Rensselaer

why not change the world?®

Information integration, life-cycle and visualization / Group Projects

Ahmed Eleish

February 19th, 2025

ITWS, ERTH, CSCI 4400/6400

Contents

- Communicating Uncertainty
- Information integration
- Information life-cycle & Management
- Information visualization
- Group Projects – Exploring Ideas and setting up Groups.

- Assignment 3

Communication of Uncertainty

- Article on “Communication of Uncertainty”

https://www.nap.edu/read/12568/chapter/8

PRESENTATION OF UNCERTAINTY

- The most widely used formal language of uncertainty in risk estimates is probability[1] (Morgan, 2009).
- As Spiegelhalter et al. (2011) stated, however, "probabilities are notoriously difficult to communicate effectively to lay audiences." **Probabilistic information, and the uncertainties associated with those probabilities, can be communicated using numeric, verbal, or graphic formats**, and consideration should be given to which approach is most appropriate.

Resource/Reference: https://www.nap.edu/read/12568/chapter/8#183

Uncertainty

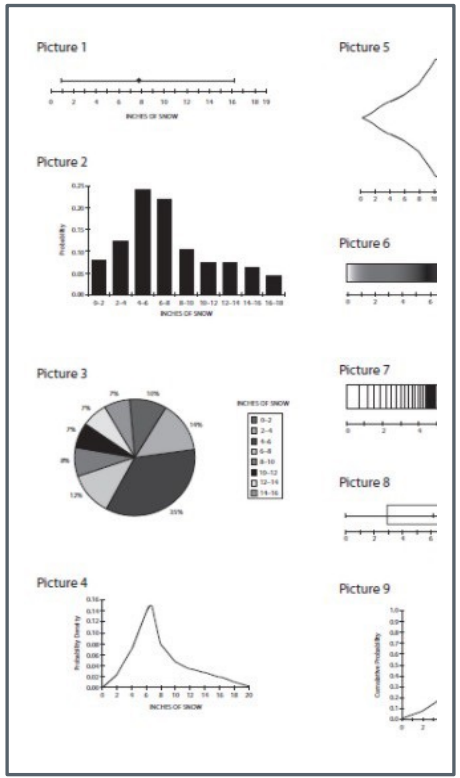
- **Graphical Presentation of Uncertainty:**

Graphical displays of probabilistic information—such as bar charts, pie charts, and line graphs—can summarize more information than other presentations, can capture and hold people's attention, and can show patterns and whole-to-part relationships (Budescu et al., 1988; Spiegelhalter et al., 2011).

Furthermore, uncertainties about the outcomes of an analysis can also be depicted using graphical displays, such as bar charts, pie charts, probability density functions,[2] cumulative density functions,[3] and box-and-whisker plots. There is some evidence that graphic displays of uncertainty can help convey uncertainty to people with low numeracy (Peters et al., 2007). A few studies have explored how well different graphical displays of quantitative uncertainty can convey information and have analyzed the effects of different graphical displays on decision making (Bostrom et al., 2008; Visschers and Siegrist, 2008).

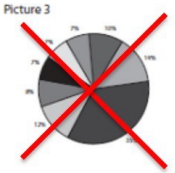
Resource/Reference: https://www.nap.edu/read/12568/chapter/8#183

- Picture 1: point estimate with an error bar;
- Picture 2: bar chart;
- Picture 3: pie chart;
- Picture 4: conventional probability density function;
- Picture 5: probability density function of half its regular height together with its mirror image;
- Picture 6: horizontal bar shaded to display probability density using dots;
- Picture 7: horizontal bar shaded to display probability density using lines;
- Picture 8: Tukey box plot modified to exclude the maximum and minimum values and to display the mean with a solid point;
- Picture 9: conventional cumulative distribution function.
SOURCE: Ibrekk and Morgan, 1987, p. 521. Reprinted with permission of John Wiley & Sons Ltd.

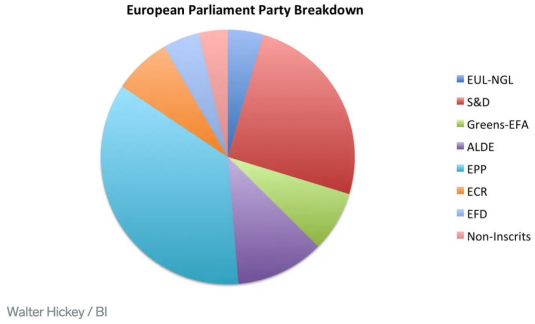


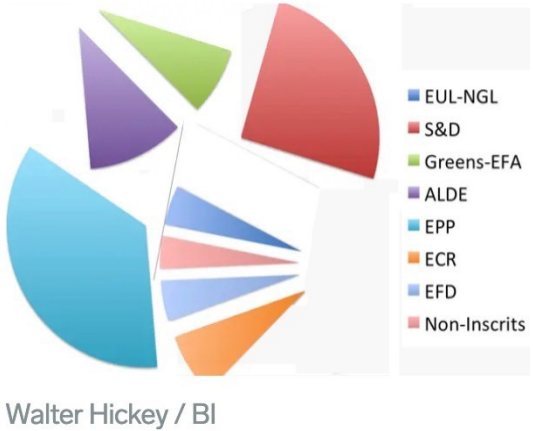
Image, Resource/Reference Credit:
https://www.nap.edu/read/12568/chapter/8[9]#183

Why Pie charts are BAD?



