Topic Maps and Beyond

Connecting Information

Michel Biezunski

Infoloom

(718) 921-0901
mb@infoloom.com

http://www.infoloom.com

Michel Biezunski

- Consultant / Innovator, Infoloom, New York
- Works on Solutions for organizing / navigating information corpora.
- Creator of the Topic Maps paradigm, with Steve Newcomb
- Background: History/Philosophy of Science

Outline

- Topic Maps In A Nutshell
- Topic Maps & RDF
- Topic Maps Applications: The Tax Map Story
- Beyond Topic Maps: the Data Projection Model

The History of Topic Maps

- Early 1990s:
  - Unix vendors. Documentation Interoperability
  - Davenport -> XML, Docbook, Topic Maps
- 1996: submitted as an ISO work item
- 2000: First Edition published. ISO/IEC 13250
- 2001: XTM, XML Version published
- 2003: XTM integrated into ISO/IEC 13250 2nd Edition.
- Work in progress: XTM2, TMRM, TMDM, TMQL, TMCL, TMGL...

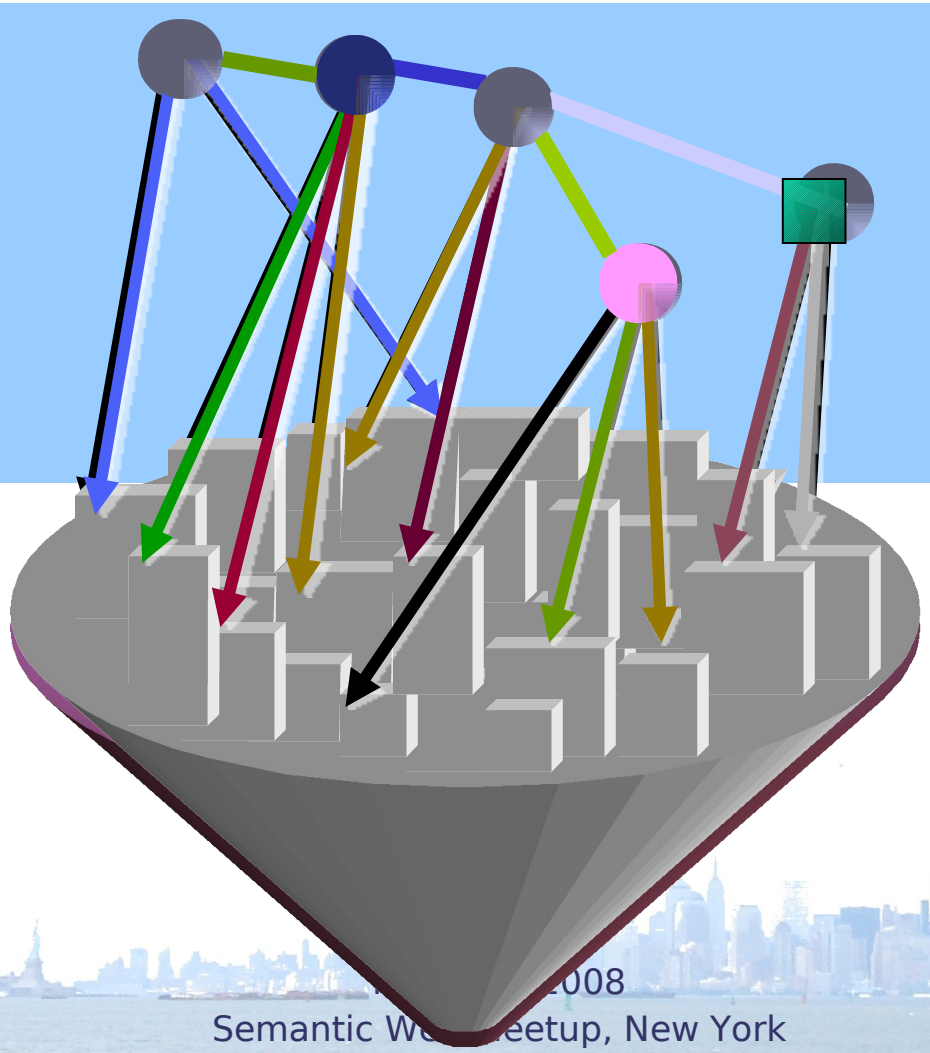
Information

Web links

Objects:
Documents,
images,
videos, etc.

