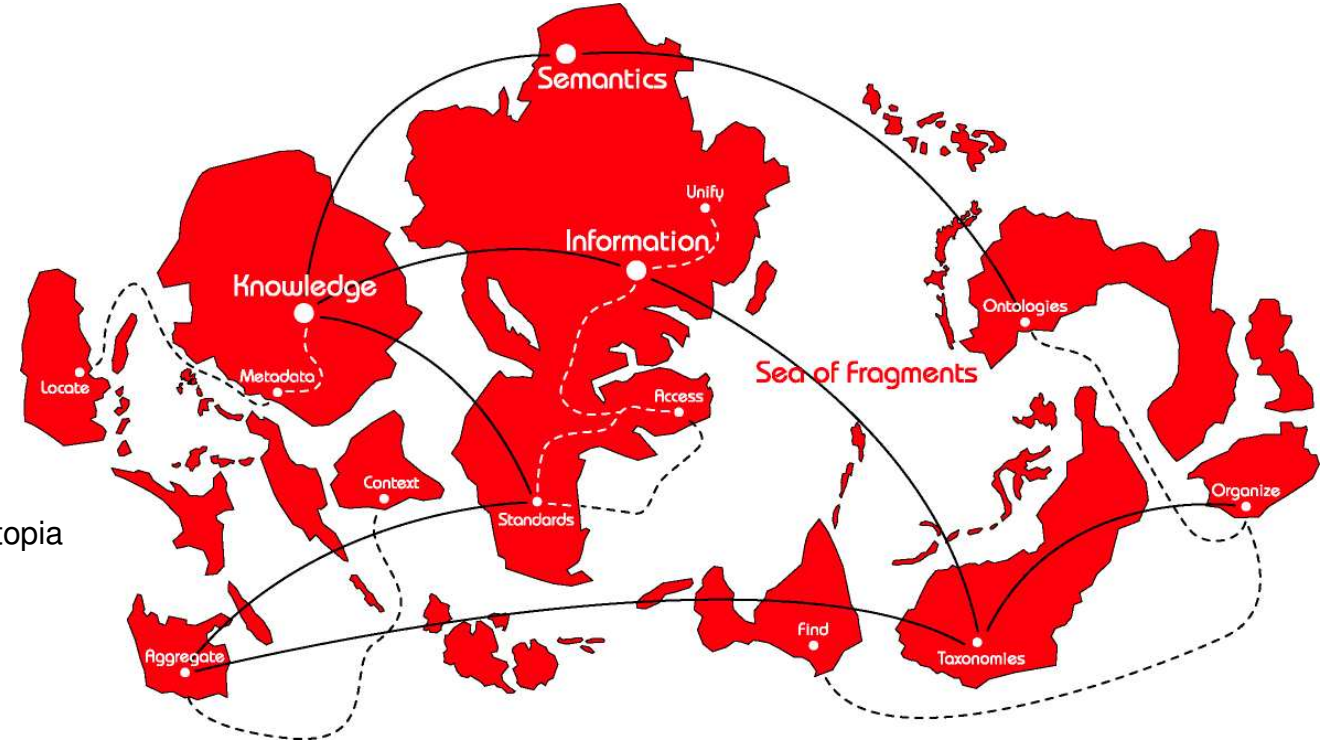
A Foundational Model for Topic Maps

Proposal for Reference Model Workshop, Montréal, August 2004



Lars Marius Garshol
Development Manager, Ontopia
<larsga@ontopia.net>

The model

Why this proposal?

- Well, basically because I had the idea, and it seemed good
- At the time it seemed to do better than existing proposals at meeting *my* goals
- And, since we're having a workshop on the subject, it seemed a good idea to present the proposal there

What goals?

- **It should be suitable as a common foundation for TMCL and TMQL**
  - it should be simpler than TMDM
  - it should be able to fully represent TMDM without loss of information
- **It should be sufficiently formal to appeal to an academic audience**

It's all quadruples

- **Instances of the model consist of quadruples**
  - (subject proxy, property, statement-id, object)
- **Mathematically,**
  - **I** is the set of all identifiers (which have no properties other than that one can be distinguished from another)
  - **L** is the set of all literals (which are strings), and
  - instances of the model are sets consisting of tuples
    $\langle s, p, q, o \rangle \in \mathbf{S} = (\mathbf{I} \times \mathbf{I} \times \mathbf{I} \times (\mathbf{I} \cup \mathbf{L}))$
  - for any given model M, neither of the following can hold
    - $\exists \langle s, p, q, o \rangle \langle x, y, z, w \rangle \mid q = z \wedge (s \neq x \vee p \neq y \vee o \neq w)$
    - $\exists \langle s, p, q, o \rangle \langle x, y, z, w \rangle \mid s = x \vee p = y \vee o = w \vee q \neq z$
- **In English,**
  - identifiers are just identifiers, local to each model,
  - different statements can't have the same identifier,
  - there can't be two equal statements with different identifiers

ontopia

That's the whole thing...