"The pie chart is easily the worst way to convey information ever developed in the history of data visualization."





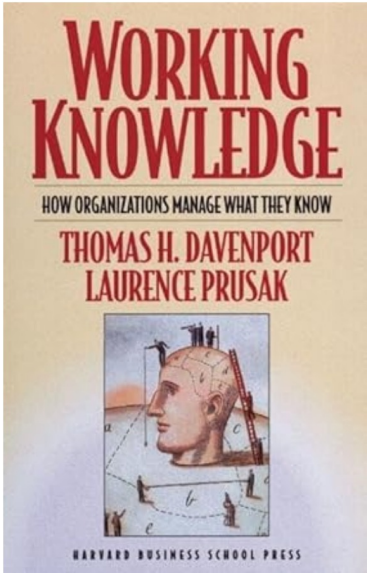
At right hand side image, the individual slices taken out of context for the purpose of comparison. Look at them, and see if you can figure out an ordering from largest to smallest.

Read: Pi charts are the worst:
https://www.businessinsider.com/pie-charts-are-the-worst-2013-6

Information Integration

For Data to become useful..

- For the data to become useful (their definition of information is data that is organized somehow) we have to do something to it. It needs to be transformed.
- Davenport and Prusak suggest that there are "5 Cs" to how we might do that. These are:



Resource/Reference: *Working Knowlege*, (HBS Press, 1998) book by Thomas Davenport and Larry Prusak,

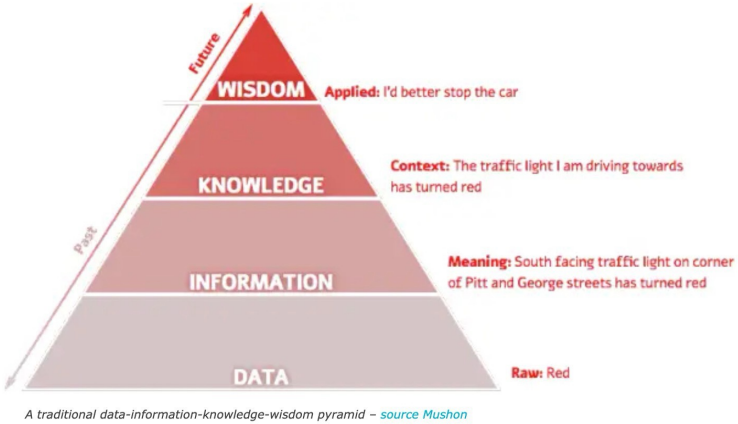
“5 Cs”

1. How is the data **contextualized**? Do we know why the data was gathered?
2. How was the data been **categorized**? Do we know the units of analysis, the key components of the data?
3. How was the data **calculated**? Have there been some mathematical or statistical analysis, such as changes over time, averages, etc?
4. What **corrections** have been applied to the data? Do we know how and whether or not errors have been removed?
5. And finally, has the data been **condensed**? Are there summaries, tables, graphics?

Resource/Reference: *Working Knowlege*, (HBS Press, 1998) book by Thomas Davenport and Larry Prusak,

From Information to Knowledge

- Turning information into action is the next step up the pyramid and what defines knowledge.
- Some call knowledge "actionable information." Another transformation is called for: Davenport and Prusak helpfully provide another list (and another set of "Cs"):



A traditional data-information-knowledge-wisdom pyramid – source Mushon

Image Credit/Resource: https://www.i-scoop.eu/big-data-action-value-context/dikw-model/

Resource/Reference: Working Knowlege, (HBS Press, 1998) book by Thomas Davenport and Larry Prusak, https://www.i-scoop.eu/big-data-action-value-context/dikw-model/

“5 more Cs”

1. The information is **compared.** How does this situation compare to other situations we have been in?
2. The **consequences** are identified. What implications does the information have for decisions and actions?
3. **Connections** have been made. How does this bit of knowledge relate to others?
4. A **conversation** is initiated. What do other people think about this information?

- Resource/Reference: *Working Knowlege*, (HBS Press, 1998) book by Thomas Davenport and Larry Prusak,

Information integration

- Involves combining information residing in **different sources** and providing users with a unified view of them.
- This process becomes significant in a variety of situations both commercial (e.g. when two similar companies need to merge their databases) and scientific (e.g. combining research results from different bioinformatics repositories).
- Integration appears with increasing frequency as the volume and the need to share existing information explodes.
- Combines information from disparate data sources and displays it in a single integrated framework

Information integration

- It has become the focus of extensive theoretical work, and numerous open problems remain unsolved.
- In management circles, people frequently refer to data integration as "Enterprise Information Integration" (EII)" (check out Wikipedia article for EII)
- Is this an information management challenge (rhetorical question)?
- Integration discussion context
  - Data Integration vs. Data Interoperability

An example – Geospatial Data

- Much of the work on information integration has focused on the dynamic integration of structured data sources, such as databases or XML data.
- **With the more complex geospatial data types, such as *imagery, maps, and vector data*, researchers have focused on the integration of specific types of information, such as placing points or vectors on maps, but much of this integration is only partially automated**.
- The challenge is that the dynamic integration of online data and geospatial data is beyond the state of the art of existing integration systems.

