

# Is there a theory of geographic information?

**Keith C. Clarke**

Professor, Geography, UCSB and  
Visiting Leverhulme Professor, Dept. of  
Information Science, City University, London

# The Leverhulme Trust

- Trust of William Hesketh Lever, first Viscount Leverhulme
- Gives awards for the support of research and education
- Emphasizes individuals and encompasses all subject areas, including Geography
- One goal is the removal of barriers between traditional disciplines
- <http://www.leverhulme.org.uk>



# Why is theory important?

- GISystems has been somewhat technically oriented
- Technology/applications have led research
- Science and technology are highly interrelated
- In many other sciences too, observation leads theory
- GIScience came after GISystems
- Theory can increase the perceived and actual value of Geography to other disciplines

# Theory and Observation/Measurement

- All observations are theory-impregnated. **There is no pure, disinterested, theory-free observation.**”
  - Science often creates theory well before application (e.g. Feynman’s brick, String theory), followed by exhaustive testing (**Observation free theory**)
  - Geographic theory that does not involve geometry or topology is outside of GIScience, (e.g. social theory borrowed from philosophy or sociology), and so will not be directly considered here
- 
- Karl Popper: pages 8-9, [\*The Myth of the Framework\*](#), 1994
  - Ideas from *Against Method: Outline of an Anarchistic Theory of Knowledge* Paul Feyerabend, 1975

# Feynman's Brick



- "I'll try to answer the professor's question if you will first answer a question from me, so I can have a better idea of what 'essential object' means. Is a *brick* an essential object?" What I had intended to do was to find out whether they thought theoretical constructs were essential objects. The electron is a *theory* that we use; it is so useful in understanding the way nature works that we can almost call it real. I wanted to make the idea of a theory clear by analogy. In the case of the brick, my next question was going to be, **"What about the *inside* of the brick?" -- and I would then point out that no one has ever seen the inside of a brick.** Every time you break the brick, you only see the surface. That the brick has an inside is a simple theory which helps us understand things better. The theory of electrons is analogous."
- Richard P. Feynman (1985) "Surely You're Joking, Mr Feynman!" Bantam / W.W. Norton Company, Inc.

# Wikipedia View of Geographic Information

“Geographic information is created by manipulating [geographic](#) (or spatial) data (geodata) in a computerized system”

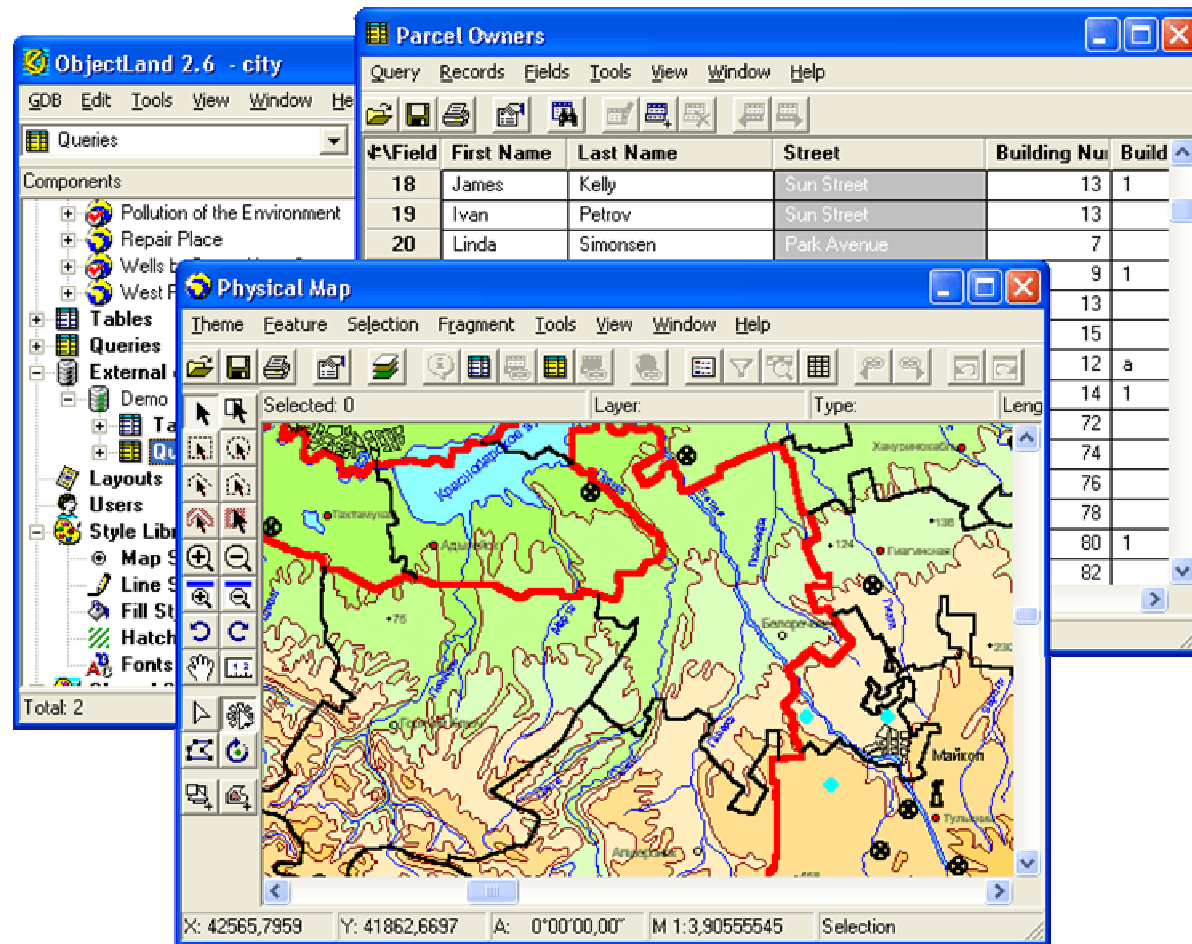
Implies:

- Can be created from nothing
- **Digital representation of space**
- “Manipulation”

# Geographical Representation (Goodchild et al., 2006)

- Necessary geometry and support
- Objects vs. Fields
- Points, Lines, Areas, Volumes
- Nominal, ordinal, interval, ratio
- Goodchild has argued for geographical atoms and di-poles
- Atom stated as the tuple:  $\langle \mathbf{x}, Z, z(\mathbf{x}) \rangle$
- Where  $x$  is a point in space-time,  $Z$  is a property and  $z(x)$  defines the property value at the point
- Objects vs. fields lie on continuum in  $m$ -dimensional space defined by properties

