Data Integration Part 2
Emerging Issues for the Data Community

Geography as an Integrating Tool?

David R. Fletcher
Geographic Paradigm Computing, Inc.

NATMEC 2000
Madison, WI
August 28, 2000

The data integration issue

- Extracting relevant information and delivering to the right people at the right time to make appropriate decisions.
- Why is this so hard?
  - We have the technology
  - We have the data
  - We're facing the decisions

Some ‘rational’ data beliefs . .

- Data is an objective view of the one, real world.
- Data captures inherent properties of objects.
- Data provides information creating knowledge.
- More, cheaper, faster, data is better.
  - i.e., higher correspondence with reality.
- Since “being” is information, the data experience is indistinguishable from “being there.”


Aristotle

If these premises are flawed, then any activity based on them is also be flawed.

. . . produces a data paradox

- more, cheaper, faster, data leads to
  - analysis paralysis
  - scarcity of attention
  - physical and mental stress
  - greater uncertainty
  - poorer decisions
  - wasted resources

. . .and consequently creates

- Info glut - drowning in data
  - AKA Data Smog, Information Overload, Infolanche
- More data than can be comprehended or used
- “Millions of pictures and no one to look at them.” - Porter Goss,
- Too much data, not enough information

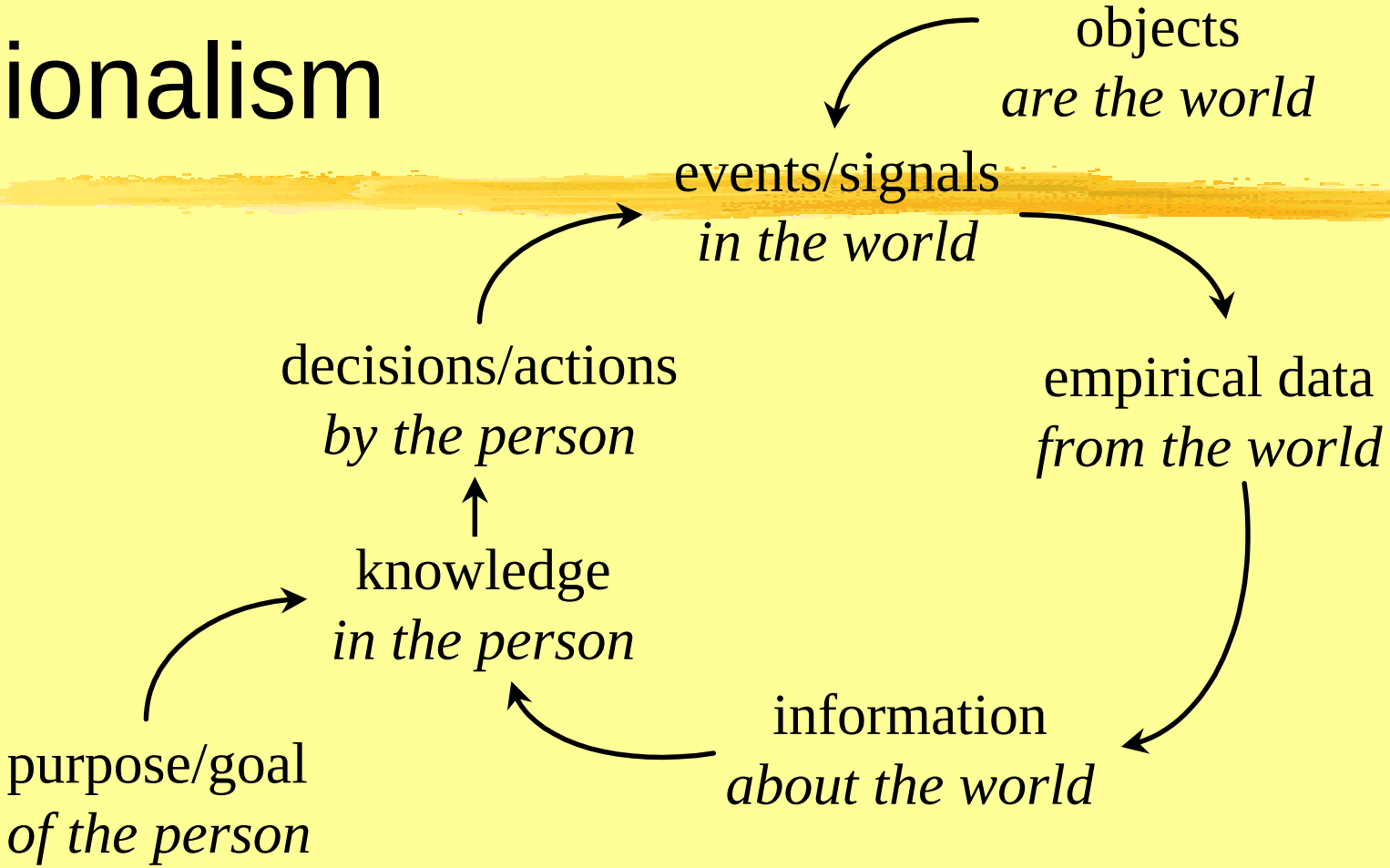
Comments on info glut

- "There are very few political [or] social . . . problems that arise because of insufficient information . . Yet [we] stand firm in believing that we need yet more information. Information has become a form of garbage." *Neil Postman*
- "Problems arise when . . . the wealth of information makes people depend on the process of collection and hyperanalysis rather than on judgment." *Forbes*
- "Ironically, in this glut of information, we may know less, not more about the world around us. At the very least, surely we know more things with less certainty than we used to." *Herald Journal*