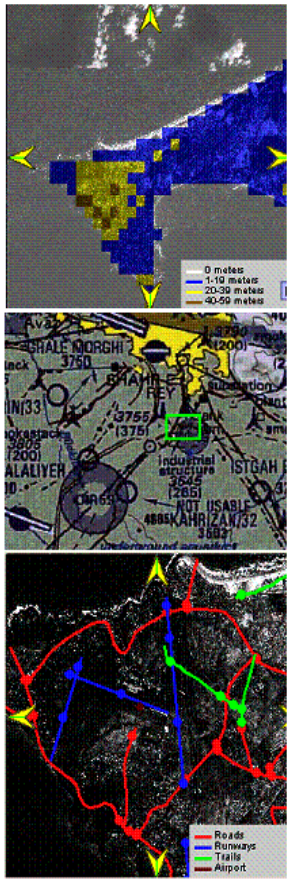


Figure 1: Image, Map, Vector, and Elevation Data Displayed in Heracles

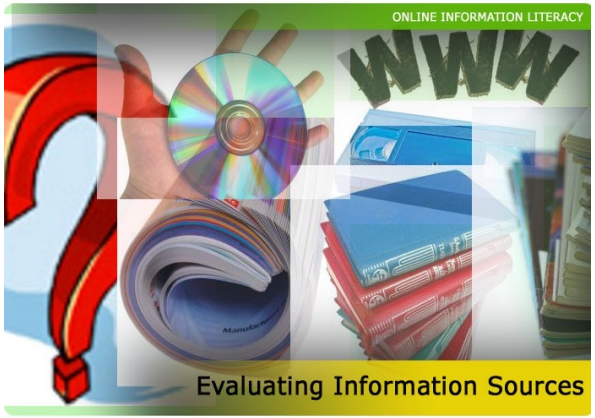
Resource/Reference/Image Credit: http://www.isi.edu/integration/TerraWorld/

Elements/ Forms of Information

- Structured/ un-structured, content, context
- Presentation and organization
- Syntax-semantics-pragmatics
- Managed, designed and architected.

Elements/ Forms of Information

- Integration poses an important challenge here
  - Two forms presented/ organized differently
  - Different structure, semantics…



Aiding integration

- Usually, an integration capability is HIGHLY curated or left entirely to the end user
- If left to the user, the results is a new product which must also be managed and shared
- “I can’t integrate what I don’t understand”
- Key idea: provide for integratability !!!
  - Standards – formats for sure but also
  - Metadata
  - Semantics

Informatics considerations

- Be aware of what it means for integration to be available and how it can actually be used

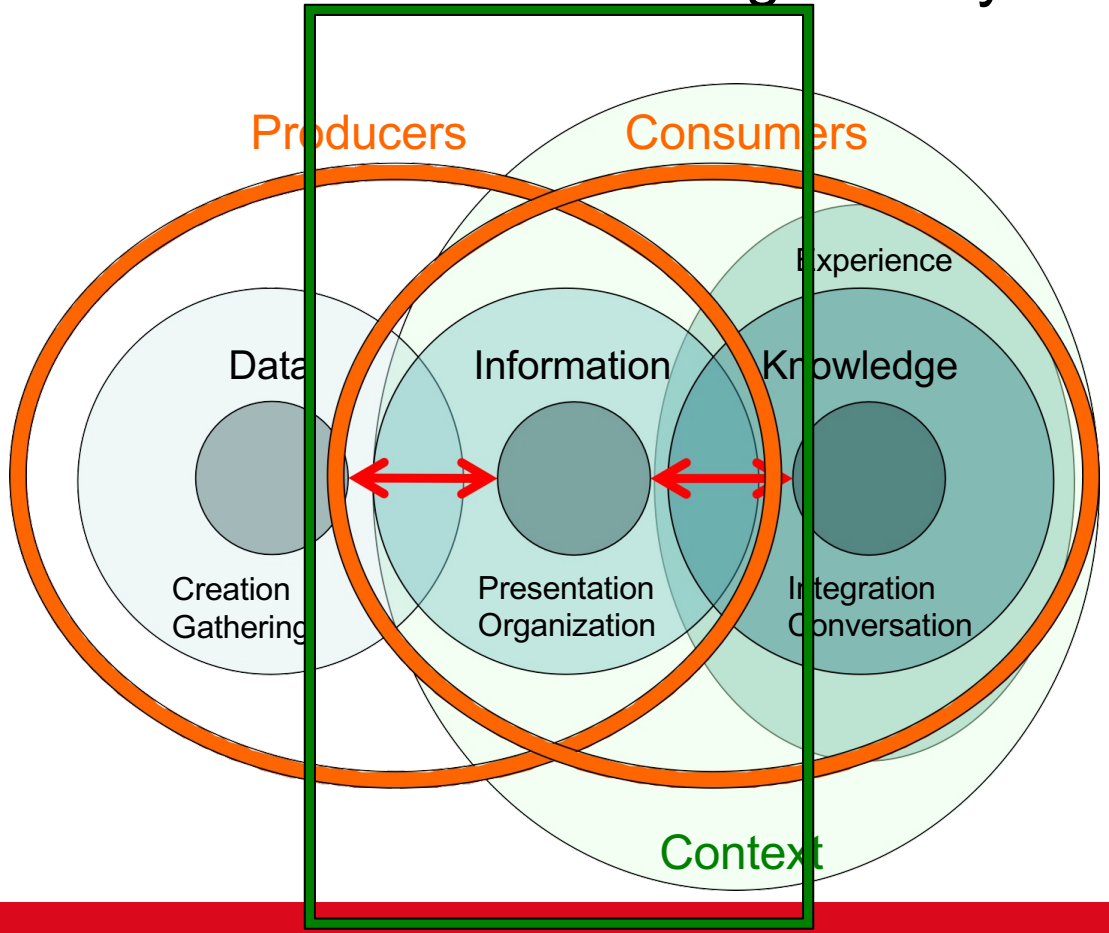
- This is more than often what leads to new findings, and abductive reasoning (one of our goals!)