# GIS seems happy with its various representations



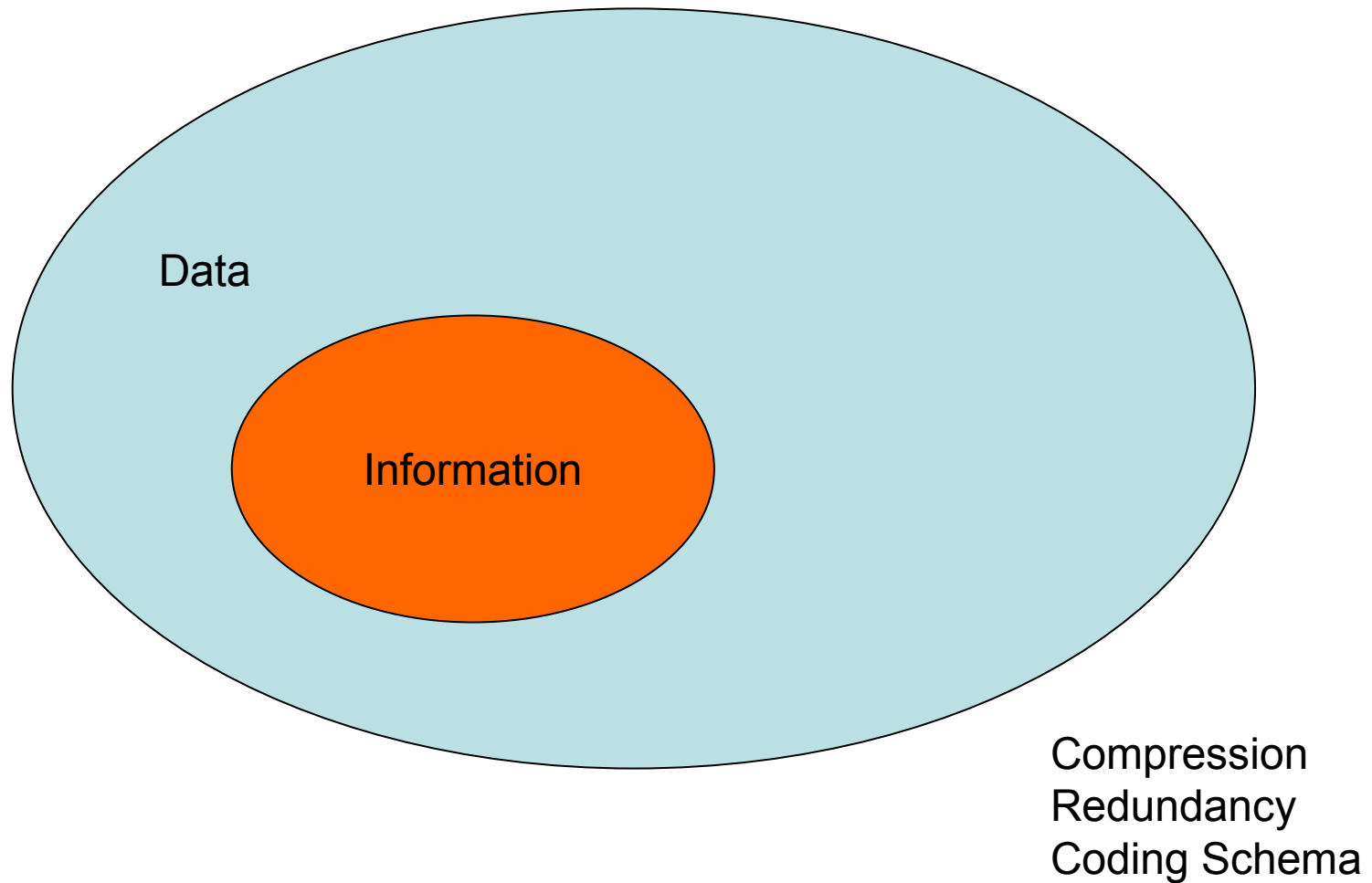
<http://www.gis-objectland.com/>

# Geographic information

- What is it?
- How can it be measured?
- What is its form and substance?
- Can it be created and destroyed?
- Does it have meaning and value beyond a single purpose?
- What does it cost?
- Is it extractable from a map?
- How can it be stored and transferred?

	0	1	00000	
	1	1	00001	
	2	2	00010	
	3	2	00011	
	4	3	00100	
decimal	5	3	00101	
number	6	3	00110	
	7	3	00111	
	8	4	01000	
	9	4	01001	
	10	4	01010	
	11	4	01011	
	12	4	01100	
	13	4	01101	bits
	14	4	01110	
	15	4	01111	
	16	5	10000	
	32	6	100000	
	64	7	1000000	
	128	8	10000000	
	256	9	100000000	
	512	10	1000000000	
	1024	11	10000000000	
	2048	12	100000000000	
	4096	13	1000000000000	
	8192	14	10000000000000	
	16384	15	100000000000000	
	32768	16	1000000000000000	
	65536	17	131071	
	131072	18	262143	
	262144	19	524287	
	524288	20	1048575	
	1048576	21	2099151	
	2099152	22	4194303	
	4194304	23	8388607	
	8388608	24		

# The simple model



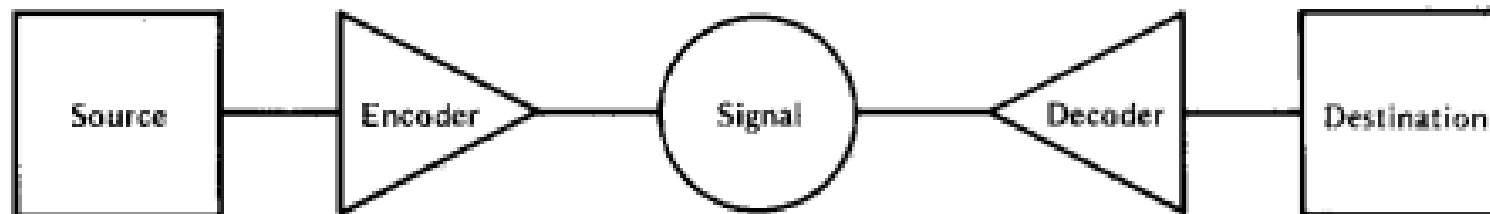
# Information Theory



**“Claude Shannon's clever electromechanical mouse, which he called Theseus, was one of the earliest attempts to "teach" a machine to "learn" and one of the first experiments in artificial intelligence.” Source: Bell Labs**

# Information channel

- **Claude Shannon's Communications Model**
- C. E. Shannon: *A mathematical theory of communication*. Bell System Technical Journal, vol. 2 pp. 379–423 and 623–656, July and October, 1948.



**Source:** *The Process and Effects of Mass Communication*. Ed. by Wilbur Schramm. Copyright © 1965 by the University of Illinois Press. By permission.

