Rensselaer

why not change the world?®

Introduction to Xinformatics

Ahmed Eleish

January 8th, 2025

ITWS, ERTH, CSCI 4400/6400

Tetherless World Constellation
Rensselaer Polytechnic Institute

Admin information

Class: ITWS/ERTH/CSCI - 4400/6400

- Hours: 11:00am – 01:50pm ET on Wednesdays
- Instructor: Ahmed Eleish
- Instructor contact: eleisa2@rpi.edu
- Instructor office hours: Thursday from 12:00 PM - 2:00 PM ET or by appointment/email
- Instructor office location: Amos Eaton 134
- TA: Binita Chinemerem - chineb@rpi.edu
- TA office hours: TBA
- TA office: Lally 205
- Web site: https://tw.rpi.edu/classes/xinformatics-2025
- LMS (http://lms.rpi.edu/)

Contents

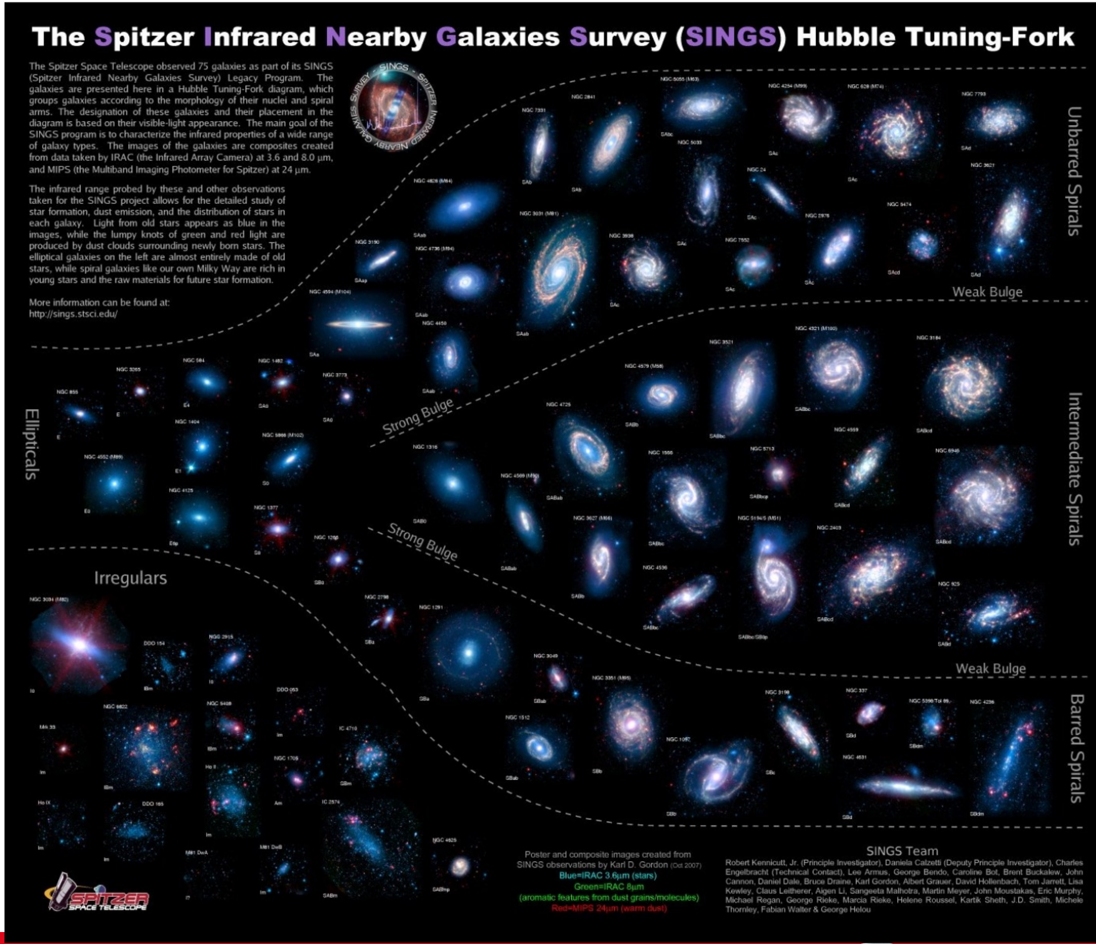
- Course Outline
- Application areas
- Assessment and assignments
- Course objectives, learning outcomes
- Introductions
- Introduction to Xinformatics
- Next modules and classes

Course Outline

- Introduction to Informatics
- Capturing the problem: Use case development and requirement analysis
- Information theory, models, tools
- Semiotics, library, cognitive and social science
- State-of-the-Art informatics applications
- Information life-cycle
- Information architectures (Internet, Web, Cloud)
- Information Visualization, Information Discovery, Information Integration
- Information Audit and Workflow
- Information quality, and bias

Application Areas

- Geoinformatics
- Astroinformatics
- Cheminformatics
- Bioinformatics
- Helioinformatics
- Healthinformatics
- Ecoinformatics



Assessment and Assignments

Assessment and Assignments

• Reading assignments

- Are given for almost every module
- Most are background and informational
- Some are key to completing assignments
- Some are relevant to the current week's class
- Others are relevant to following week's class (i.e. pre-reading)
- Summaries to be submitted by next class

– 6000 Level Students:

- are tested on these as part of assignments,

i.e. extra questions

• You will progress from individual work to group work

Current assignment structure (no final exam!)

- Assignment 1: Use case and user requirements development - 10% (written)
- Assignment 2: Information uncertainty and uncertainty reduction - 10% (written)
- Assignment 3: Analysis of cognitive, collection, and social/cultural aspects of information systems in signs - 15% (written) / 5% (presentation)
- Assignment 4: Information models and information architectures - 10% (written)
- Assignment 5: Development of use case, information architecture, model and prototype implementation for informatics area of your choice (Group Project) – 25% (written) / 10% (presentation)

- Participation: Attend class, submit summaries of readings, discuss in class – 15%
    - Short written summaries for pre-readings to be submitted by email

Objectives

- To instruct future information architects how to sustainably generate information models, designs and architectures
- To instruct future technologists how to understand and support essential data and information needs of a wide variety of producers and consumers
- For both to know tools, and requirements to properly handle data and information
- Will learn and be evaluated on the underpinnings of informatics, including theoretical methods, technologies and best practices.