Beliefs leading to info glut

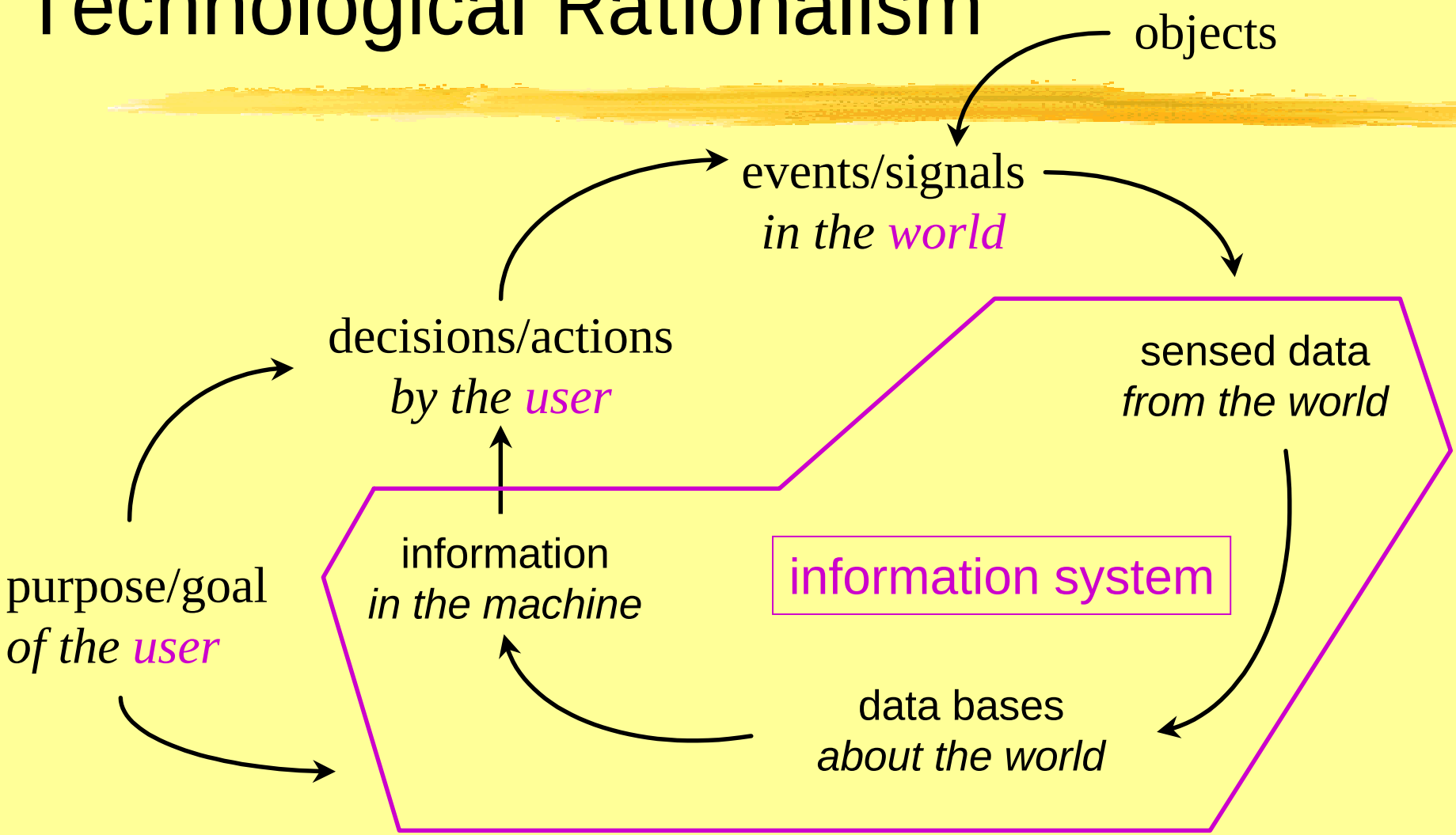
- Ideological - technopoly
  - data contains information
  - information is meaning
  - human beings are information processors
  - technological innovation is human progress
- Philosophical - rationalism
  - the world reduces to geometry and number
  - the representation is the thing
    - all meaningful objects in the world are spatial-temporal.
    - faithful representation of metrical data is an accurate reconstruction of reality - correspondence is truth.
    - a one-to-one correspondence between the earth and the world

Rationalism



The notion that a thinking subject exists apart from an independent and measurable spatial object world

Technological Rationalism



But, what if rationalism is illusionary?