# Shannon's premise

“The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point.” (Shannon, 1948)

# The Quantity of Information

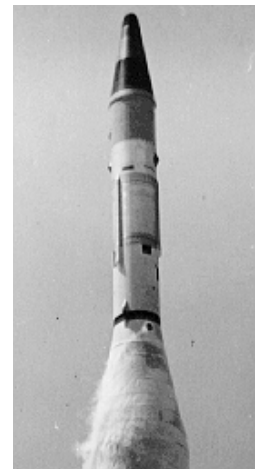
- Assume binary states at source and recipient
- Amount of information is the latency or persistence of entropy in the signal

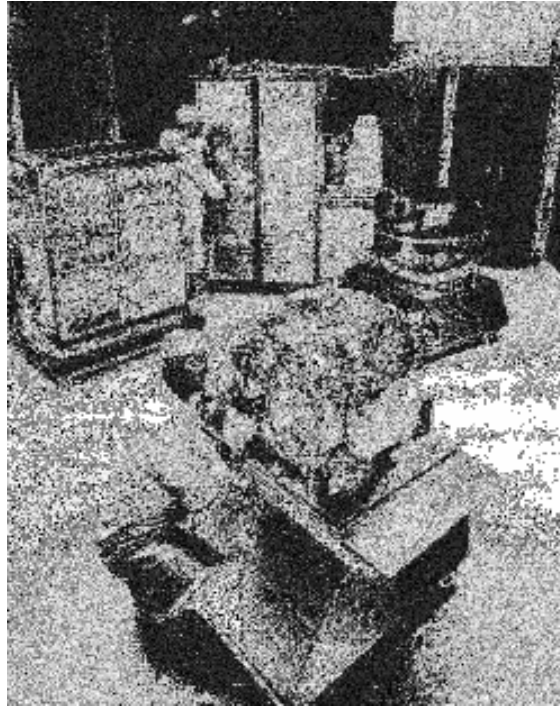
Shannon's entropy equation:

$$H(X) = - \sum_{i=0}^{N-1} p_i \log_2 p_i$$

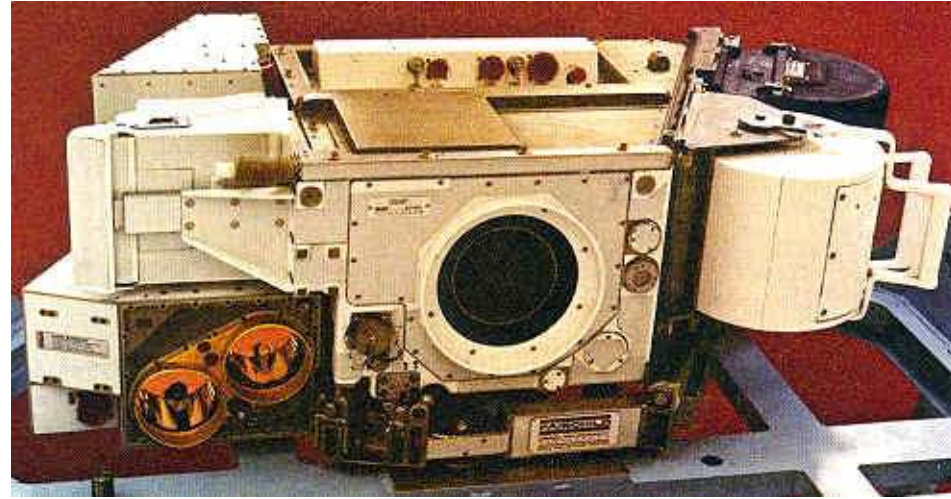
# Channels of Geographic Information

- 1950s RAND project FEEDBACK
- Implemented as SAMOS
- Goals was to reduce remotely sensed images into an analog radio signal, transmit it from satellite to earth, then reconstruct it
- Program went “Deep Black” mid 1950s





# SAMOS

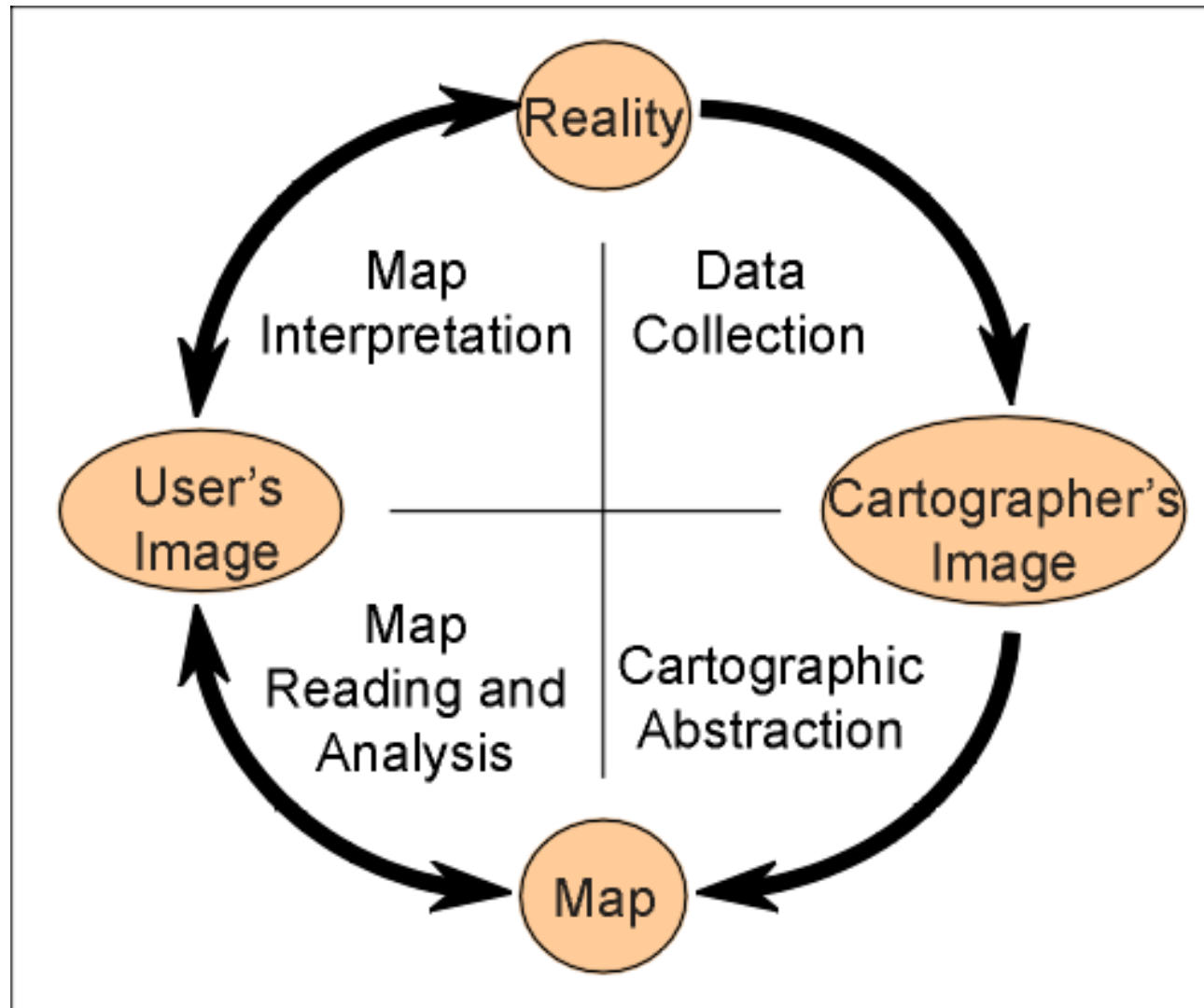


- 12 launches, 3 readout 9 recovery
- Resolutions 100, 20, 5 and 8 feet
- Film developed and scan converted
- 5 days to 4 months missions
- 1958-1961 then cancelled
- 1966-7 went to the Moon as the Apollo Lunar Mapping Camera
- Prototype for next generation systems