Learning Outcomes

- Through class lectures, practical sessions, written and oral presentation assignments and projects, students should:
  - Develop and demonstrate skill in Participation and Organization of multi-skilled teams in the application of Informatics
  - Demonstrate the development of Conceptual and Information Models and explain them to non-experts (6400 level only)
  - Demonstrate the application of information theory and design principles to information systems
  - Evaluate, discuss and be able to apply (6400 level only) Informatics Best Practices and Standards
  - Demonstrate knowledge (advanced for 6400 level) and application of Informatics Standards
  - Develop and demonstrate (advanced for 6400 level) skill in Informatics Tool Use and Evaluation

Skills needed

- Modelling, theory, architecture experience?
  – Nah, we'll cover that

- Literacy with computers and applications that can handle information

- Ability to access internet and retrieve/ acquire data

- Presentation of assignments (both individual and group presentations)

What is expected

- Attend class, complete assignments (esp. reading, submit summaries, be prepared to discuss in class)
- Participate (e.g. reading), discuss, ask questions
  – Work constructively in group and class sessions
- And follow the assignment naming scheme

Academic Integrity

- Student-teacher relationships are built on trust. For example, students must trust that teachers have made appropriate decisions about the structure and content of the courses they teach, and teachers must trust that the assignments that students turn in are their own. Acts, which violate this trust, undermine the educational process. The Rensselaer Handbook of Student Rights and Responsibilities defines various forms of Academic Dishonesty and you should make yourself familiar with these. In this class, all assignments that are turned in for a grade must represent the student's own work. In cases where help was received, or teamwork was allowed, a notation on the assignment should indicate your collaboration.

- Submission of any assignment that is in violation of this policy will result in a penalty. If found in violation of the academic dishonesty policy, students may be subject to two types of penalties. The instructor administers an academic (grade) penalty of full loss of grade for the work in violation, and the student may also enter the Institute judicial process and be subject to such additional sanctions as: warning, probation, suspension, expulsion, and alternative actions as defined in the current Handbook of Student Rights and Responsibilities.

- Second violation will result in failure of the course.

- **If you have any question concerning this policy before submitting an assignment, please ask for clarification.**

Questions so far?

Peter Arthur Fox, Ph.D.
1959 - 2021

Introductions

- Who you are
- Why you are here
- What you expect to learn
- Your interests/hobbies

Introduction to Informatics

- **E.g. Bioinformatics**

– Over the past few decades, major advances in the field of molecular biology, coupled with advances in genomic technologies, have led to an explosive growth in the biological information generated by the scientific community. This deluge of genomic information has, in turn, led to an absolute requirement for computerized databases to store, organize, and index the data and for specialized tools to view and analyze the data.

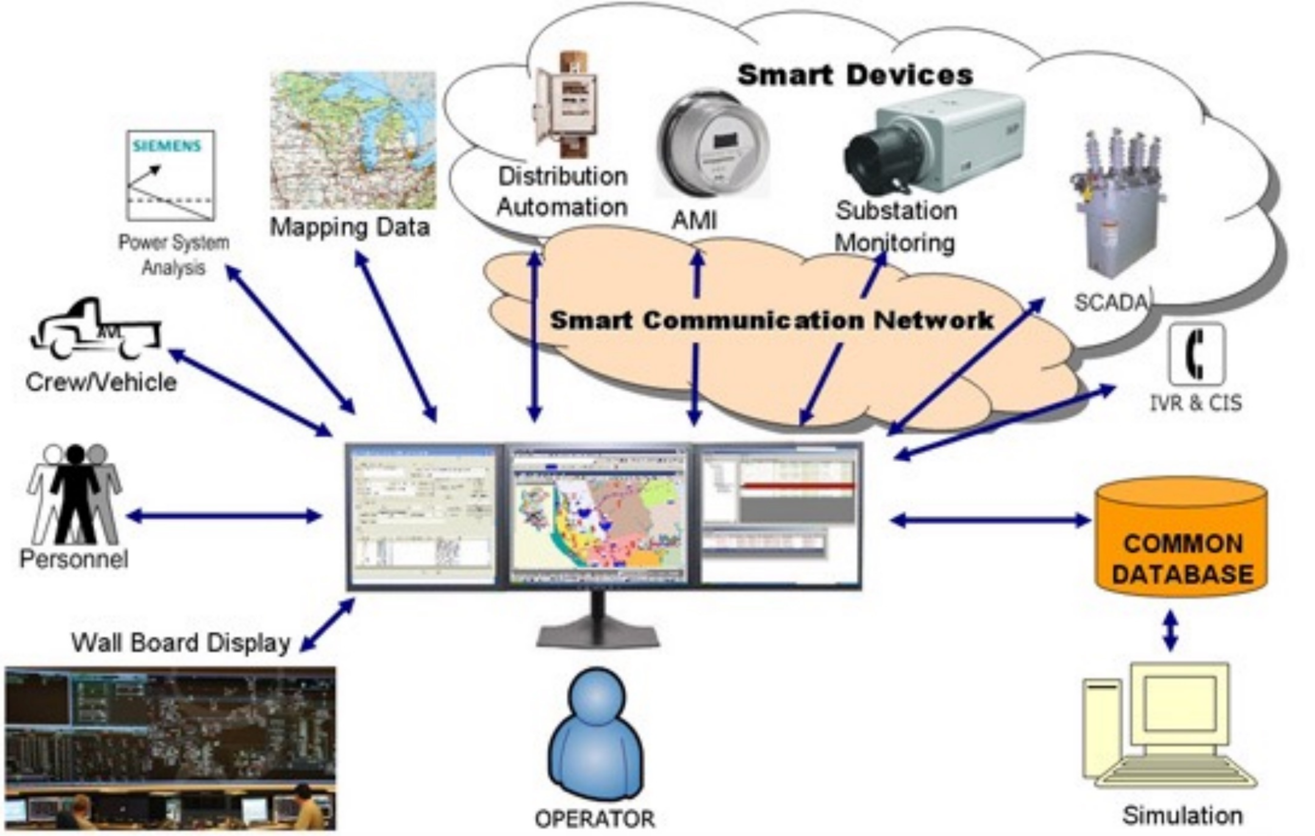
– What is informatics? (What is Bioinformatics)

– Read: https://www.amia.org/fact-sheets/what-informatics

Tell us more…

- Bioinformatics is the field of science in which biology, computer science, and information technology merge to form a single discipline.
- The ultimate goal of the field is to enable the discovery of new biological insights as well as to create a global perspective from which unifying principles in biology can be discerned.
- At the beginning of the "genomic revolution", a bioinformatics concern was the creation and maintenance of a database to store biological information, such as nucleotide and amino acid sequences.
- Development of this type of database involved not only design issues but the development of complex interfaces whereby researchers could both access existing data as well as submit new or revised data.