Reasoning

- Deductive: top-down, start with theory, examine evidence, form definitive conclusion
- Inductive: bottom-up, start with observations, find patterns, form general conclusion
- Abductive: start with observations, estimate most likely conclusion (with uncertainty)

Information Lifecycle

Data-Information-Knowledge Ecosystem



Life cycle - definitions

- Life-cycle elements
  - **Acquisition**: Process of recording or generating a concrete artefact from the concept (see transduction)
  - **Curation**: The activity of managing the use of data from its point of creation to ensure it is available for discovery and re-use in the future (https://www.dcc.ac.uk/about/digital-curation)
  - **Preservation**: Process of retaining usability of data in some source form for intended and unintended use
  - **Stewardship**: Process of maintaining integrity across acquisition, curation and preservation

Definitions ...

- **Management**: Process of arranging for discovery, access and use of data, information and all related elements.

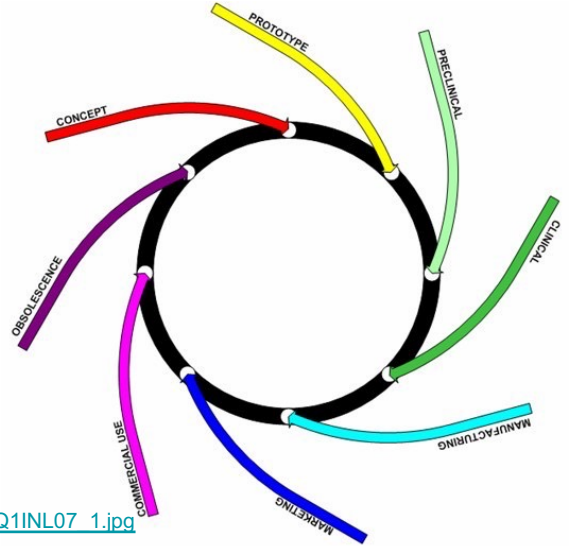


Image Credit: http://zone.ni.com/cms/images/devzone/tut/Q1INL07_1.jpg

The nature of the challenge

- To architect information systems today:
  - You may play many roles
  - **You may not get all the metadata or information you need even if you get the data**
  - You will need skills that you were not taught
- To work with end-users today
  - You may have lots of technical experience
  - You will need new skills in addressing the changing use of data and information
  - One ‘size’ does not fit all

Acquisition

- Learn / read what you can do about the means of acquisition
  - Documents may not be easy to find
  - **Bias is everywhere**!!!

- **Document** things as you go

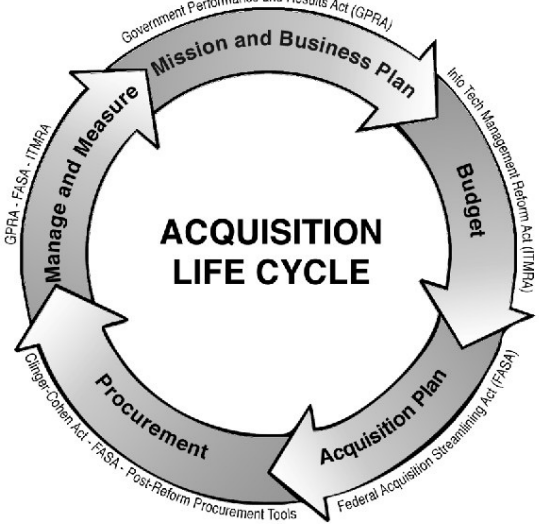


Image Credit: http://www.p2c2group.com/images/pto_lifecycle_medium.gif

Curation

- From *producer* to *consumer*!
- Consider the organization and presentation of the data as information
  - Design factors to reduce uncertainty
  - Making use of semiotics – (from last week lecture)
- Document what has been (and not) done
  - Look to add metainformation

Preservation

- ‘Archiving’ is only **one** component
  - Where are your class notes from last term?
  - This term?
- Involves steps that may not be conventionally thought of
- Think 10, 20, 50, 200 years forward. Looking historically gives *some* guide to future considerations
- …So, how would you preserve your class notes from this class?