# Cartographic Communication

- Shannon model influenced cartography directly
- 1960s/70s Communications school
- Framed by Robinson & Petchenik 1976
- Source-channel-recipient
- “World” as input, mental model of world as decoded signal (Muehrcke)
- Heritage link to SAMOS, SAGE, cryptology

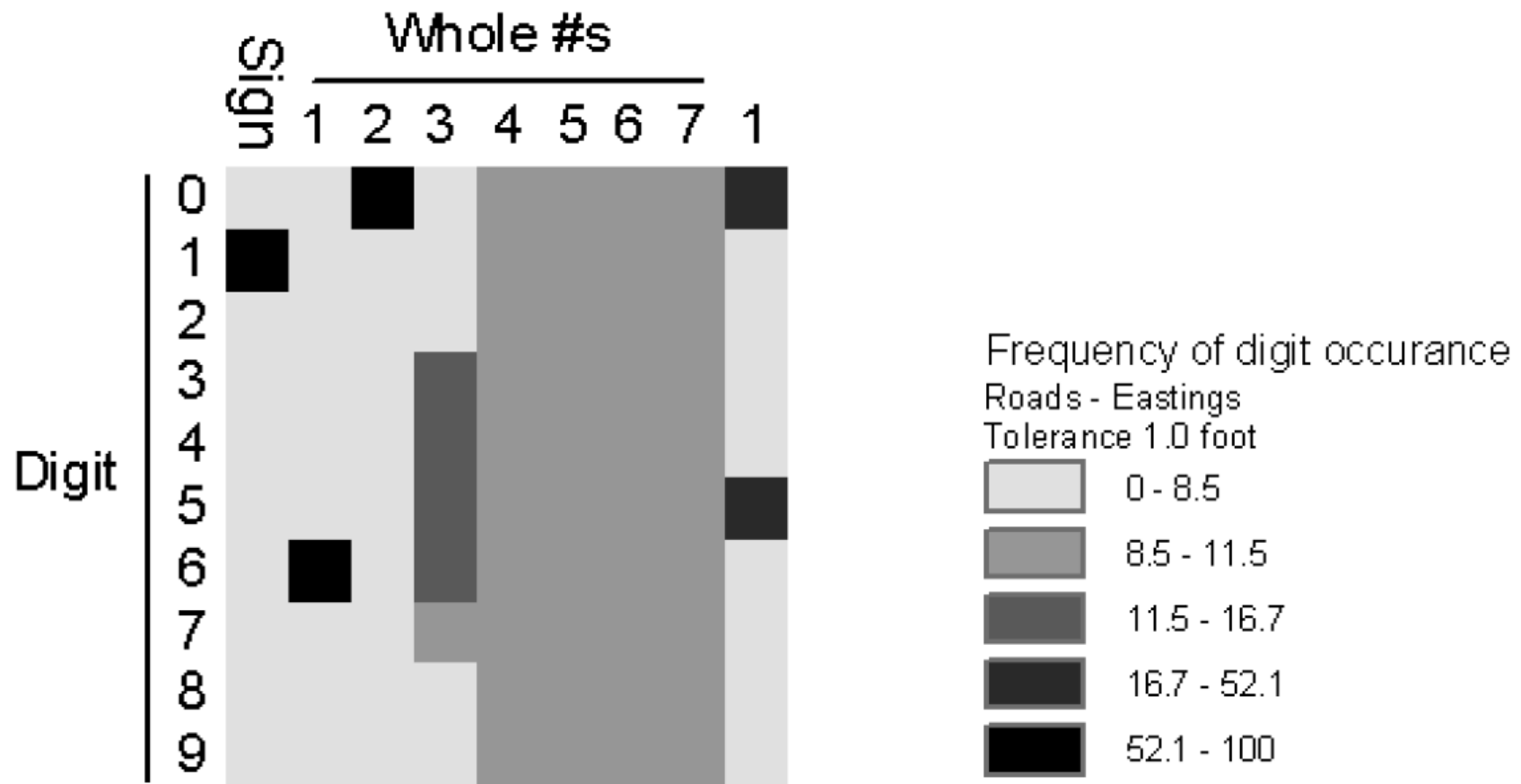
# Cartographic Communication



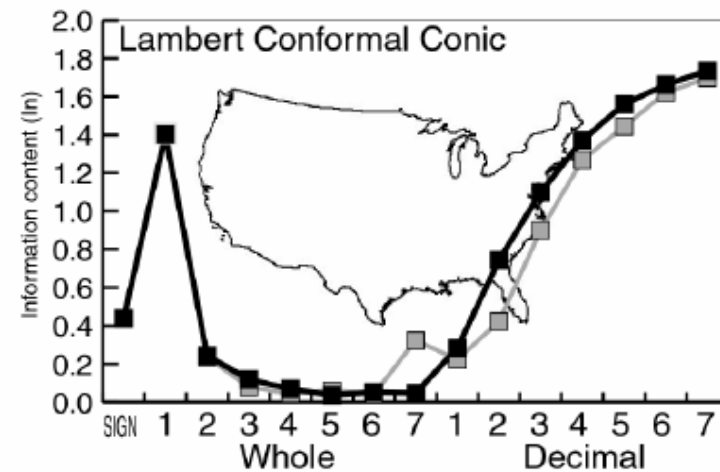
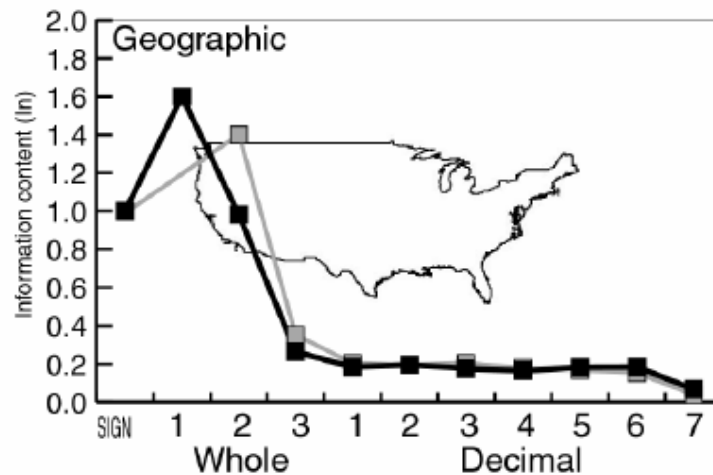
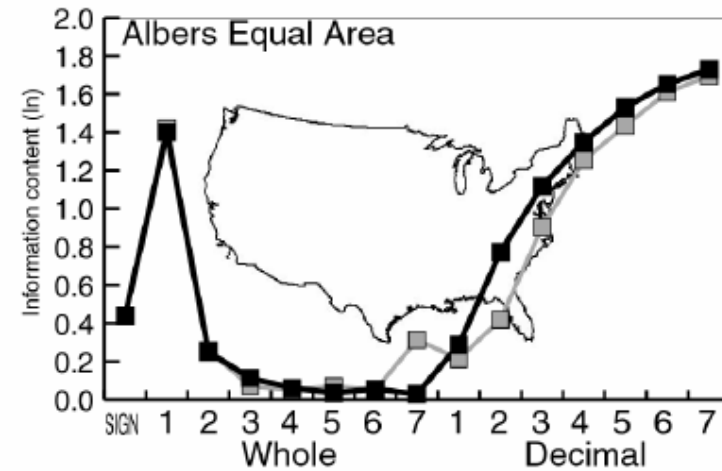
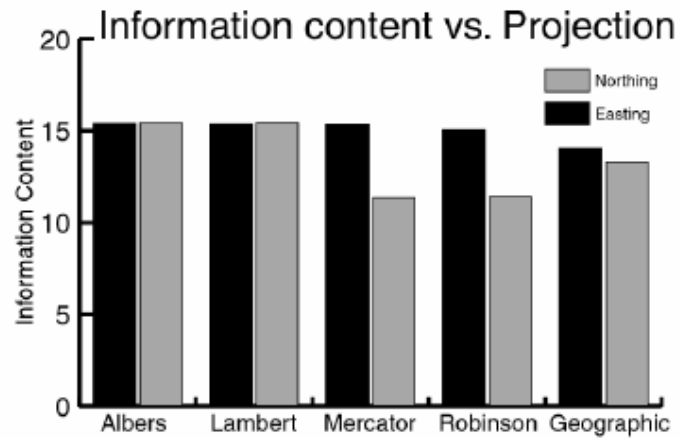
# Shannon and Mapping

- SAMOS and sending remotely sensed data to earth via radio signals
- Robinson & Petchenik's simple communications model
- Tobler's transformational view of cartography
- Shannonian theory waxes then wanes

# Clarke & Battersby (2003): CDDF



# The CDDF and Projections



# Cartographic transformations (Tobler, 1979)

- Transformations can be of :