Data as Infostructure



Definitions

- Data - are encodings that represent the qualitative or quantitative attributes of a variable or set of variables.

- Data (plural of "datum", which is seldom used) - are typically the results of measurements, computations, or observations and can be the basis of graphs, images of a set of variables.

- Data - are often viewed as the lowest level of abstraction from which information and knowledge are derived***

Definitions

- Information

– Representations (of data) in a form that lends itself to human use

– data in context

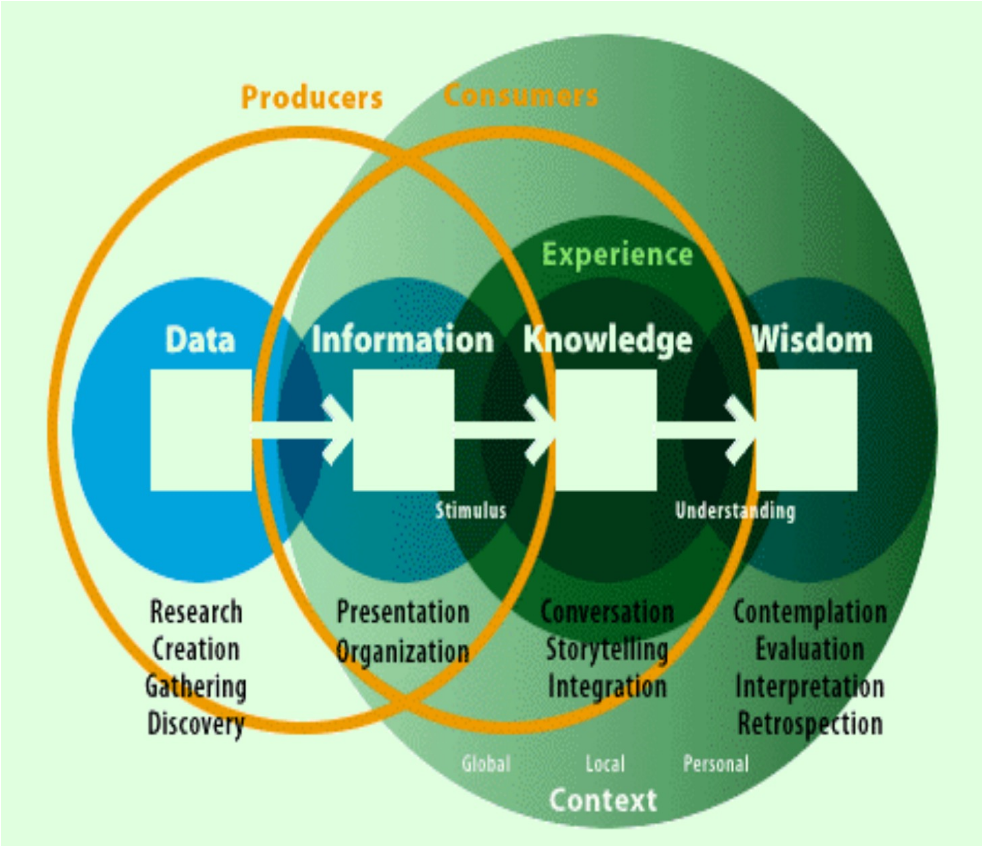
– The word information derives from the Latin informare (in+formare) meaning to give form, shape, or character to. It is therefore to be the formative principle of, or to imbue with some specific character or quality.

Definitions

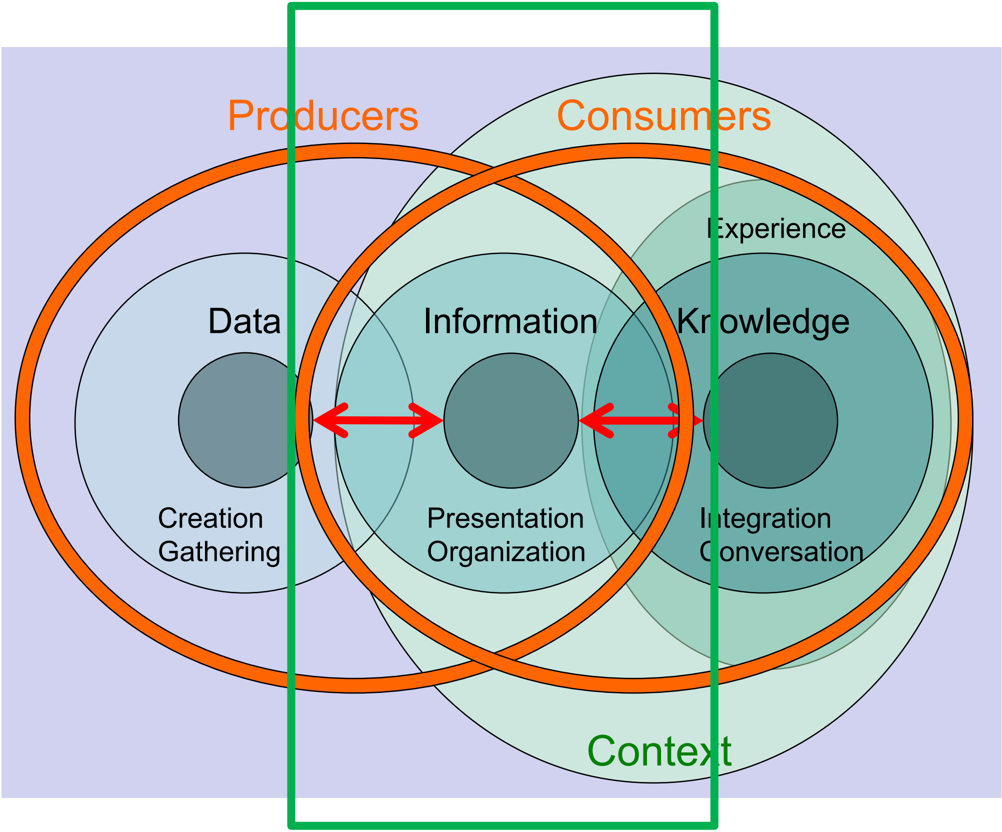
- **Metadata** – data about data
- Metainformation – information about information
- **Documentation** – integrated collection of information and metadata intended to support all aspects of data (find, access, use…)