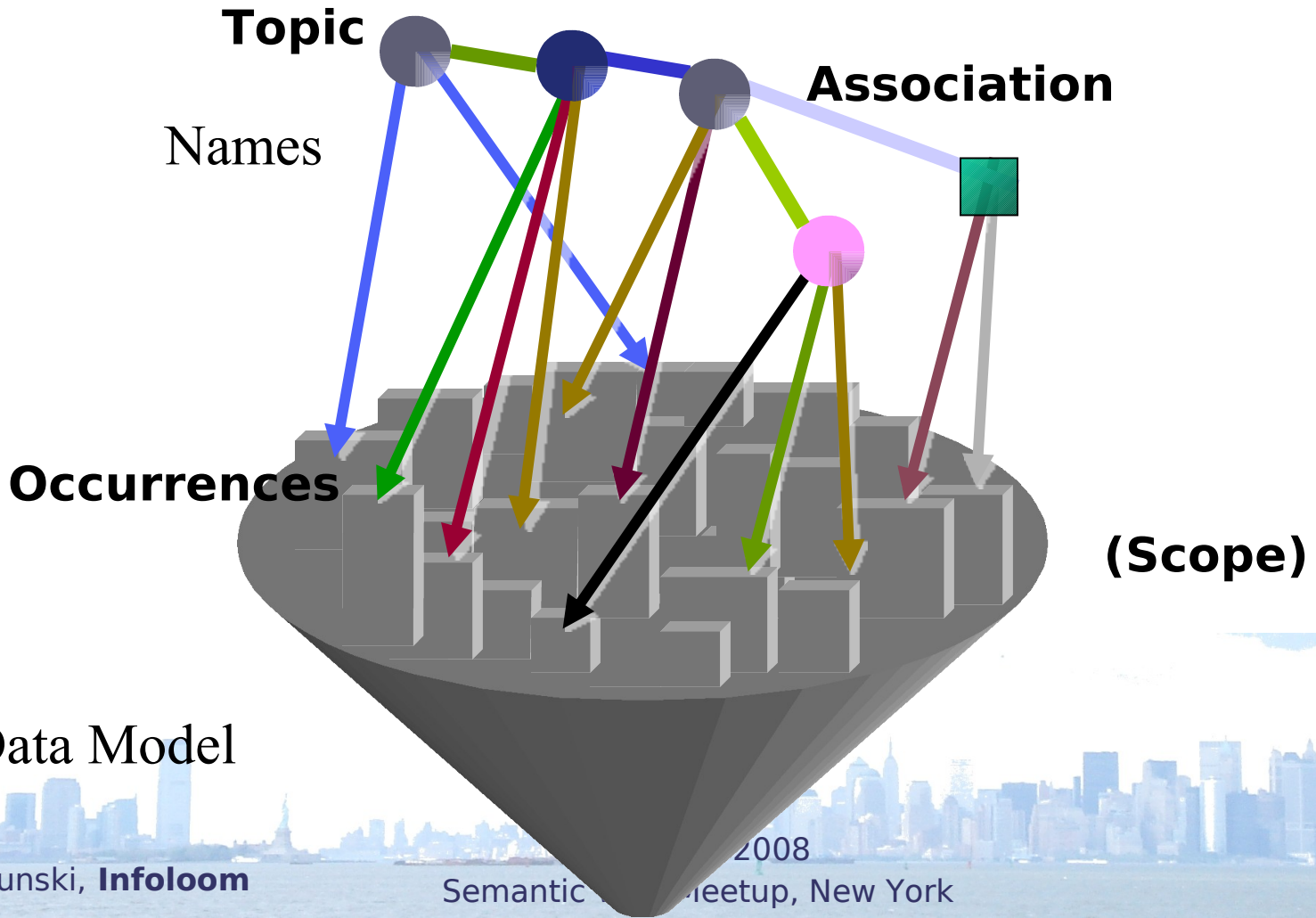
Topics

Topic Map



User-defined semantics

Basic Topic Maps Constructs



TM Data Model

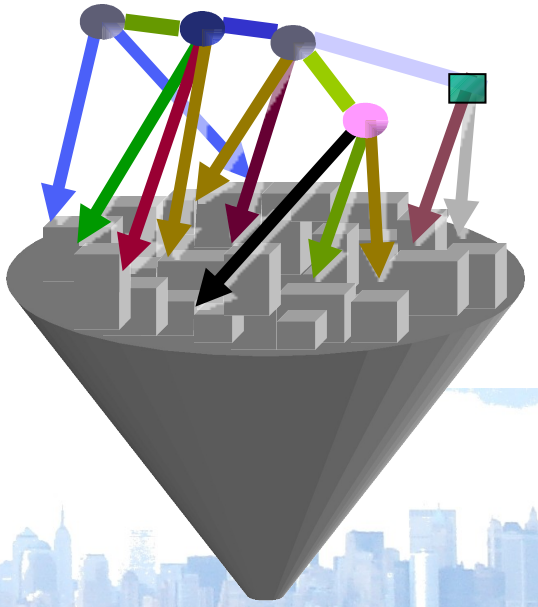
Why Topic Maps?

Before

Infoglut

After

Information / Knowledge Management



Topic Maps & RDF: Similarities

- Graph based
- Web enabled
- XML Serialization
- Generic languages
- Knowledge based
- Ontology based
- Different layers:
  RDF > RDFS > OWL >SPARQL
  TMRM >? TMDM > TMCL > TMQL

Topic Maps & RDF: Differences

- n-directional links
- Topics are subject reifiers
- Focus on Human perspectives on subjects
- Web not required

- 1-directional links
- URIs are subject reifiers
- Focus on Machine intelligence
- Web required

Topic Maps & RDF: Problems

Topic Maps

- Topic Map Data Model too narrow.
- Satellites standards in construction: insufficiently driven by user requirements.

RDF

- Learning curve very high. Perceived as “academic”
- Current applications tend to use “elementary particles” (RDF statements) as if they were “molecules” (Complex sets of relationships treated as units)

Topic Maps Applications

- Encyclopaedias
- Government Applications:
  - US: DOE, IRS, DOD
  - Europe: Norway, Netherlands, Germany
- Topic Maps for Learning (TM4L)
- Libraries: Vanderbilt University
- Free Topic Map Browsers (Ontopia), Topic Map Engines (TM4J), Current Research Work (Germany, Norway)

The Tax Map Story

Tax Map is an editorial product.

Used in the IRS call centers

Project started in 2001.

Automatic year-to-year updates.