  - Attribute Data
  - Locational properties
  - Graphics
  - Information content of maps
- Basic questions:
  - Is a transformation quantifiable?
  - Can the transformation process be automated?
  - Is a transformation invertible?
  - Is a transformation stable?

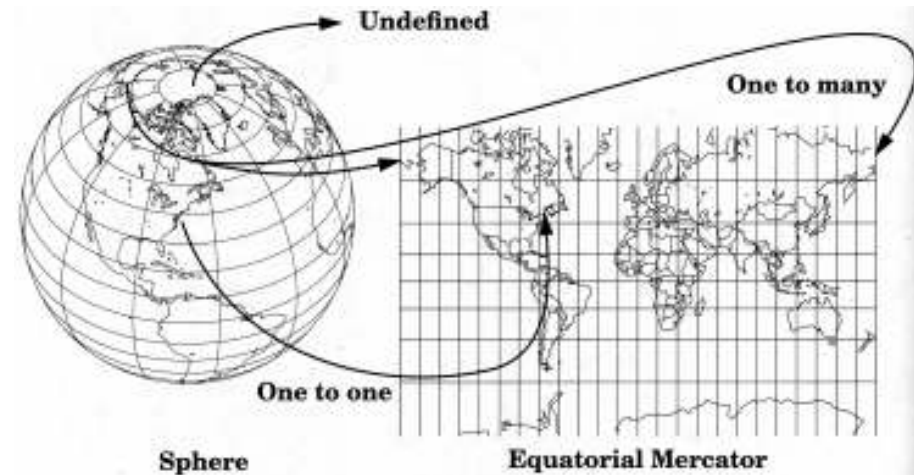
# Cartographic Invertibility

$$X = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_n & y_n \end{bmatrix}$$

$XT = X'$  Normal transformation.

$X'T^{-1} = X + E$  Inverse transformation.

$TT^{-1} = E$        $TT^{-1} = I$  Transformation both with and without error.



# Problems with Shannon's Theory

- Requires complete description of possible states at outset (Representational schema)
- Ignores issues of meaning to sender or receiver
- Is scale-less
- Is indifferent to order, organization, or coding schema
- Is hard to apply in complex channels (e.g. images, networks)