- The map is NOT the territory
- The map does not show all of the territory
- The map reflects the map maker's POV

- The data is NOT the object
- Attributes describe processes, NOT objects
- Data reflects the designer's POV, not the user's.

A different perspective

- Knowledge is constructed out of use, not data.
  - Knowledge emerges from human interactions with the world.
  - Knowledge is *human*, not data centered.
  - Knowledge leads to data.
- The earth is *not* the world.

Genuine, practical knowledge is forged by involvement in the world.

Knowledge by experience

- Direct experience of objects or events
  - direct sensation (e.g., seeing, hearing, touching, manipulating)
  - counting (i.e., naming or listing a group of objects)
  - measuring (i.e., a quantitative comparison of a thing to a standard
- The quality of experience is influenced by the
  - condition of the observer (e.g., knowledge, experience, understanding, diligence, skill, motivation, physical abilities)
  - context of the object (e.g., association, meaning)
  - quality of the interaction (e.g., urgency, setting)

Knowledge by communication

- Communication: Reporting or receiving verifiable notations or descriptions of objects or events that we have seen, heard, felt, counted, measured, witnessed or experienced with another person.
- Communication depends on the trustworthiness and shared understanding of the participants.
- Metadata may conceal disagreement and ambiguity.

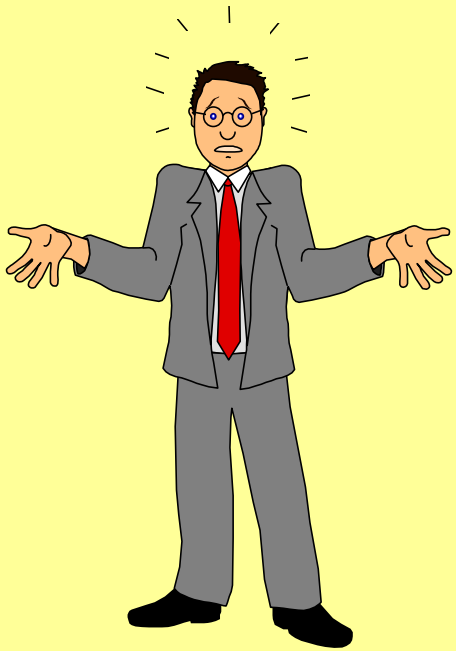
The problem with metadata

- Definitions are not about things but about use.
- Standard definitions are not the same as shared meaning
- Meaning is established by context
- Data never has exactly the same meaning twice.
- Metadata may be a necessary but not a sufficient component of shared meaning.

Knowledge by reasoning

- Reasoning: Making conclusions, inferences and judgments about the world.
  - inference: statements about the unknown based on what is known
  - judgment: a conclusion based on facts and inferences or approval or disapproval of events, other people or objects
- Decisions relate directly to the quality of the underlying facts *and* the knowledge and skills of the reasoning person.
- Not all knowledge can be derived from reasoning.
- Reasoning under uncertainty will always be uncertain.

So what does all of this have to do with traffic data?!?



Let’s get practical!

- Traffic emerges as a consequence of the dynamic interplay between travel demand and transport choice. Traffic data is an incomplete, static trace of this process.
- Travel demand is a complex human desire.
- Transport choice is a complex human judgment.
- Traffic is a complex human behavior.
- These factors are tied inextricably to places in the world that is larger than just physical geography.
- This implies that traffic data is contextual and not merely physical. That is, an understanding of place is essential in understanding traffic data.

Place based knowledge

- Places in the world encompass physical, social, economic, and environmental factors.
- Every ‘place’ is subject to multiple POV’s
  - Political, planning, engineering, quality of life, etc.
- Each POV (planning, engineering, constituent, etc.) ‘creates’ different overlapping worlds.
- Places can be local – intersections, neighborhoods, projects
- Or very large – metropolitan areas, corridors, states or regions.
- Every place is complex, interconnected and interdependent with other places.

Place based reasoning

- Information about places needs to be integrated, synthesized and understood from multiple POV's.
- Place based reasoning is fundamentally different from layer based data strategies.
- Layer approaches suggests that traffic data has location attributes.
- Place based reasoning suggests that places have traffic - and many other characteristics – in a holistic, scalable representation.

Conclusions

- Shared understanding of traffic data results from mutual use and not from either data or meta data.
- Human judgments such as trust, competence, and reliability are essential elements in determining data quality.
- We don’t integrate data because we don’t trust it or don’t understand it, NOT because we don’t have it.
- Developing human centered approaches enhances trust.
- Providing a world context enhances understanding.
- Places offer better ways to think about traffic data.

"You live in a universe and within you, you form pictures of the universe as it appears to you. And you know nothing of that universe except for pictures. But the pictures within you of the universe are not the universe. . ."

A. E. van Vogt

Some References

- Coyne, R., *Designing Information Technology in the Postmodern Age*
- Hayakawa, S. I., *Language in Thought and Action*
- Heidegger, M., *Being and Time*
- Postman, N., *Technopoly*
- Shenk, D., *Data Smog*