Provenance - origin or source from which something comes, intention for use, who/what generated for, manner of manufacture, history of subsequent owners, sense of place and time of manufacture, production or discovery, documented in detail sufficient to allow reproducibility

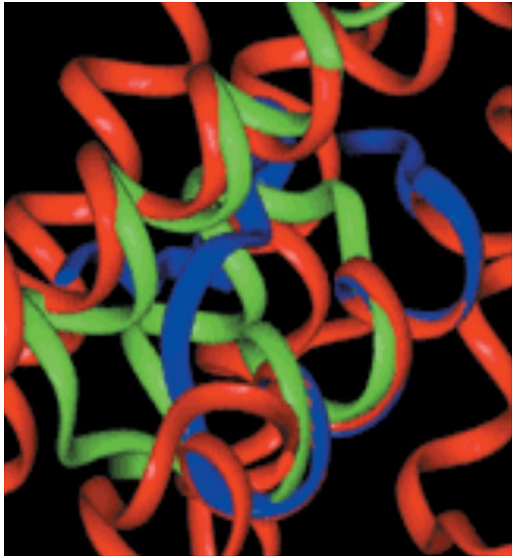
Could it really be this linear?



Data-Information-Knowledge Ecosystem

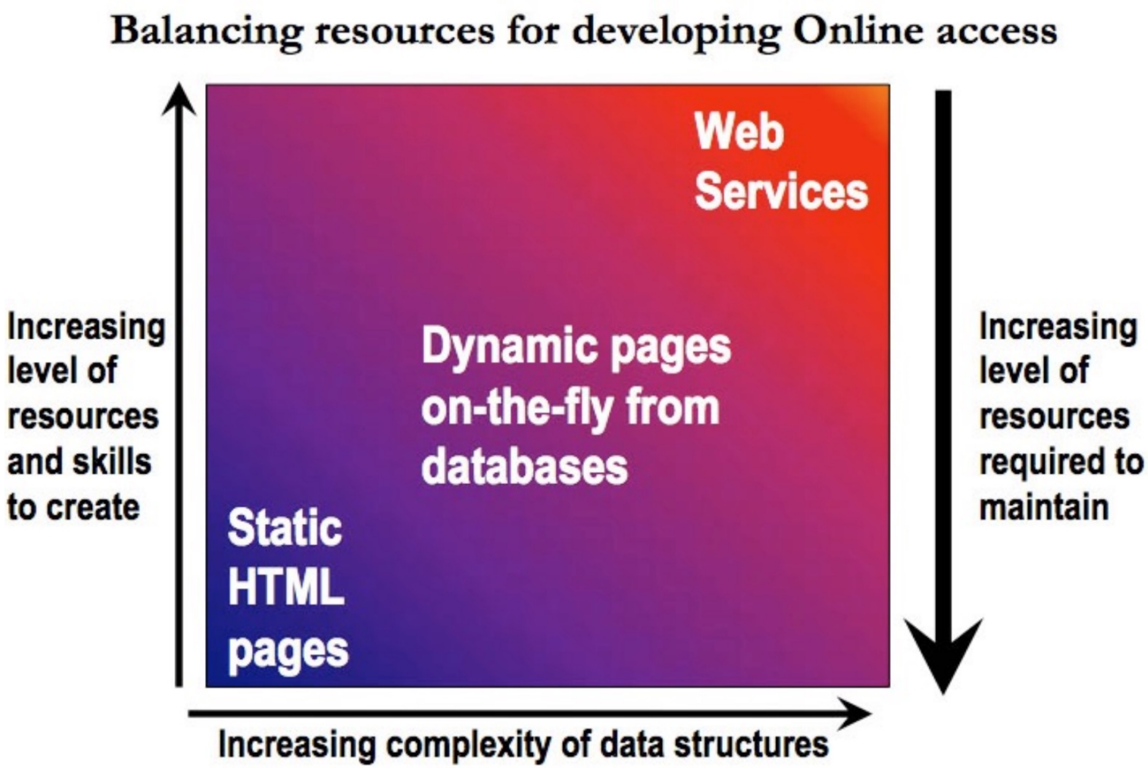


The Information Era: Interoperability



- Modern information and communications technologies are creating an "interoperable" information era in which ready access to data and information can be truly universal.
- Open access to data and services enables us to meet the new challenges of understanding complex systems:
  - managing and accessing large data sets
  - higher space/time resolution capabilities
  - rapid response requirements
  - data assimilation into models
  - crossing disciplinary boundaries.

Shifting the Burden from the User to the Provider



Other forms of information

IoT!