# Contemporary Geocommunications



Globally networked



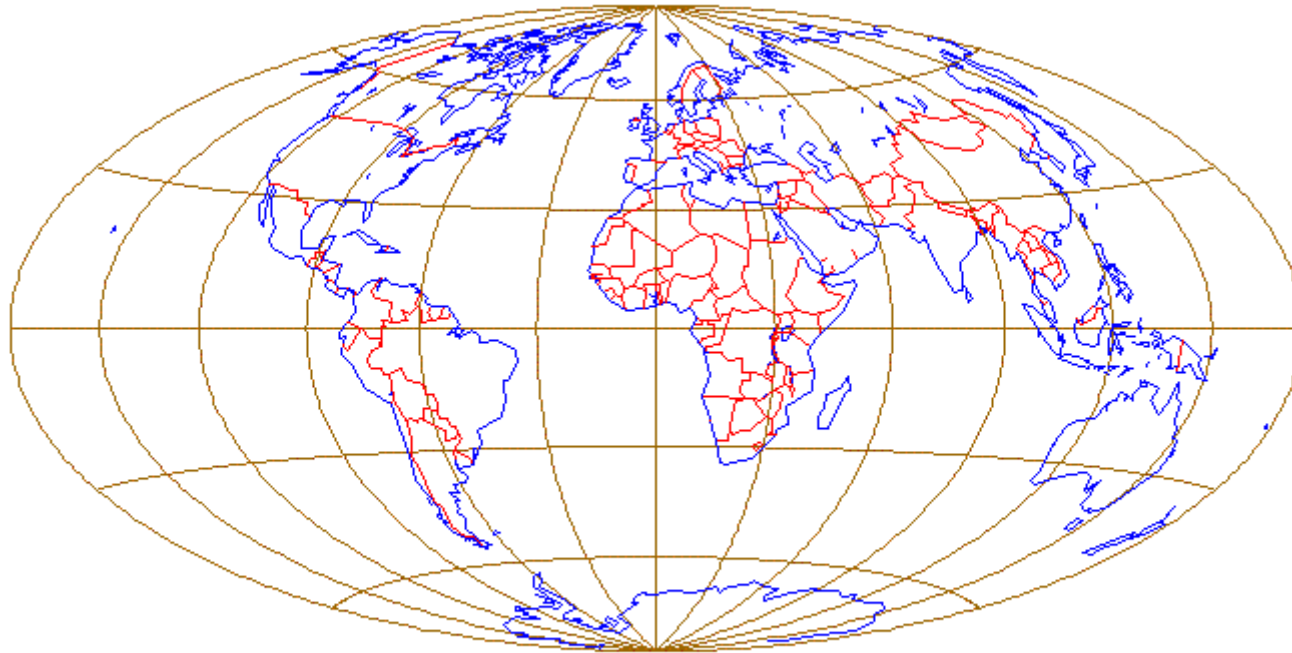
Overcapacity

# How efficient is the channel of communication?

- Is there information that can't be expressed, e.g. in text
- What are the limits of a GIS as a communication channel?
- What information about a place can't be expressed in GIS?
- Is the message optimally expressed in the coding system?
- What if the sender and receiver can't understand each other?
  - different language
  - different alphabet
  - different GIS
  - different data model

Please stand by for  
information transfer

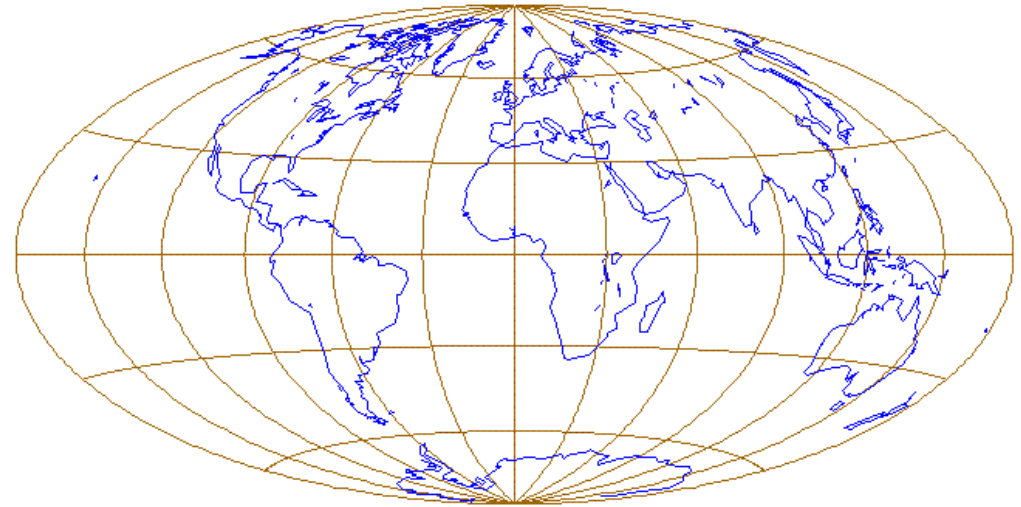
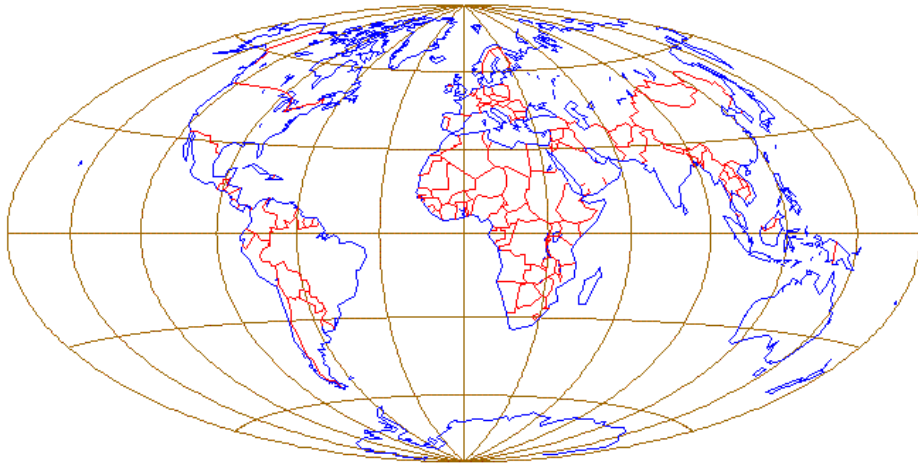
# Source-channel-receiver



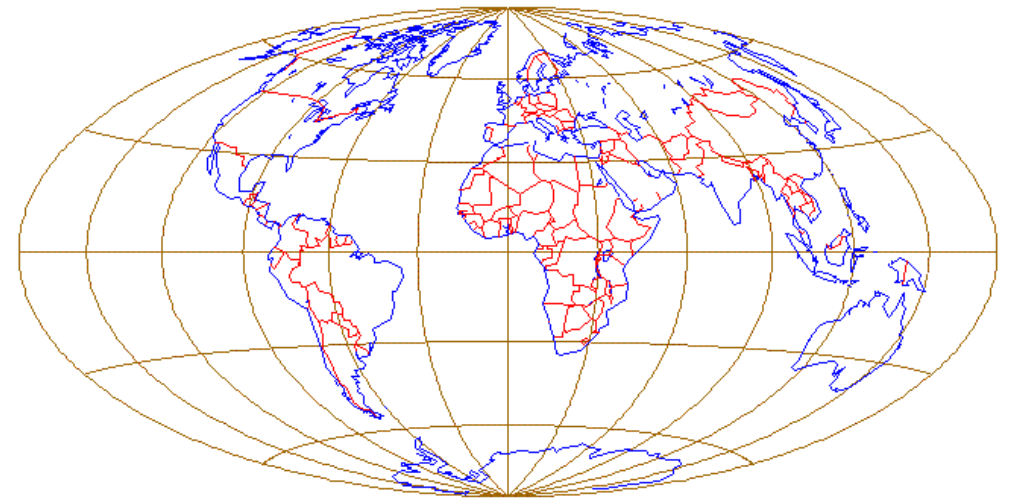
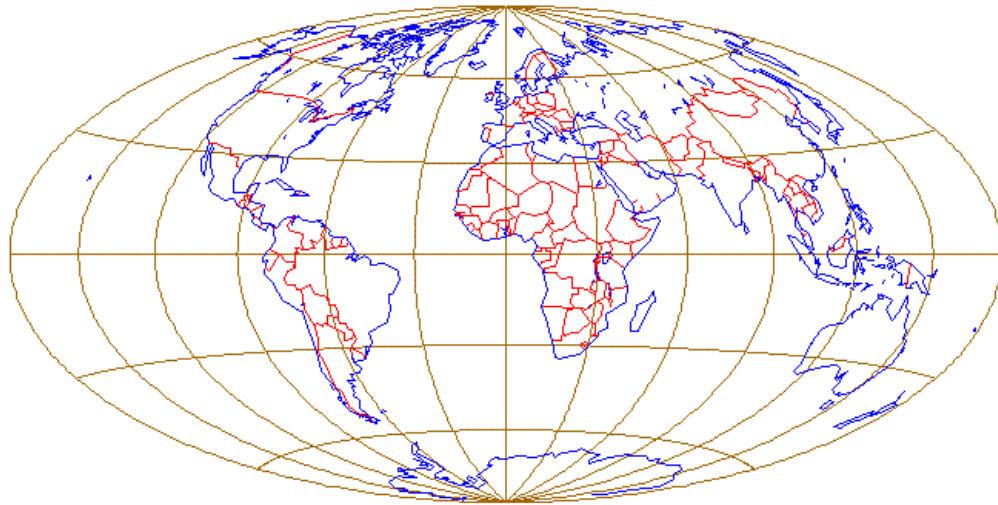
# Not one channel

