



**Building of Research Team in the
Field of Environmental Modeling
and the Use of
Geoinformation Systems with the
Consequence in Participation in
International
Networks and Programs**

**Registry number: CZ.
1.07/2.3.00/20.0170**

StatGis Team



More Than Big Data

Wasn't geographic information
always big?

**Wasn't geographic
information always big?**

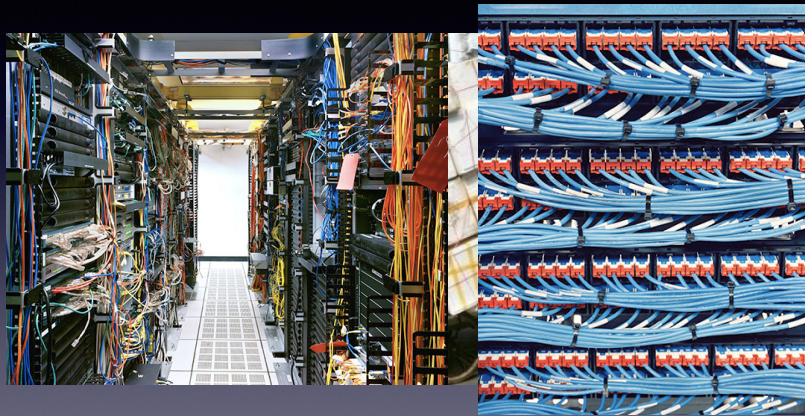
Some similarities, some differences, some
thoughts about big data and
cartographic mapping

Francis Harvey
Department of Geography, Society and Environment
University of Minnesota
fharvey@umn.edu
@fhap13

Key Points

- Maps are big data with “lies”
- Many areas of science are moving to become data-intensive
- Media focus on “Big Data” overlooks the distinction to data-intensive science
- Big amounts of data change how we can approach scientific query
- There are many scientific challenges

Big Data



Variety

- Structured
- Unstructured
- Semi-structured
- All the above

3 Vs of Big Data

- Terabytes
- Records
- Transactions
- Tables, files

- Batch
- Real-time
- Streams
- Near-time

Volume

Velocity

<http://kavyamuthanna.files.wordpress.com/2013/01/picture-11.png>

Big Data isn't just BIG

Outline

- Mapping and maps
- Big data and data intensive science
- Science/knowledge
- Knowledge Ecology
- Cartographers as knowledge ecologists



Mapping and Maps

- the *map* - the universal metaphor (Bowker)
- unparalleled applications
 - make discoveries
 - analysis of relationships
 - consider error and uncertainty
 - spatial enablement
 - attempt to control and gain power over a situation
- To work, "all maps lie" (Monmonnier)

Mapping

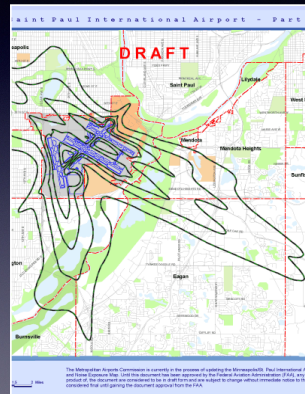
- The mysterious connection of experience with the unexperienced, unexperience-able, and knowledge
- A functional definition: The geographical and cartographic attempt to control and gain power over situations

The “Deceit” of Maps

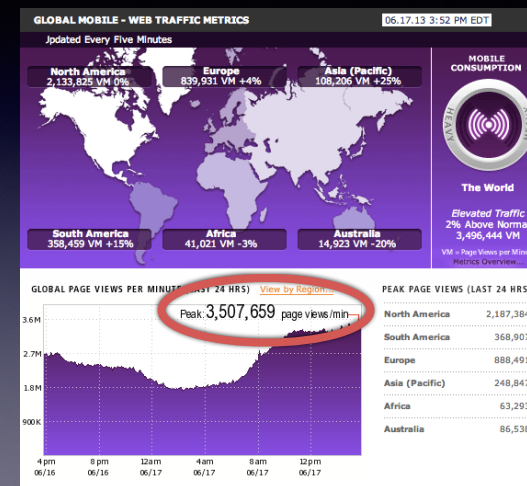
- Maps are NOT the territory, but have similar structures to the territory. This accounts for their usefulness, at least when they are correct (based on Korbrzyski, 1933)
- Geographic representation
- Cartographic representation
- Conventions

Map as Exemplars of Big Data?