Enhanced and improved regularly.

Now available on the Internet

User Requirements for TaxMap

- Enable research by subject.
- Be easy to use.
- Not require changes to IRS workflows.

Tax Map CD on the Internet

Combines Automation and Human Expertise

Automatic processing:

- Batch process from sources to Web
- Comes from independently maintained sources.
  - Products of independent workflows within IRS
    Pubs, Forms and Instructions, FAQs, Tax Topics
- Some editorial operations are globally automated:
  - Extracting subjects from product database and XML/SGML documents,
  - Automatic Rules for making relationships, synonyms, deleting, combining, etc.

Human input

- Other editorial operations are specified by tax experts.

Where does this come from?

- Why are two topic names associated?
  - Sometimes it looks weird.
  - Does it result from the action of some algorithm?
  - Is it the decision of tax experts which may be biased in some way?
  - Is it a bug in the production process?
- Answer:
  - We don't know.
- Request:
  - We should know.

Auditing the Topic Map

...Using an approach similar to RDF

The Data Projection Model

- is based on the idea that no information item is ever isolated.
- Any semantic can be expressed as a set of binary relations.
- The semantics of operators are not constrained:
  - They could be simple assertions
  - They could be processes
- Binary relations comprise graphs.

Methodology: 2 steps

1st step: decompose a system into its elementary processes expressed as binary relationships