- Content (World outline map, political boundaries, graticule, Hammer projection)
- Confirmation (All details present)
- Level of detail (extent, granularity, periodicity)
- Specific relations (e.g. larger than, east of)
- Spatial patterns (plate tectonics)

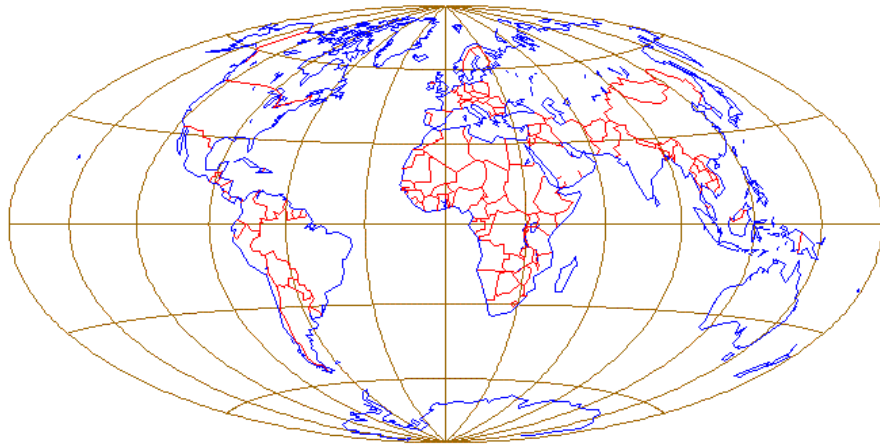
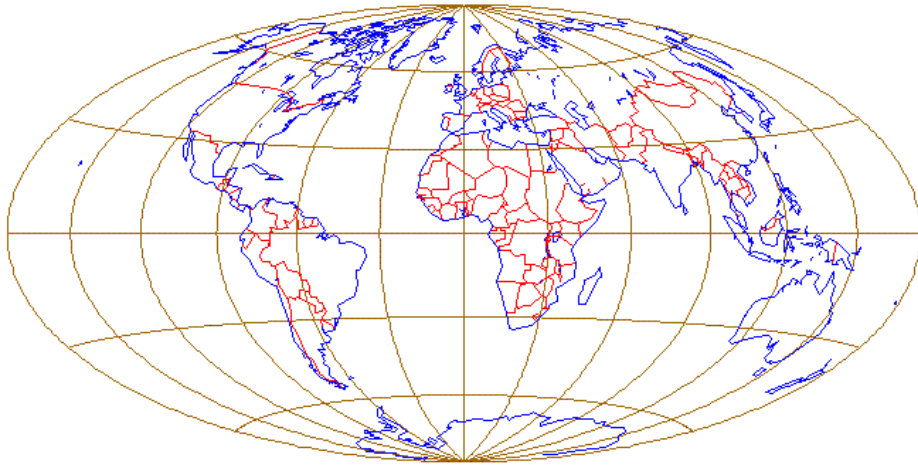
# Content



# Confirmation



# Level of Detail

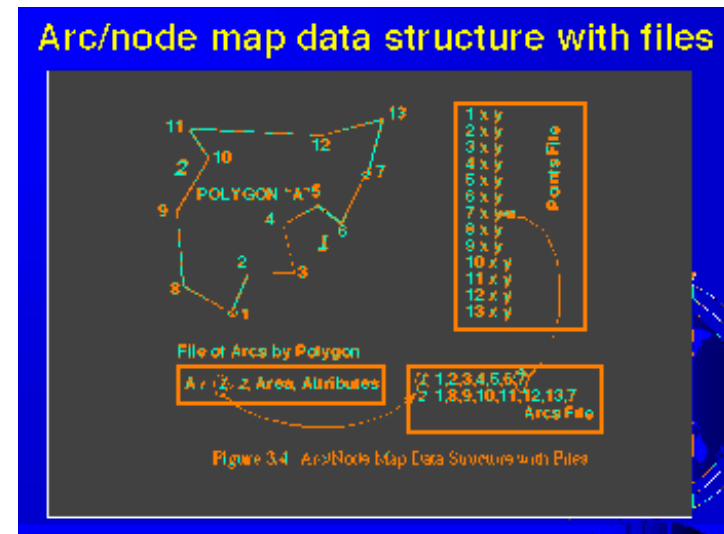


# Goodchild (2003)

- Information consists of linkages among properties and concepts
- Atomic elements are known as “facts”
- GI exists when facts include coordinates or placenames
- Value of a fact to the receiver depends on the number of facts already known
- Number of facts can be overwhelming

# A more complete representational model

- Metadata
- Ontology
- Geometry
- Relations
- Data structure
- Procedures and algorithms



# Goodchild (1998) “The geolibrary”

- Goodchild (1998) georeferenced vs. geoinformation
- Metaobjects, composed of atoms  
BLOBs
- Granularity: heirarchical, collection, “clump”
- GIBO: Geographic information-bearing objects

## Raper et al. “A Framework for evaluating geographical information” IJGIS 2003

- Information has 2 components
  - Representational
    - Ontology
    - Modeling
    - Systems
  - Communication
    - Relevance (+/-), public/private, **quality**
    - Commodity (**value**)
    - Exploration
    - Management