Yottabytes

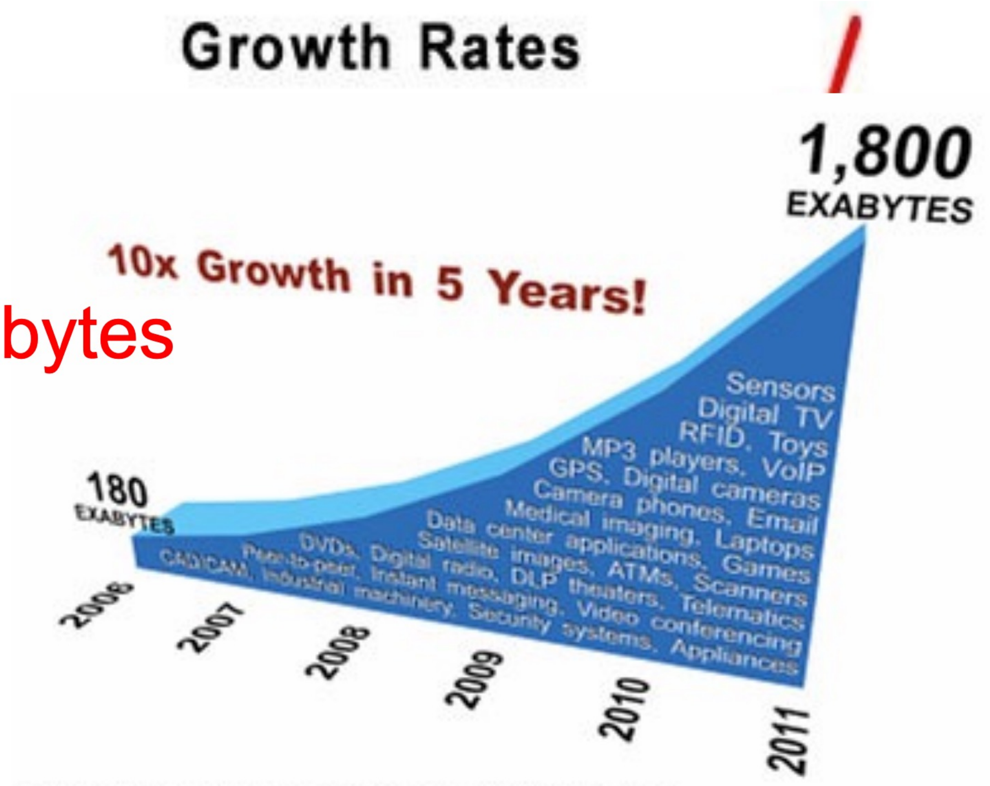
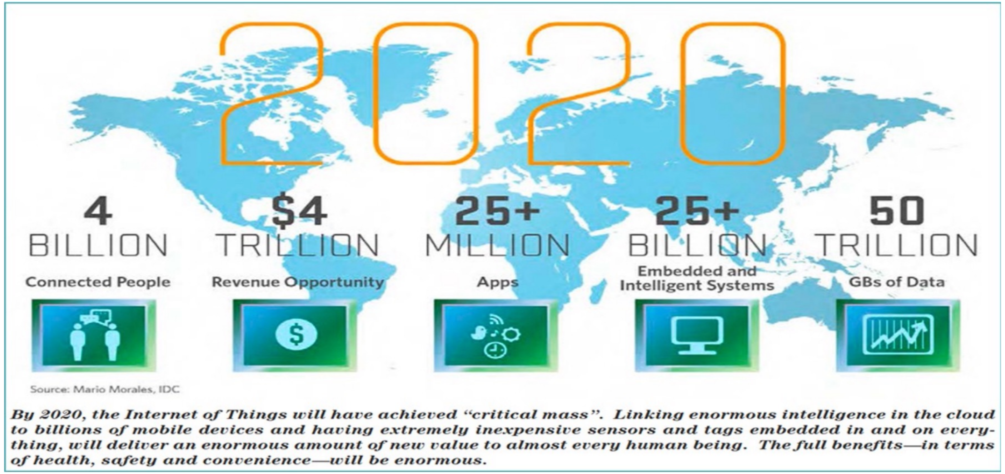
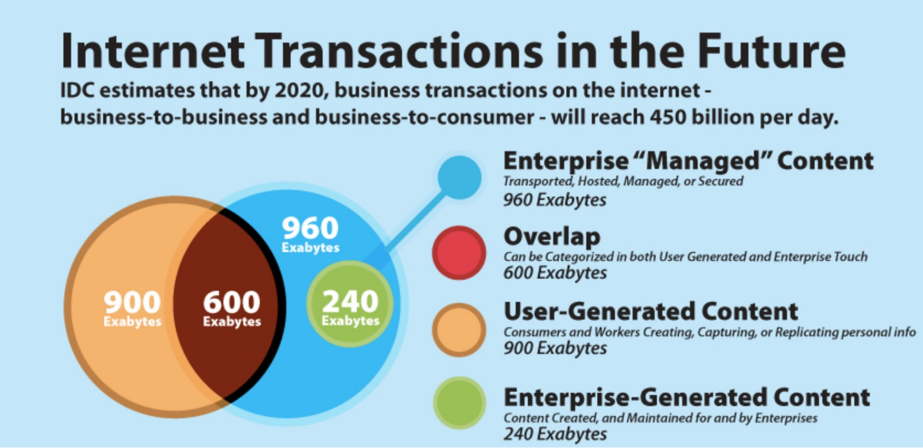


Figure 1

Information explosion





The key is

- As volume, complexity and heterogeneity increase…
– Suddenly information may look more like a continuum
– All known methods, algorithms will not scale (except for very simple operations)
– And because it is information, humans are part of the loop

- Thus – we need to understand and apply the theoretical foundations

- Problem: most theory to date are developed in an analog world, not a digital one!!

Mind the gap

- As capabilities and needs grow on both sides: science/ medicine/ engineering – and technology:
    - There is still a gap between science and the underlying infrastructure and technology that is available

- Cyberinfrastructure is the new research environment(s) that support advanced data acquisition, data storage, data management, data integration, data mining, data visualization and other computing and information processing services over the Internet.

Mind the gap

- Informatics - information science includes the science of (data and) information, the practice of information processing, and the engineering of information systems.
- Informatics studies the structure, behavior, and interactions of natural and artificial systems that store, process and communicate (data and) information. It also develops its own conceptual and theoretical foundations. Since computers, individuals and organizations all process information, informatics has computational, cognitive and social aspects, including study of the social impact of information technologies.