- non-essentialist understanding
- irreductionist
- maps as artifacts, boundary objects, that connect people



Big Data?



And bigger?

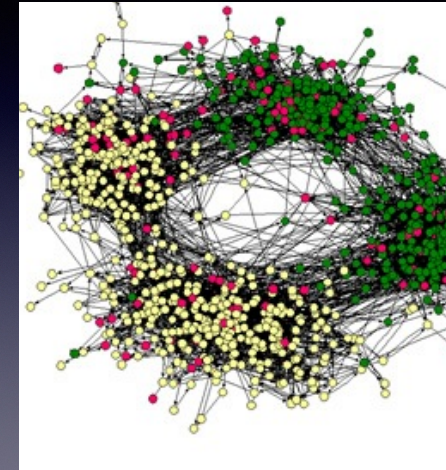
- Everyone uses spatial computing
- Everyone is a mapmaker
- Every computing device can be location aware
- Everyone has rising expectations and awareness of risks



<http://blogs.lawyers.com/wp-content/uploads/2012/01/Phone-tracking-300.jpg>

Data Intensive Science

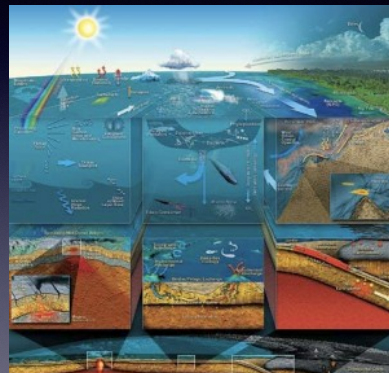
- Jim Gray: The fourth paradigm
- Jeanette Wing: Computational thinking
- Creative discovery



Friendship network of children in a US school
<http://hoipio.blogspot.com/2010/01/information-revolution-fourth-paradigm.html>

Scientific progress will come from working with data

- GIS becoming a fundamental technology for working with data
 - Maps remain central
 - because of memes
- Big Data in the media
- IT meets science

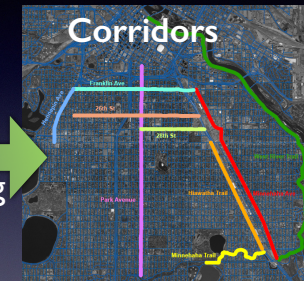


Complexity of planetary life-interactions
<http://www.earthzine.org/2010/03/22/observing-the-oceans-a-2020-vision-for-ocean-science/>