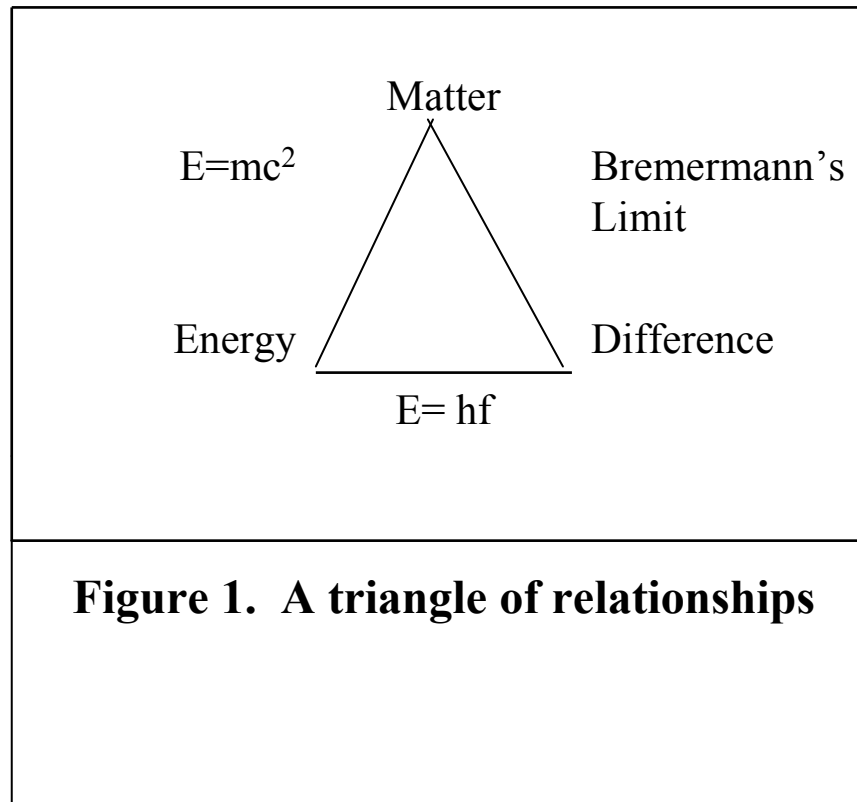
# Other approaches?

- Shannonian information theoretic approach does not have value beyond explaining a single channel with imperfect communications
- TCP/IP etc. removed issues of imperfect transmission by use of packets, at least for the Internet
- Internet has replaced channel with network
- Yet more than content needs transfer, and more than a receiver needs to be present
- What other approaches are there?

# PHYSICAL RELATIONSHIPS AMONG MATTER, ENERGY AND INFORMATION

## By Stuart A. Umpleby

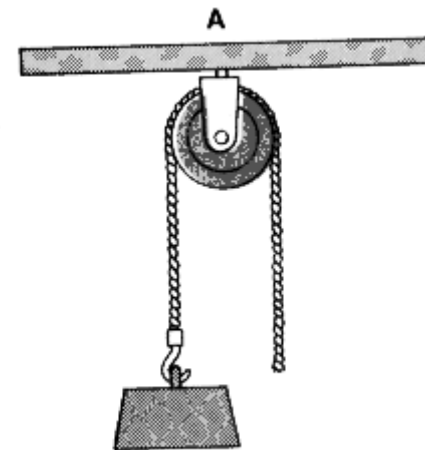
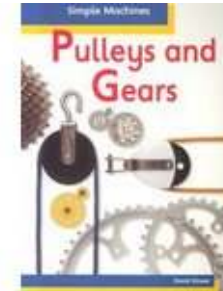
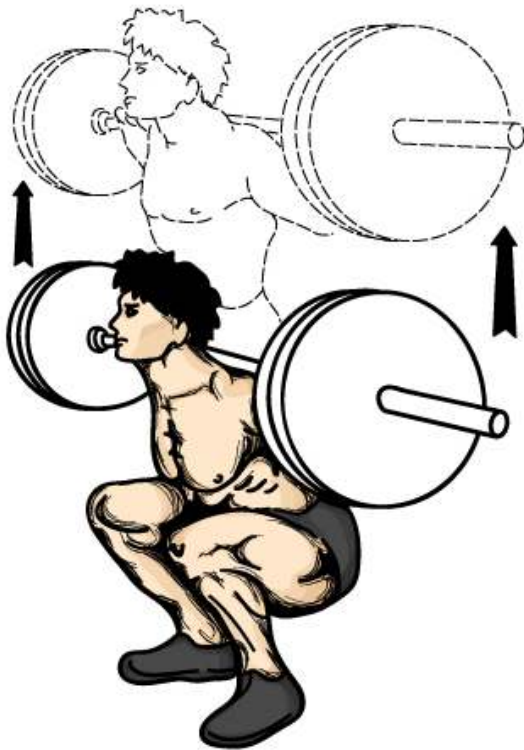
- [http://www.gwu.edu/~umpleby/recent\\_papers/2004\\_physical\\_relationships\\_among\\_matter\\_energy\\_information\\_umpleby.htm](http://www.gwu.edu/~umpleby/recent_papers/2004_physical_relationships_among_matter_energy_information_umpleby.htm)



# Matter and information

- Bremermann suggested that there is an upper bound on the rate at which symbols can be processed by matter
- They can be processed at speeds not exceeding  $10^{47}$  bits/gram/s
- Implications: “pure information”, atomic geoinformation? Energy, particles?
- How is geoinformation separable from pure information?

# Is this so crazy an idea?



Without information,  $E + 0 = e_1$

With information  $E + I = e_2 < e_1$

So  $I$  creates  $e_2 - e_1$  Or  $I$  lifts  $M_2 - M_1$

# For example: Solar Power



# Select one site

- “Cost” is the loss in power-generation capacity due to current versus “best” or better site chosen with information from the map
- Choices are among viable options, not optima
- Can be expressed in dollars, watts, carbon emissions savings, etc. (Negawatts)
- So information has a maximum value (E, \$, M) at the time of choice, afterwards value diminishes
- What are the trade-offs among E, M and I?

# Information value (quality?)

- Information increases value with access if it is non-competitive
- Map can be centrally stored, in which case each use is a Shannonian transfer (Client-server, e.g. GoogleEarth, MapQuest)
- Value does vary with prior knowledge, and with utility of the information
- Can be (is always?) highly temporally and spatially localized, i.e. obeys Tobler's First Law (Motive for LBS)



# In summary: **No**, or at least not yet

- A theory of geographic information is desirable
- Shannon's model has been taken as far as it will go
- Geographical representation seems unified, and almost fully formalized (e.g. OGIS)
- Theory must account for relations, process and time over spatial geometry
- Components are content, confirmation, LOD, relations and patterns, i.e. more than content
- Once again, theory may lag observation and technology
- Only fragments of a theory exist to date
- It is worth pursuing observation-free theory
- Lack of understanding impacts energy and materials
- Information "gap" filling, theory gap is not