Image Credit: http://www.planning.dc.gov/planning/lib/planning/preservation/hp_image_folder_/historic_preservation_3_x_3.jpg

Information Life Cycle

- The life cycle applies **within**, **before** and **after** your use case…

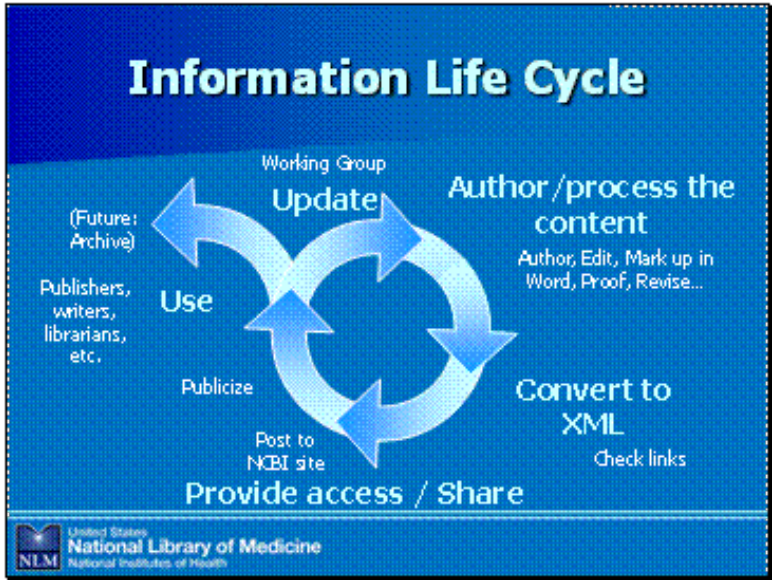


Image Credit: http://www.nlm.nih.gov/pubs/techbull/mj08/ppt/graphics/cm/image020.gif

Information Lifecycle Governance (ILG)

- **Information lifecycle governance (ILG) helps you manage your business information throughout its lifecycle from creation to deletion**.
- It automates critical data operation requirements like records management, electronic discovery, compliance, storage optimization and data migration initiatives.

Reference/Resource: https://www.ibm.com/analytics/information-lifecycle-governance?

How the information is created

- Systemic
- Environmental
- Trial-and-error (or ad-hoc)

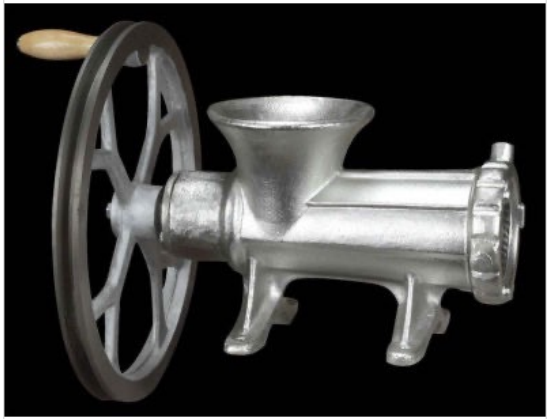
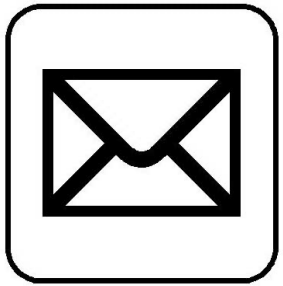


Image Credit:
http://common1.csnimages.com/lf/1/hash/2174/225740/1/Hand-Crank+Professional+Meat+Chopper+(20+lbs.+per+minute).jpg

How is information delivered?



- White paper (a document)
- Web site FAQ
- Web site informational
- Web site directed (link sent with e-mail, and so on) to a specific Web site
- One-to-one presentation:
  - Word of mouth / communication

How the information is managed

- Complexity of the information
- Complexity of the creation process
- Complexity of the management system

Complexity = Uncertainty?

- Financial impact of creation

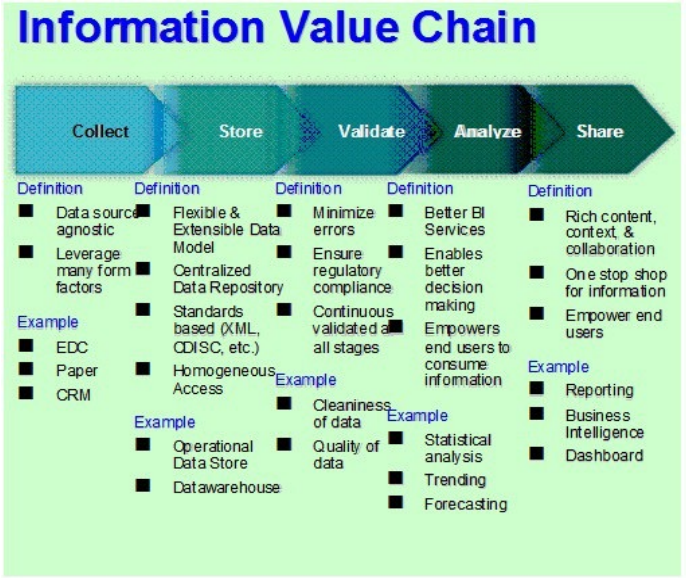
Type of information created

- Tacit (created and stored informally):
  - Human memory
  - Localize, e.g. hard drive of the computer
  - Movement of tacit information into a formalized structure
- Explicit (created and sorted formally):
  - Network shared
  - Network Web site/intranet
  - Document-management system
  - Formal Knowledge Management system

