Small volume data sets should be supported with dynamic browse capabilities
- Browse should be real, relating to the data set being queried
- A browse item may suffice as data to some users
- Browse is a marketing investment to increase product(s) exposure

PRESERVATION



There was an article in the *Washington Post* newspaper by Joel Achenback, April 19, 1998, which addressed the approaching singularity of technology.

It states:

"the moment in the future when so many technologies have converged - computers, miniaturization, bio-medicine - that they become "auto-catalytic", driving one another to yet greater sophistication, "hyper-accelerating". Predictions will be worthless because everything is changing so fast - an event horizon beyond which we can detect nothing!"

This is the wave of the future!