PERSPECTOR : < x | o | y >

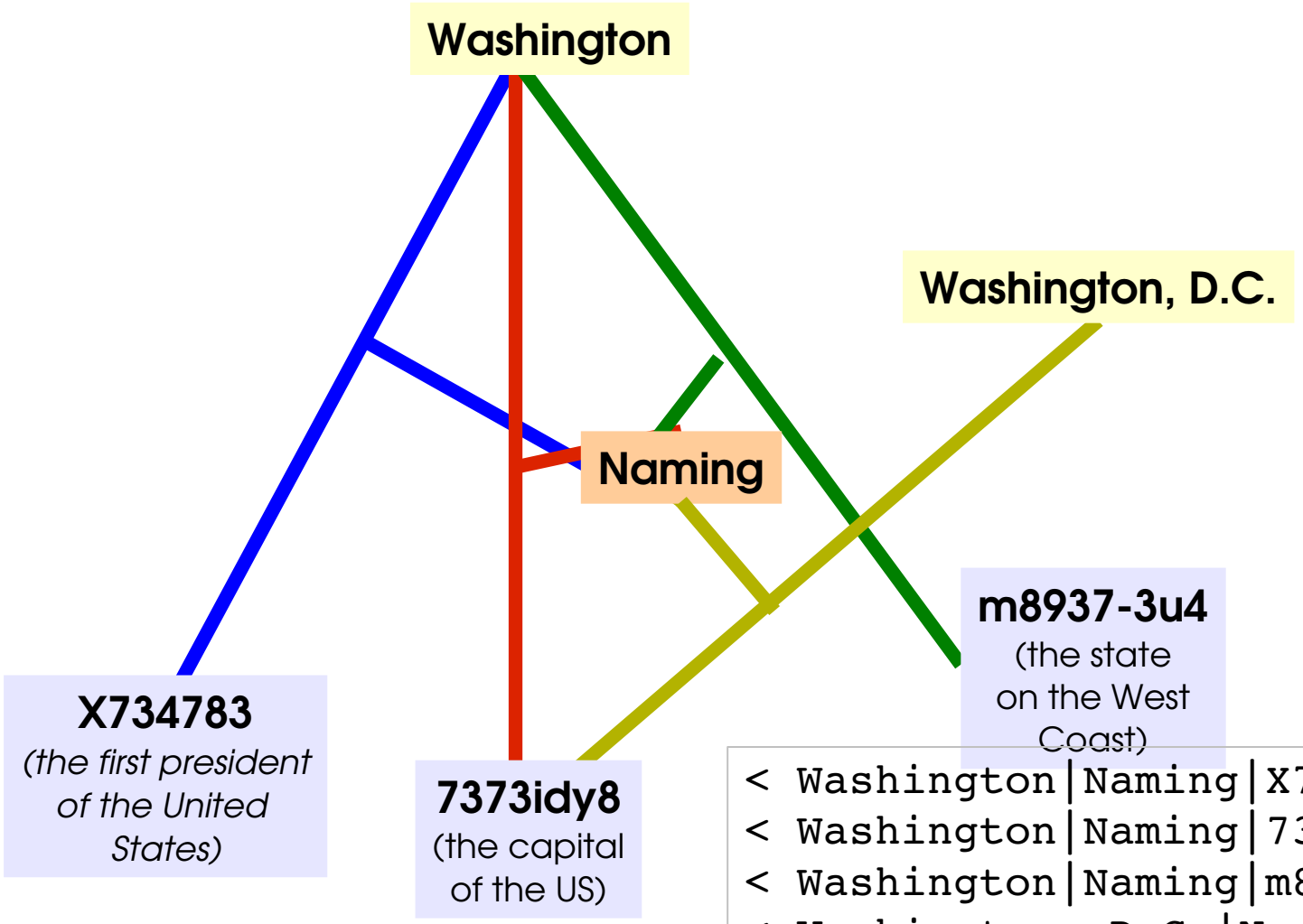
x operand, operator, *y* operand

Example : < 2 | + | 3 >

2nd step: rebuild views integrating the components.