For information creation:

- Consider the
  - Value of the source
  - Age of the information
  - Source of the information, and previous interactions with that specific source

- Value of the source
- Age of the information
- Source of the information, and previous interactions with that specific source

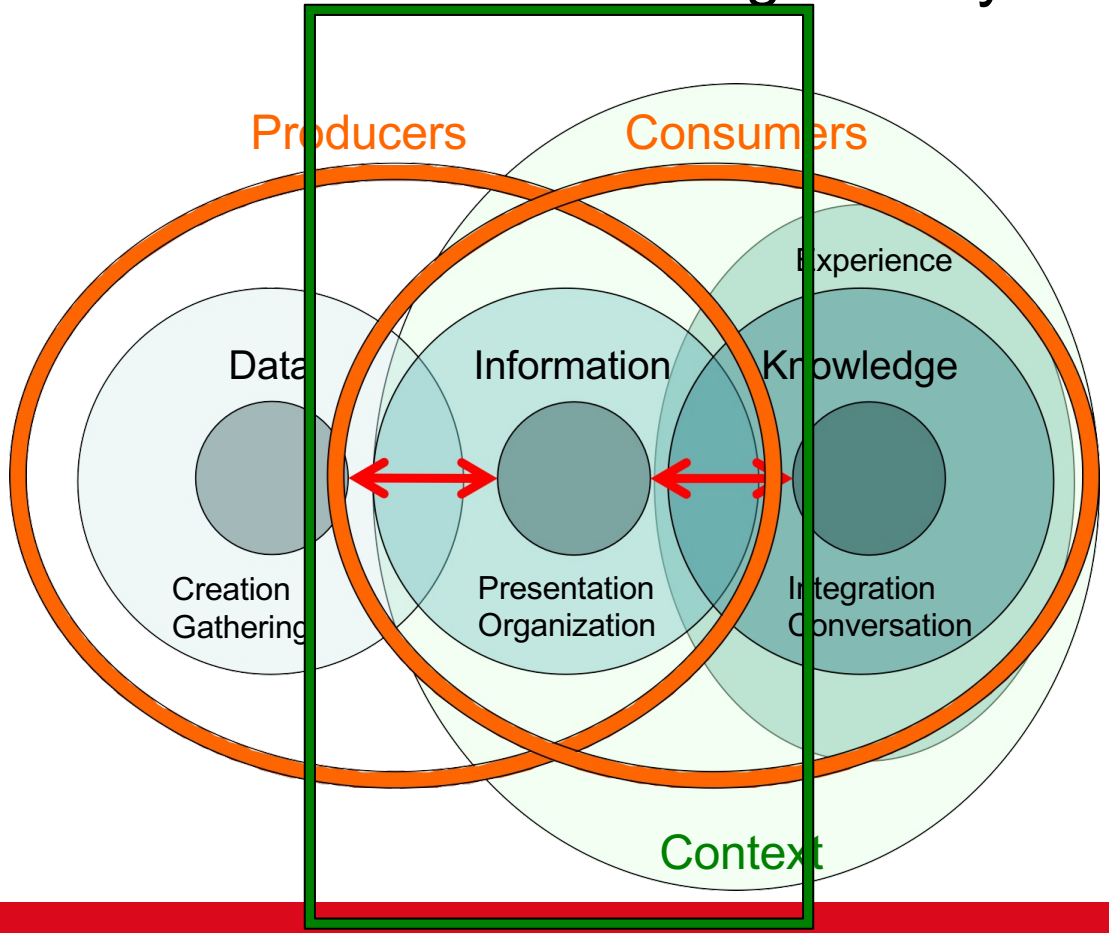


Life cycle is a complex issue

- Must be managed
- Documented
- As part of the use case, but also often outside it

Image Credit: http://www.bittbox.com/wp-content/uploads/2007/04/complex_circular_vectors.jpg

Data-Information-Knowledge Ecosystem



Information Visualization

Information Visualization

- Definition: "to form a mental vision, image, or picture of (something not visible or present to sight, or of an abstraction); to make visible to the mind or imagination" [The Oxford English Dictionary, 1989]

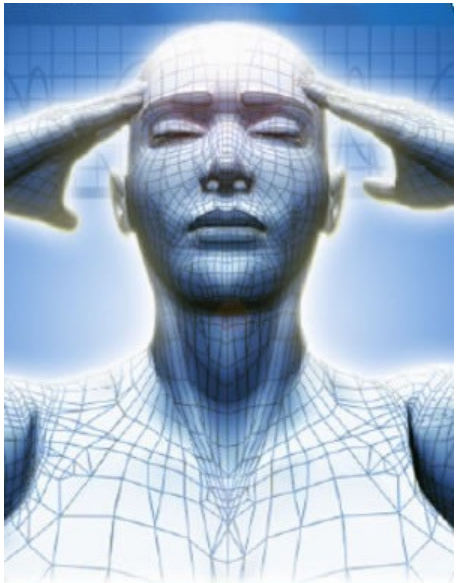
- Direct link to cognition and mental representation