Example of bike commuters

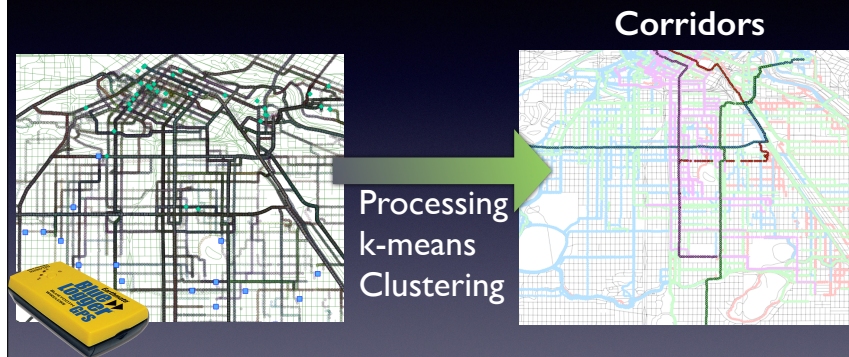


Processing Analysis



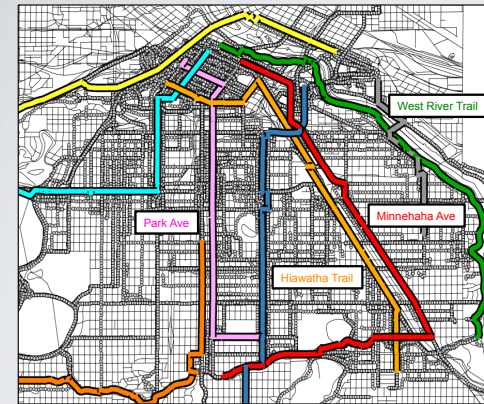
Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
GPS Route Speed (m/s)	110	.41	13.80	4.5121	1.36005
Speed on Hwaehwa (m/s)	21	3.63	6.79	5.3790	.84994
Speed on Park (m/s)	32	3.24	6.21	4.4485	.78912
Speed on Minnehaha (m/s)	4	4.42	5.68	5.1676	.57973
Speed on WestRiver (m/s)	9	4.68	7.07	6.0360	.72864
Speed on MinneTrail (m/s)	8	4.83	6.60	5.9409	.61032
Speed on Hennepin (m/s)	4	3.99	6.56	5.4339	1.22731
Speed on 26th (m/s)	2	3.63	5.91	4.7726	1.61355
Speed on Franklin (m/s)	6	3.46	6.13	5.1469	.90623
Speed on 28th (m/s)	6	4.00	6.85	4.9500	1.06658
Valid N (listwise)	0				

Example of bike commuters

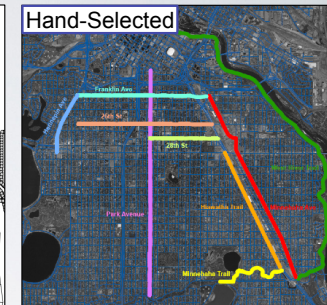


17

8-PRIMARY CORRIDORS: METS VS HAND-SELECTED



Matrix-Element Track Similarity
w/ PAM, $k = 8$



Similarities:
Many common corridors
Park Ave, Hiawatha, Minnehaha Ave, West River Trail

Differences:
METS missed East/West corridors
Few trajectories that direction (k-medoid)
Hand-Selected did not put a corridor in N/NW Minneapolis

18

Research Challenges

- Processing large data sets
- Error detection
- Automated learning
- Spatio-temporal analysis
- Visualization
- Capture, curation, analysis



Scientific Challenges

- David Weinberger
 - limits of empiricism
 - the brickyard analogy
 - when does the complex become simple enough to understand
 - the role of universals
 - the continued importance of networks

Brickyards of the Imagination

- Sherry Turkle
 - Simulation comes first now; reality comes second
 - **Alone Together** (book and TED talk)
 - Simulation as a barrier to real messy and demanding relationships
 - Less ability to self-reflect
 - Passive connections as friends



Science/Knowledge



Science - the organized pursuit of knowledge
(Library of Congress)

Knowledge/Science

- William Butler Yeats' Symbolism
 - The visible world is no longer reality, and the unseen world is no longer a dream. (Symons, 1900)
- Michael Foucault's philosophical excavations
 - examine discourses through archaeology to understand empowerment/disempowerment
- Thomas Kuhn's Paradigma
 - science understood through systems of thinking, thought, institutions, equipment called paradigma
- Paul Feyerabend's *Against Method*
 - scientific advances can only be understood in historical context
- Karl Popper
 - world and reality (Worlds 1, 2, 3)

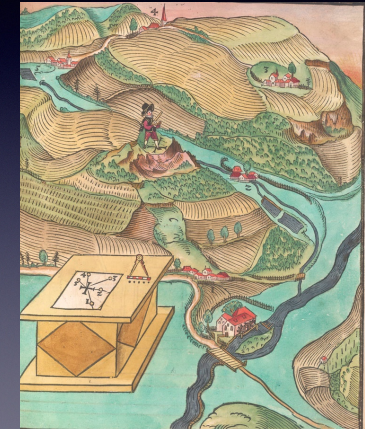
Knowledge Ecologies

- network of interactions, often essentialized, but always with limits
 - "the pattern that connects" G. Bateson
- big data ontologies and epistemologies and para-empirics
 - Rumsfeld
- different approaches: meaning is always arising in use, memes persist