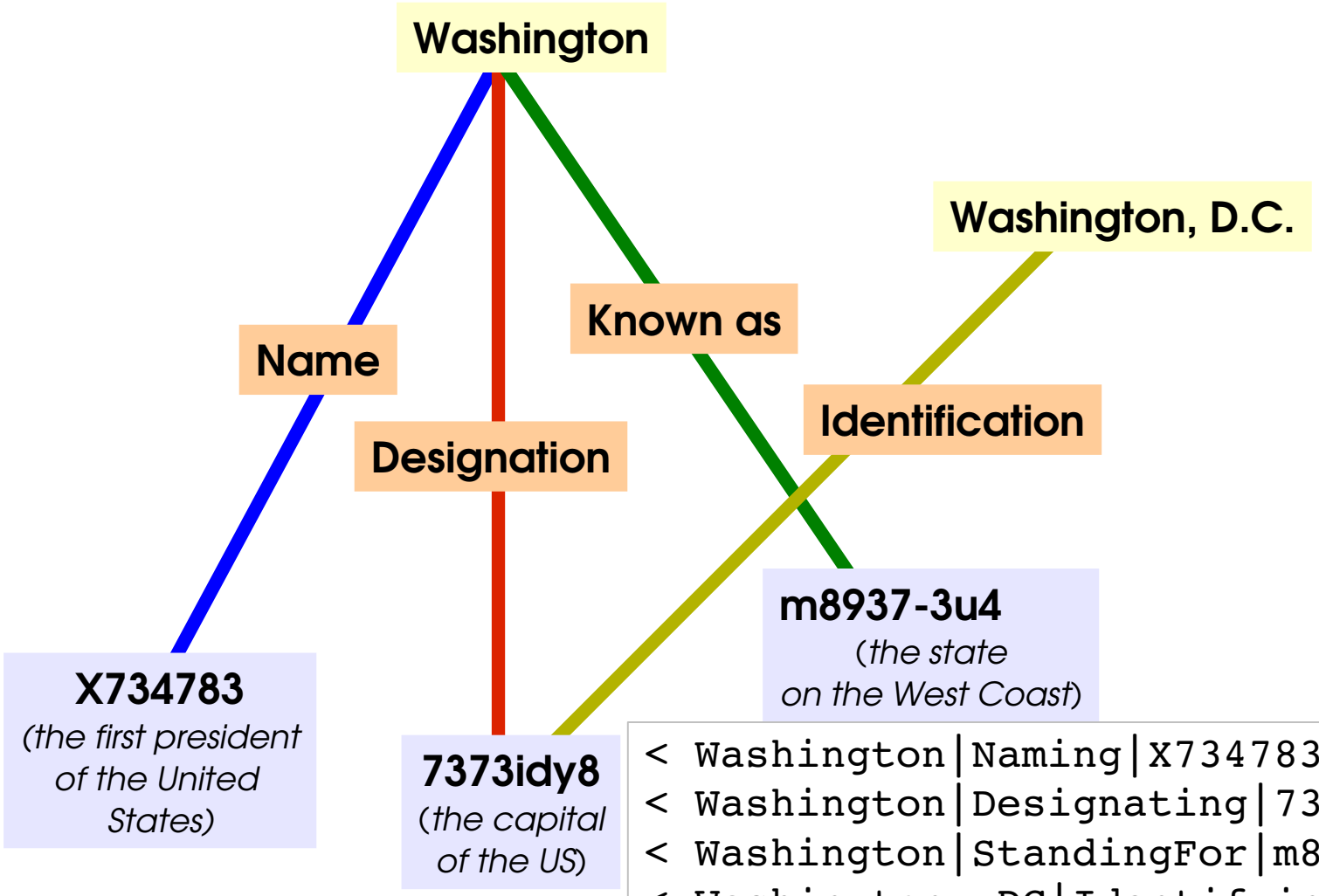
- Each view conforms to a given perspective.
- Multiple perspectives possible.

Example: Naming as Perspectives




```
< Washington|Naming|X734783 >
< Washington|Naming|7373idy8 >
< Washington|Naming|m8937-3u4 >
< Washington, D.C.|Naming|7373idy8 >
```

Multiple “Naming” Perspectives




```
< Washington|Naming|X734783 >
< Washington|Designating|7373idy8 >
< Washington|StandingFor|m8937-3u4 >
< Washington, DC|Identifying|7373idy8 >
```

Data Projection Demos

Demo 1: Simple perspectors:

http://www.infoloom.com/dpm3.html#demo1

Demo 2: Multiple Names

http://www.infoloom.com/dpm3.html#demo2

Topic Maps and RDF, revisited

- Common core: express graphs of relations.
- In Topic Maps, names are different from the subjects they designate. (Names themselves can be subjects.)
- An RDF statement can be unfolded into many RDF statements. There is a potential to do this in many RDF applications.
- An RDF graph can express a Topic Map.

What's next?

- Distinguish Connections from Semantics
- Distinguish Connected items from Processes operating upon them
- Resulting information models are transparent and auditable.
- Information can be expressed using multiple perspectives, and can be retrieved using other perspectives.