A moment of history

- In the late 1950’s (actually around 1957-1958 or 1962 depending on what you read) the modern informatics term was coined
- Existed for a while but then split into library science and computer science and developed their own fields, became disconnected
- Now coming back to be relevant to science+…
- Informatics IS NOT just having a someone work with an “IT” person (NOT, NOT, NOT)

Core informatics

- The realm of computer science (for the most part, also librarians)
- Strongly influenced by science (and engineering)
- Must work with the scientists, sustainably

Science Informatics

- Where science meets the underlying technical capabilities and methods
- Must be expressible in science terms; use cases
- The people in this area are multi-lingual and both interdisciplinary and multi-disciplinary
- Team work, or really a community of practice (CoP)

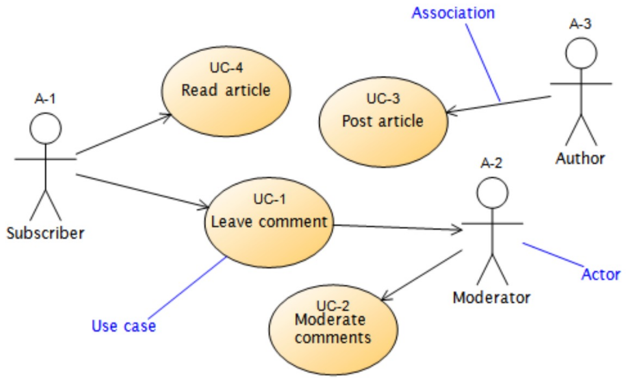
Use Case

… is a collection of possible sequences of interactions between the system under discussion and its actors, relating to a particular goal.

– The collection should define all system behavior relevant to the actors to assure their goals will be carried out properly.

– Any system behavior that is irrelevant to the actors should not be included in the use cases.

- is a prose description of a system's behavior when interacting with the outside world.
- is a technique for capturing functional requirements of business systems and, potentially, of an IT system to support the business system.



Information theory

- Semiotics, also called semiotic studies or semiology, is the study of sign processes (semiosis), or signification and communication, signs and symbols, into three branches:

– Syntactics: Relation of signs to each other in formal structures

– Semantics: Relation between signs and the things to which they refer; their denotation

– Pragmatics: Relation of signs to their impacts on those who use them

Library science

- Curates the artifacts of knowledge but increasingly: (yes) information
- Organizes and manages them for consumers

    - Cataloging and classification

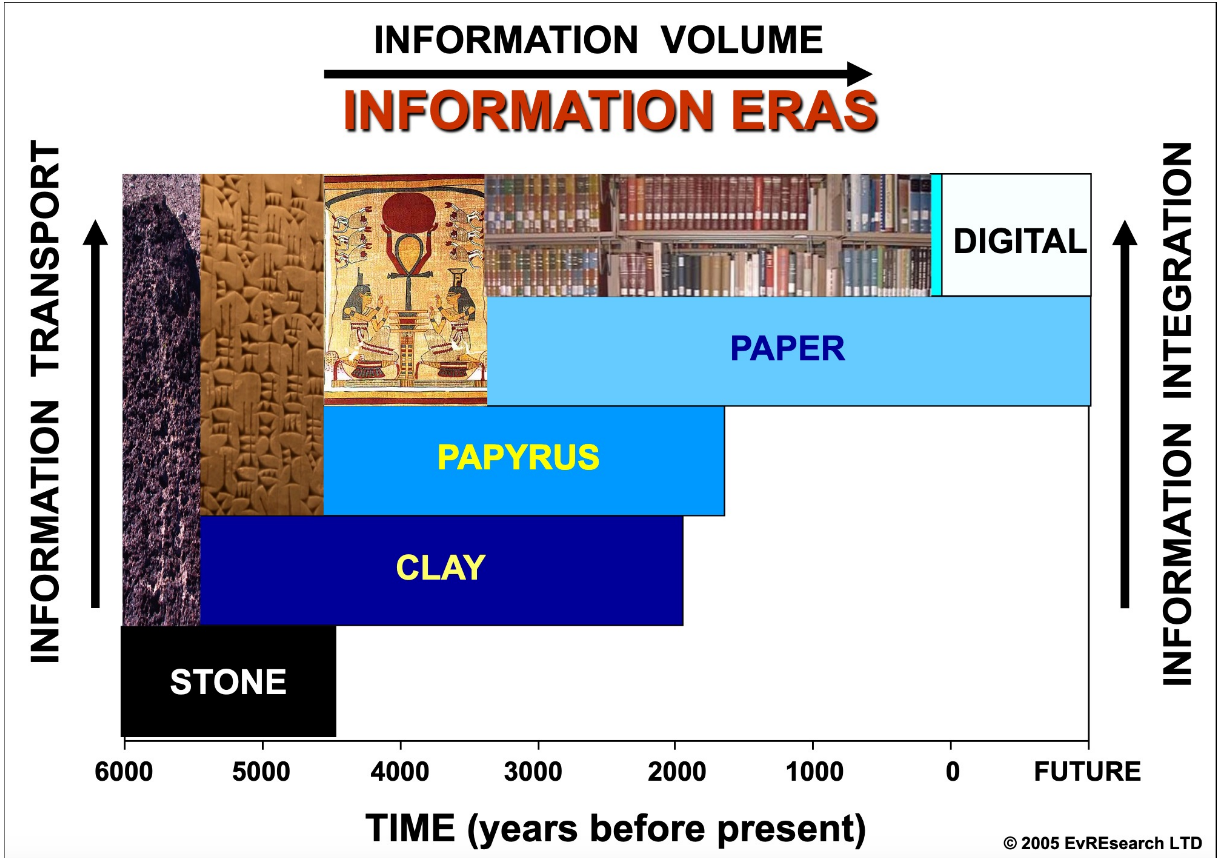
- Preservation

    - 'maintaining or restoring access to artifacts, documents and records through the study, diagnosis, treatment and prevention of decay and damage'