- Once again, Semiotics

Image Credit: http://thestartingfive.net/wp-content/uploads/2008/02/question.jpg

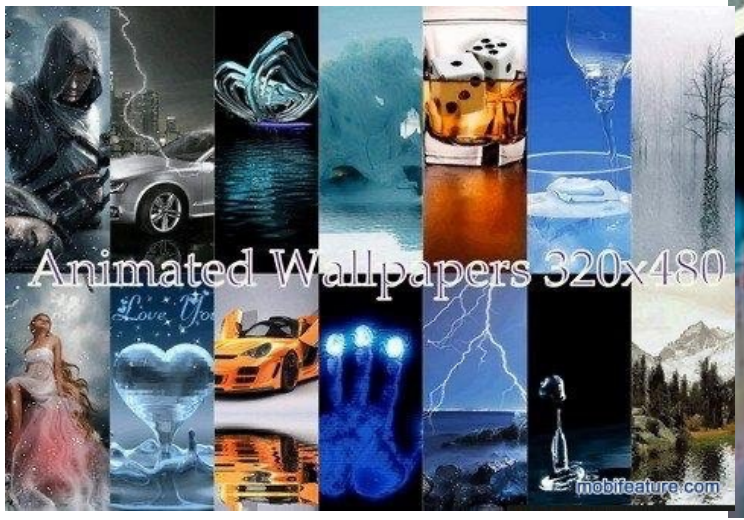
Why visualization?

- Reducing amount of data
- Patterns
- Features
- Events
- Trends
- Irregularities
- Exit points for analysis



Visualization formats

- Many – vector, raster (image), animation, multi-dimensional,



Images Credits: http://www.webmedia-solutions.com/web-development-blog/wp-content/uploads/2009/01/graphic-file-formats.jpg

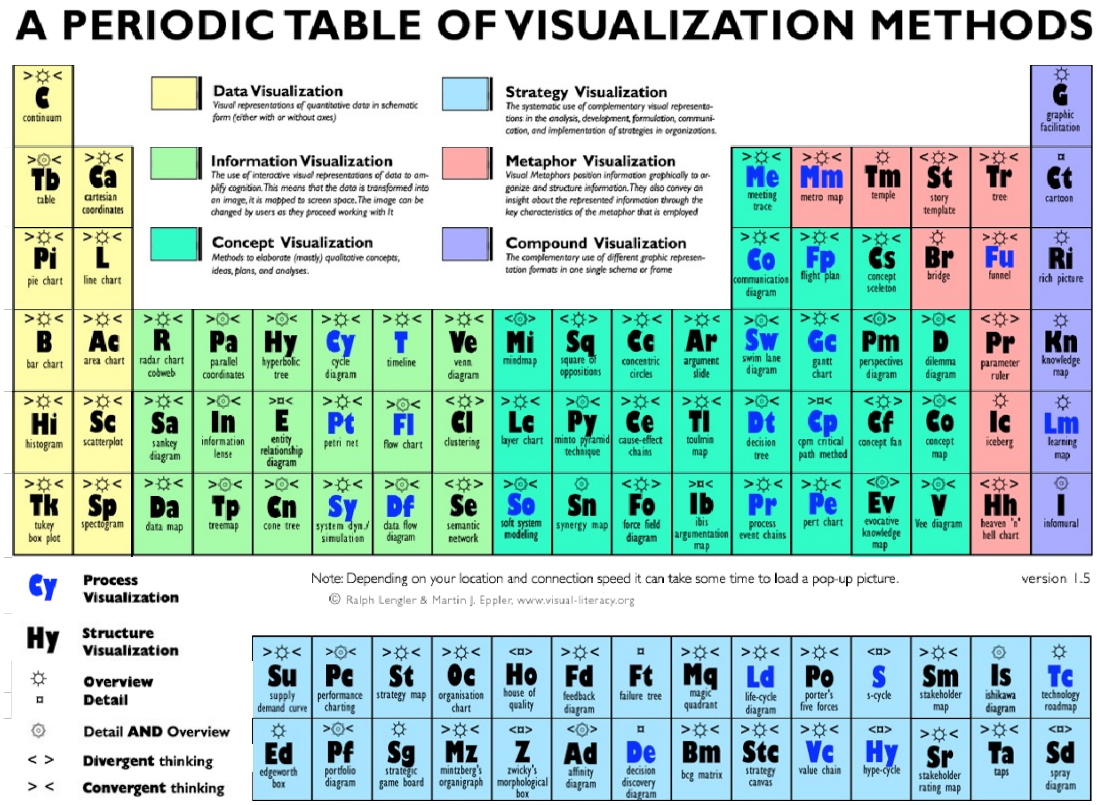
However, information, data..

- Assignment 3 - your presentations will be on semiotics and the visual representations of information systems – both good and bad
- Not just a matter of the ‘producer’ view… consider the ‘consumer’ view, i.e. what is the goal of the visualization?
- This is a time when
  - Experience helps a lot
  - But so does **listening** and gaining external **feedback**

Remember metadata!

- Many formats already contain metadata or fields for metadata, use them!
- How do you visualize Metadata?