Cartographers as Knowledge Ecologists

Knowledge Ecology of Cartography, or skills and conventions, and concepts from hundreds and thousands of years

- Always dealing with big data
- Some of cartography's memes
 - Scale
 - Generalization
 - Regions



16th Century Surveying Instructions - British Library

Handling geographic information for mapping

- Scale
- Generalization
- Regions

Challenges: Traditions and modernity

Known Challenges in handling big data

- Online Mapping
 - Reliable and accurate data transmission and representation
- Topology
 - Relations and efficiencies in data storage/processing
- Knowledge
 - The brickyard problem

Para-empirical challenges

- Linda Kurgan, *Close Up, At A Distance*
- Para-empirics: data is never facts, but representations
- Measurements based in conventions, aesthetics, and rhetorics we associate with images
- Data is para-empirical
- Room for everyone to participate/engage



Conclusions

- Yes, geographic information was always big
- Approaches are very different
- Big data is different, but data-intensive science is far more different
- Cartography's memes are central to big data mapping
- Knowledge and meaning come through use in ecologies of knowledge

References

- Berggren, J. L., & Jones, A. (2000). *Ptolemy's Geography, An Annotated Translation of the Theoretical Chapters*. Princeton, NJ: Princeton University Press.
- Claval, P., Nardy, J. P., & Vidal, d. L. B., Paul, (1968). *Pour le cinquanteaire de la mort de Paul Vidal de La Blache*. Paris: les Belles lettres.
- Evans, Michael R., Dev Oliver, Shashi Shekhar, and Francis Harvey. (2013) Fast and Exact Network Trajectory Smilarity Computation: A Case-Study on Bicycle Corridor Planning. In Second ACM SIGKDDO International Workshop on Urban Computing, Chicago, IL, USA: ACM.
- Finch, V. (1933). Montfort: A study in landscape types in southwestern Wisconsin. *Geographic Society of Chicago, Bulletin* 9.
- Finch, V. C. (1939). *Geographical Science and Social Philosophy*. *Annals of the AAG*, 29(1), 1-28.
- Hartshorne, R. (1939). *The Nature of Geography: A Critical Survey of Current Thought in the Light of the Past* (Second Edition ed.). Lancaster, PA: Association of American Geographers (reprinted with corrections, 1961).
- Hartshorne, R. (1959). *Perspective on the Nature of Geography*. Chicago: Rand McNalley and Co.
- Martin, G. J., & James, P. E. (1993). *All Possible Worlds. A History of Geographical Ideas*. New York: John Wiley and Sons.
- Harvey, F. (2008). *A Primer of GIS. Fundamental Geographic and Cartographic Concepts*. New York: Guilford.
- Harvey, F. (2009). *Of Boundary Objects and Boundaries: Local Stabilization of the Polish Cadastral Infrastructure*. *The Information Society*, 25(5), 315-327.
- Harvey, Francis. (2013) A New Age of Discovery: The Post-Gis Era. In *Gi Forum. Creating the Gisociety*, edited by Thomas Jekel, Adrijana Car, Josef Strobl, and Gerald Griesebner. Berlin: Wichmann Verlag.
- Livingstone, D. N. (1992). *The Geographical Tradition*. Oxford: Blackwell.
- McHarg, I. (1969). *Design with Nature*. New York: Natural History Press.
- Monmonier, M. (1991). *Ethics and map design. Six strategies for confronting the traditional one-map solution*. *Cartographic Perspectives*, 10, 3-8.
- Monmonier, M. (1991). *How to lie with Maps*. Chicago: University of Chicago Press.
- Wood, D. (1992). *The Power of Maps*. New York: Guilford Press.

Thanks and Acknowledgements

Colleagues and students at the University of
Minnesota for comments, examples, and
inspiration particularly Shashi Shekhar and
Mike Evans.

The StatGIS Team