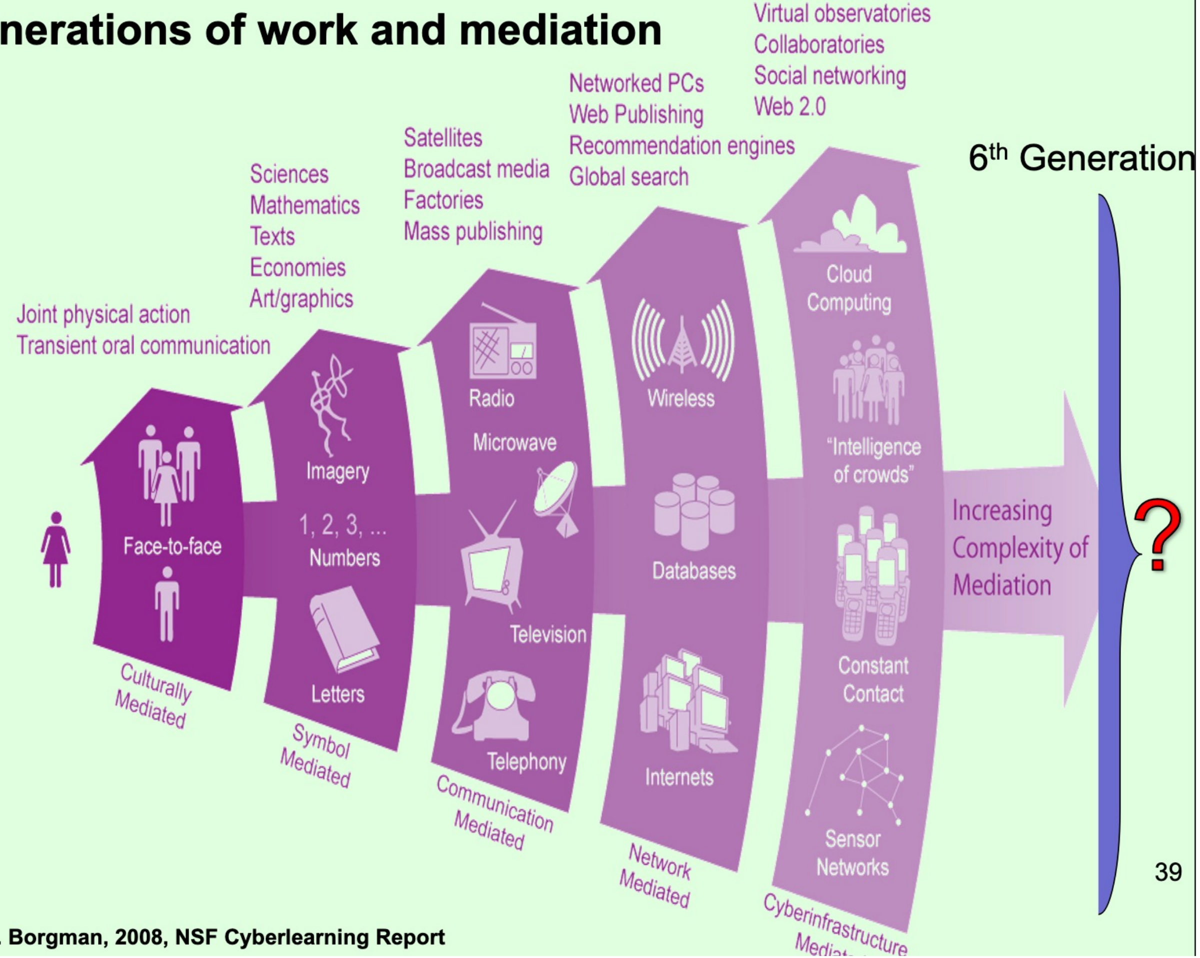
- In our digital age

    - Curation and preservation

HISTORY OF INFORMATION THRESHOLDS



5 generations of work and mediation



From: C. Borgman, 2008, NSF Cyberlearning Report

Social Science

- Branch of humanities
- – Especially as it relates to networks of people
- Exploits sociology of groups, teams
- Cultural norms as well as discipline norms

Cognitive Science

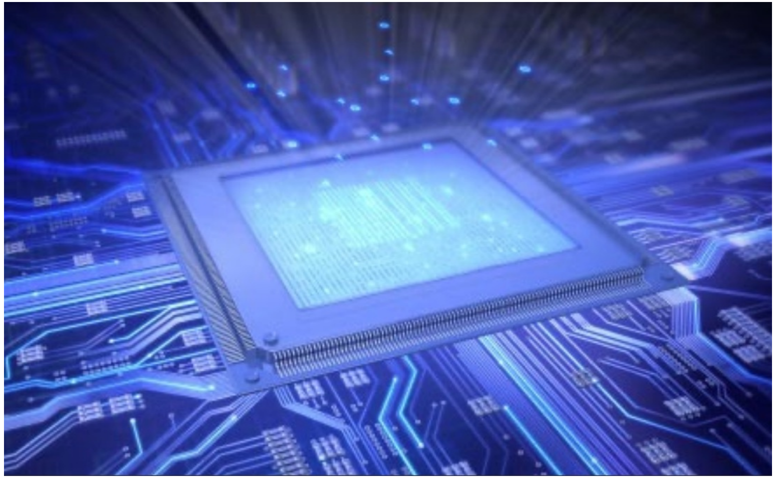
- Interdisciplinary study of the mind and intelligence

- operates at the intersection of psychology, philosophy, computer science, linguistics, anthropology, and neuroscience.

- Of relevance for data and information science are three significant theoretical underpinnings
    – mental representation,
    – the nature of expertise,
    – and intuition
- Very relevant to models, modeling, metamodel choice

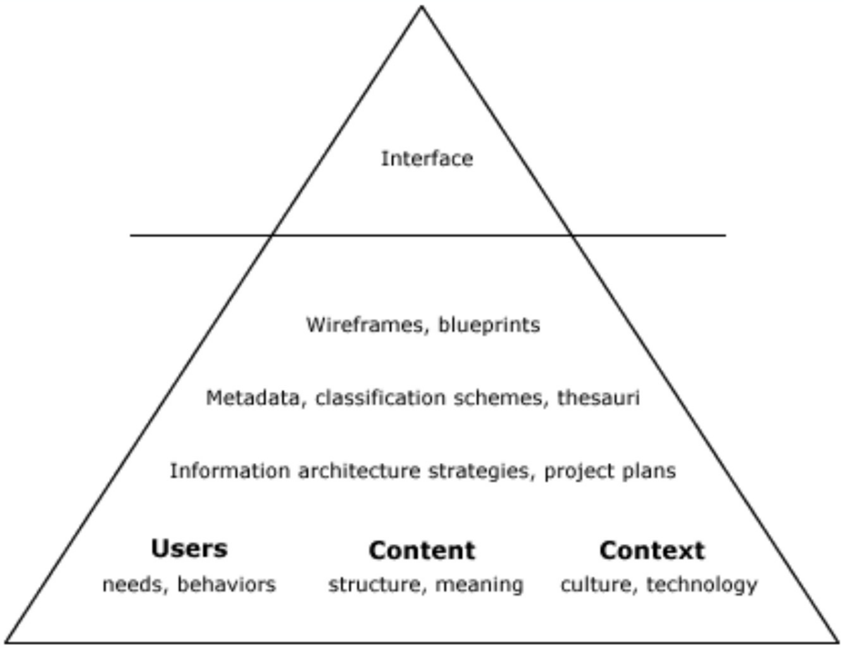
Computer Science

- Data Structures
- Knowledge/ Information
- Representation
- Algorithms!! Encoding/ coding
- More…