Visualization



http://www.visual-literacy.org/periodic_table/periodic_table.html

Managing visualization products

- The importance of a ‘self-describing’ product
- Visualization products are not *just* consumed by people
- How many images, graphics files do you have on your computer for which the origin, purpose, use is still known?
- How are these logically organized?

Reference

- http://www.smashingmagazine.com/2007/08/02/data-visualization-modern-approaches/
-
- http://agbeat.com/business-marketing/piktochart-simple-infographic-creator-online-for-the-busy-professional/

- https://d3js.org

Group Project

- This is the Term Project that you are going to work on till the end of the semester
- 4-6 members in a team (No more than 6 students in a group)
- Some Ideas to Explore on:
  - Weather information
  - Disaster and Risk Management
  - Disease outbreaks
  - Early Warning Systems

Group Project

- 1 student from each group should send me the list of group members (and cc group mates) by next class Feb 26th
- Project assignment instructions on Feb 26th

Brainstorm ideas on Information System for the Group Project

- Identify the area(s) in informatics that you want to work on (Ex. Geo, Astro, Urban, Healthcare informatics…)
- Create mind maps…
- **Develop or refine a use case around a particular area of informatics that you choose as a group**
- Take Notes during the group discussions
- Use collaborative editors such as Google Doc and use a Google Drive/Dropbox/Box to share your notes, data and other resources with the group members.

- Discuss how a prototype implementation will address areas defined in lecture materials covering information uncertainty, semiotics, cognition, and architectures.
- Develop a conceptual model for the use case you chose as a group. This model should include relations among the “content” (things) and application of information theory and architecture principles (e.g. interfaces) and include diagrams

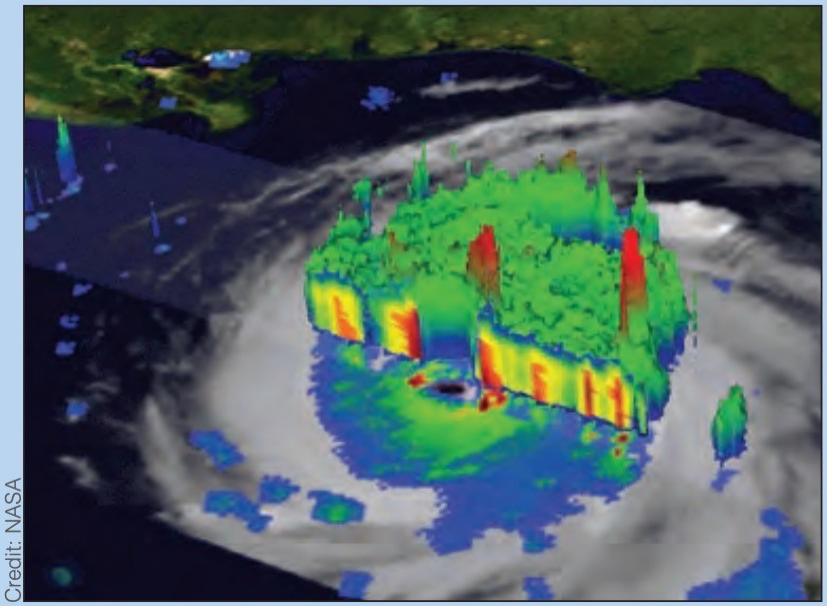
Consider following Application Areas..

- Ask Questions: How do we use the information to build a better future?

Credits (left to right): NASA, Orange County Archives, Kevin Conners/morgueFile.com, David Fine/FEMA, Beachgranny/morgueFile.com

Information on Global Precipitation

Monitoring and Predicting Hurricanes



Credit: NASA

https://pmm.nasa.gov/GPM

Image Resource: https://pmm.nasa.gov/sites/default/files/document_files/GPM%20Mission%20Brochure.pdf

Information on Weather & Climate

- Enhanced Prediction Skills for Weather and Climate

Photo credit: Ben Grader

- Improve Forecasting Capabilities for Floods, Drought and Landslides

https://pmm.nasa.gov/GPM

Image Resource: https://pmm.nasa.gov/sites/default/files/document_files/GPM%20Mission%20Brochure.pdf

Better Agricultural Crop Forecasting

- The agricultural community needs to know the timing and amount of precipitation to forecast crop yields and warn of freshwater shortages that might affect irrigation and production.
- Satellite data from the GPM mission will provide global precipitation estimates over land that can be incorporated into forecast models

Photo credit: Kevin Connors/morgueFile.com

https://pmm.nasa.gov/GPM

Image Resource: https://pmm.nasa.gov/sites/default/files/document_files/GPM%20Mission%20Brochure.pdf

Jefferson Project at RPI

Monitoring Freshwater Resources

Water resource managers rely on accurate precipitation measurements to monitor freshwater resources necessary for human activities including public consumption, irrigation, sanitation, mining, livestock and powering industries. Global observations of precipitation from the GPM constellation of satellites will allow scientist to better understand and predict changes in freshwater supply.

Photo credit: Jonah G. S.

Jefferson Project at Lake George

Image Credit: https://jeffersonproject.rpi.edu/

https://pmm.nasa.gov/GPM

Image Resource: https://pmm.nasa.gov/sites/default/files/document_files/GPM%20Mission%20Brochure.pdf

Thanks!!