Type system

- Well, in practice we'll need a type system as well
- That is, a subdivision of L with properties and interpretations for the various subsets
- We don't need that right now, however, and so will skip it

TMDM representation

Representing TMDM

- **The point of this exercise, of course, is to represent TMDM and to allow us to define operations on it**
- **In the following we use these conventions:**
  - UPPER_CASE_IDENTIFIERS are identifiers used to model TMDM
  - lower_case_identifiers are identifiers for instance objects/nodes
  - “strings” are literals

Starting simple

```
<topicMap xmlns="...”>
  <topic id="lmg"/>
</topicMap>
```

(t1, SRCLOC, s1, “file://...#lmg”)

Adding a base name

```
<topicMap xmlns="...”>
  <topic id="Img">
    <baseName>
      <baseNameString>Lars Marius
     Garshol</baseNameString>
    </baseName>
  </topic>
</topicMap>
```

(t1, SRCLOC, s1, “file://...#Img”)

(t1, TOPNAME, s2, “Lars Mariu...”)

Adding a variant name

```
<topicMap xmlns="...”>
  <topic id="lmg">
    <baseName>
      <baseNameString>Lars Marius
     Garshol</baseNameString>
      <variant>
        <parameters>
          <subjectIndicatorRef sort...
        </parameters>
        <variantName>
          <resourceData>garshol...
        </variantName>
      </variant>
```

(t1, SRCLOC, s1, “file://...#lmg”)
(t1, TOPNAME, s2, “Lars Mariu...”)
(s2, VARIANT, s3, “garshol, ...”)
(s3, SCOPE, s4, t2)
(t2, SUBJID, s5, “http://www.to...”)

Handling occurrences

```
<topicMap xmlns="...”>
  <topic id="Img">
    <occurrence>
      <instanceOf>
        <subjectIndicatorRef desc...
        <resourceData>Description...
    </occurrence>
  </topic>
</topicMap>
```

(t1, SRCLOC, s1, “file://...#Img”)
(t1, t2, s2, “Description...”)
(t2, SUBJID, s3, “http://www.on...”)

Associations

```
<topicMap xmlns="...”>
 <association>
   <instanceOf>
    <subjectIndicatorRef emp-by...
   </instanceOf>
   <member>
    <roleSpec>
     <subjectIndicatorRef empr...
    </roleSpec>
    <topicRef href=”#ontopia"/>
   </member>
   <member>
    <roleSpec>
     <subjectIndicatorRef empe...
    </roleSpec>
    <topicRef xlink:href="#Img"/>
   </member>
  </association>
</topicMap>
```

(a1, TYPE, s1, t1)
(t1, SUBJID, s2, “http://...#emp-by”)
(a1, t2, s3, t3)
(t2, SUBJID, s4, “http://...#empr”)
(t3, SRCLOC, s5, “...#ontopia”)
(a1, t4, s6, t5)
(t4, SUBJID, s7, “http://...#empe”)
(t5, SRCLOC, s8, “...#Img”)

Operations

Subscripting

- For $s \in S$, s[n] = the nth element of the quadruple
- For $m \in M$, $m[n] = \{x \mid x = s[n] \wedge s \in M\}$
- That is,
  - subscripting can be applied to a quadruple, or to a model
  - in the first case it yields a single value, in the second a set

Filtering

- **For m ∈ M, s, p, q ∈ I ∪ *, and o ∈ I ∪ L ∪ *,**
- **Φ(m, s, p, q, o) = {z** |
  - z ∈ M
  - s = * or s = z[1]
  - p = * or p = z[2]
  - q = * or q = z[3]
  - o = * or o = z[4]
- }

Variable bindings

- V = set of all variables
- A *variable binding* is a tuple (v, w) ∈ V x (I ∪ L)
- A *match* is a set of variable bindings
- A *result set* is a sequence of matches
  - <{(v1, w1), (v2, w2), ...}, ...>

Problems

Not a reference model

- **The TMDM mapping here does not provide conceptual guidance**
- **E.g., it's not clear how occurrences are related to associations**
- **This is not here because**
  - it wasn't a goal,
  - meeting this goal requires complicating things

Doesn't do RDF interoperability

- **In this model there is still a difference between RDF and TMDM**
- **Again, this is because**
  - it was not a goal,
  - meeting this goal requires complicating things

No type system

- **Necessary, but not at this time**

Operations are too simple

- Can't handle AND, OR, NOT, ...
- No support for variable bindings yet
- This is necessary, but not completed yet