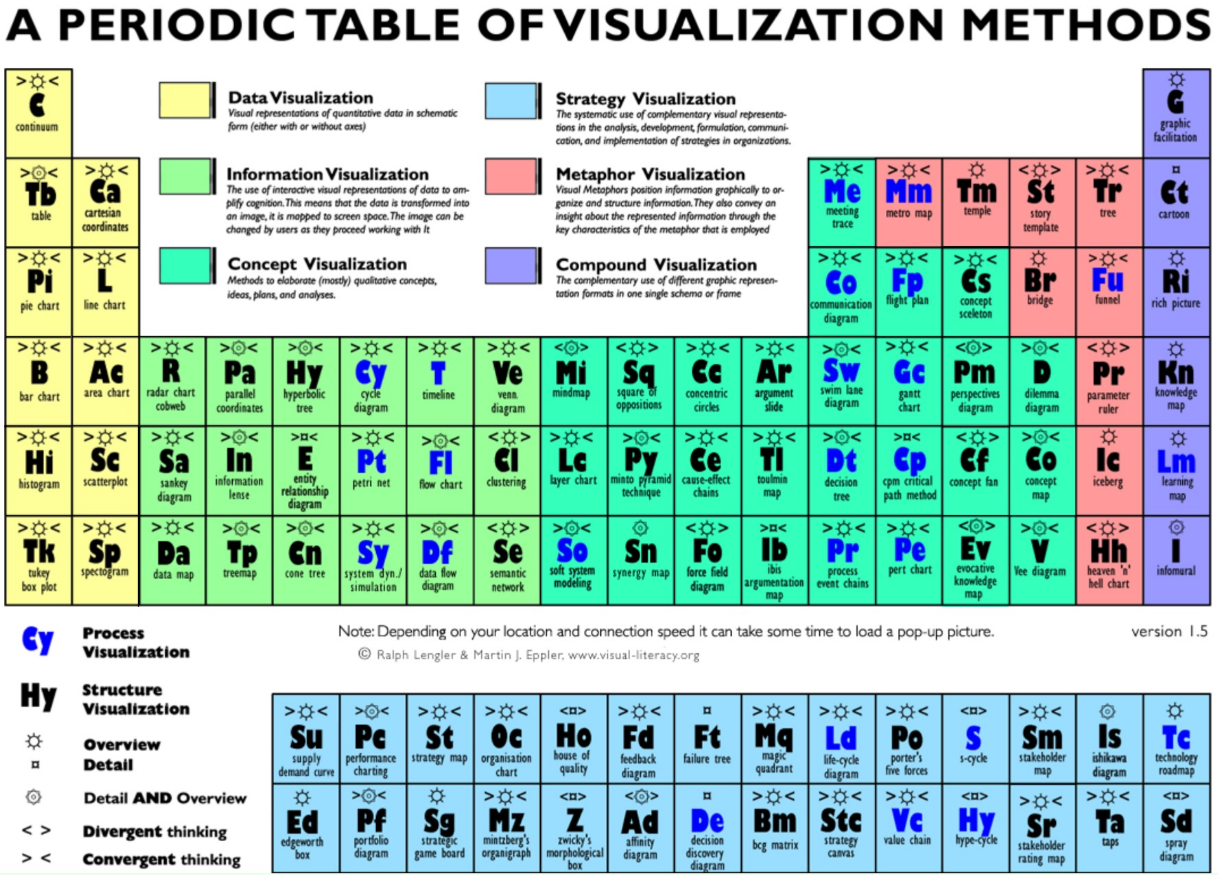


Information Architectures

- Building on content, context, and users, some illustrate information architecture as an iceberg.
- Just like an iceberg, the majority of information architecture work is out of sight, "below the water."
- The work includes the creation of plans, controlled vocabularies, and blueprints all before any user interfaces are created.



Visualization



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

Information Audit

- Analysis and evaluation of a organization's information system (whether manual or computerized) to detect and rectify blockages, duplication, and leakage of information.
- The objectives of this audit are to improve accuracy, relevance, security, and timeliness of the recorded information.
- Related to workflow management

Information Quality, and Bias

- Quality is perceived differently by information providers and information recipients
- There are many different qualitative and quantitative aspects of quality
- Methodologies for dealing with qualities are just emerging
- Little resources are allocated to quality
- Each organization handles quality differently
- Bias detection is often an expert skill, and bias correction even more-so

Reading (Prior to next week)

- BioInformatics: http://www.wired.com/wiredscience/2008/07/researchers-tra/
- Biomedical informatics for proteomics http://www.nature.com/nature/journal/v422/n6928/full/nature01515.html?free=2
- Clinical Research Informatics: https://jclinbioinformatics.biomedcentral.com/articles/10.1186/2043-9113-5-S1-A1
- Urban Informatics: http://www.creativeclass.com/_v3/creative_class/2009/04/09/now-emerging-urban-informatics/
- Geo-Informatics: https://en.wikipedia.org/wiki/Geoinformatics
- Astro-Informatics: https://asaip.psu.edu/Articles/astroinformatics-in-a-nutshell
- Use Case: https://en.wikipedia.org/wiki/Use_case
- What is Biomedical & Health Informatics? : https://brand.amia.org/m/3cb085297670d4a9/original/What-is-Informatics.pdf

Reading (Prior to next week)

- Read AND summarize: Use Case + 3 out of 7 other readings.
- SHORT summaries, 2-3 paragraphs
- Submit summaries by email before next class (part of your participation grade)

Next week

- Capturing the problem: use cases and requirements as-a basis for Assignment 1

Thanks!