# *Bridging the Gaps:*
# *Adaptive Approaches to Data Interoperability*

Michael K. Bergman


keynote presentation at

DC-2010 Conference

Pittsburgh, PA

*October 22, 2010*

# *The Iconoclast Cometh*

# Outline of Talk

- Linked Data
- Data Web, Structured Data and Semantic Web
- Players and Roles → DCMI
- Conclusions

# Three Overall Assertions

<LinkedData> <isA> <Valuable*Technique*>

<DataWeb> <hasNeedOf> <Semantics>

<DCMI> <hasRole> <Unique>

# Linked Data

# Three Linked Data Assertions

<LinkedData> <isA> <PreferredTechnique>

<Techniques> <doNotSolve> <RootChallenges>

<RDF> <hasBestRoleAs> <CanonicalDataModel>

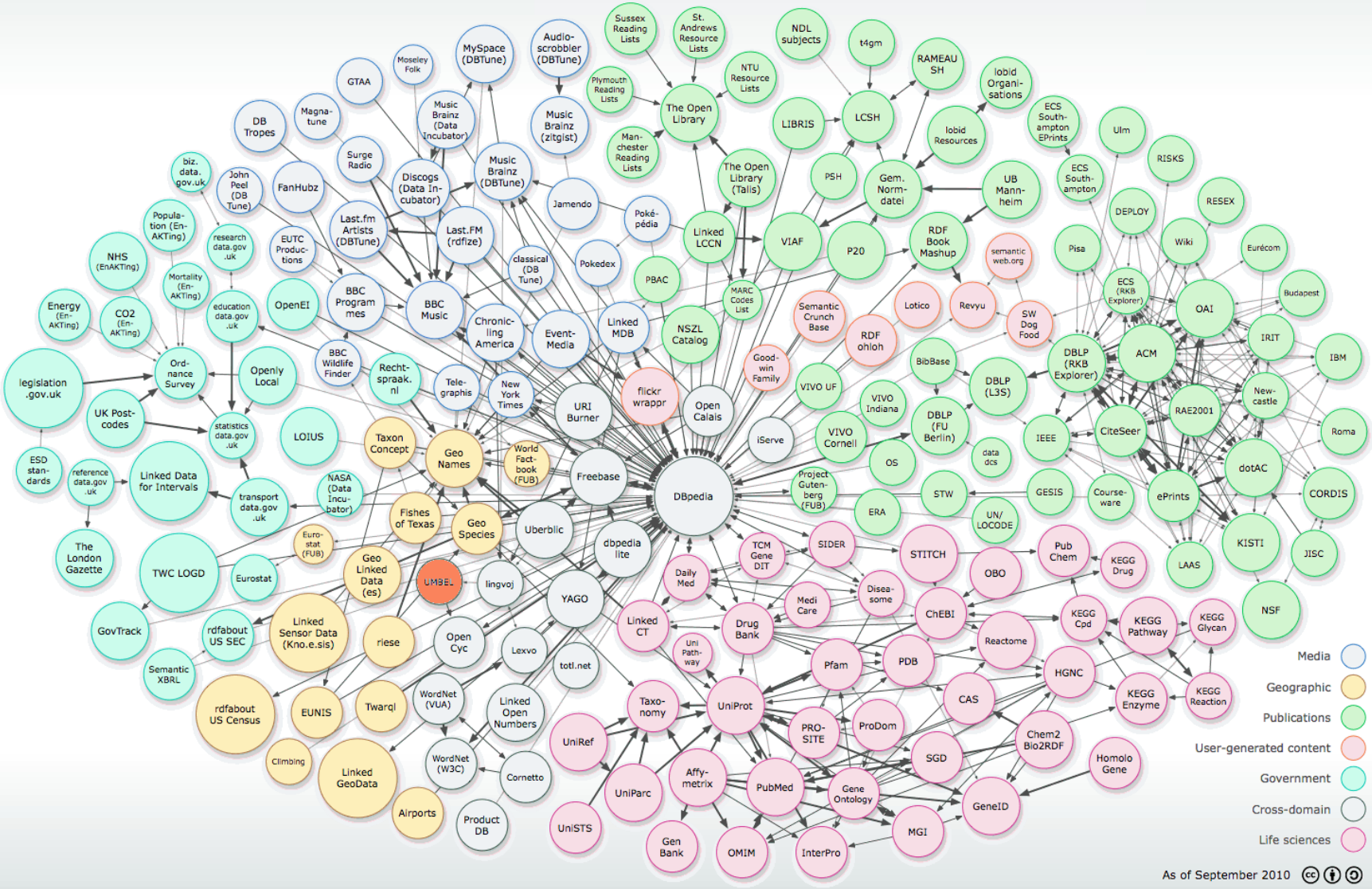
# Three More Linked Data Assertions

<LinkedData> <hasGrowing> <Triples>

<LDUsers> <wronglyUse> <ManyPredicates>

<LinkedData> <hasLack> <MajorUptake>

# 25 Billion Linked Data Triples



# Bad Results from sameAs Misuse

# The State of Linked Data

- **Growing, but not as fast as promise would suggest**
- **Not used much, except curated settings**
- **Few actual dataset linkages**
- ***NO* true interoperability, except curated** (life science, some others)
- **Difficult to publish**
- **If done right, best form to consume**

# Data, Structure and Semantic Web

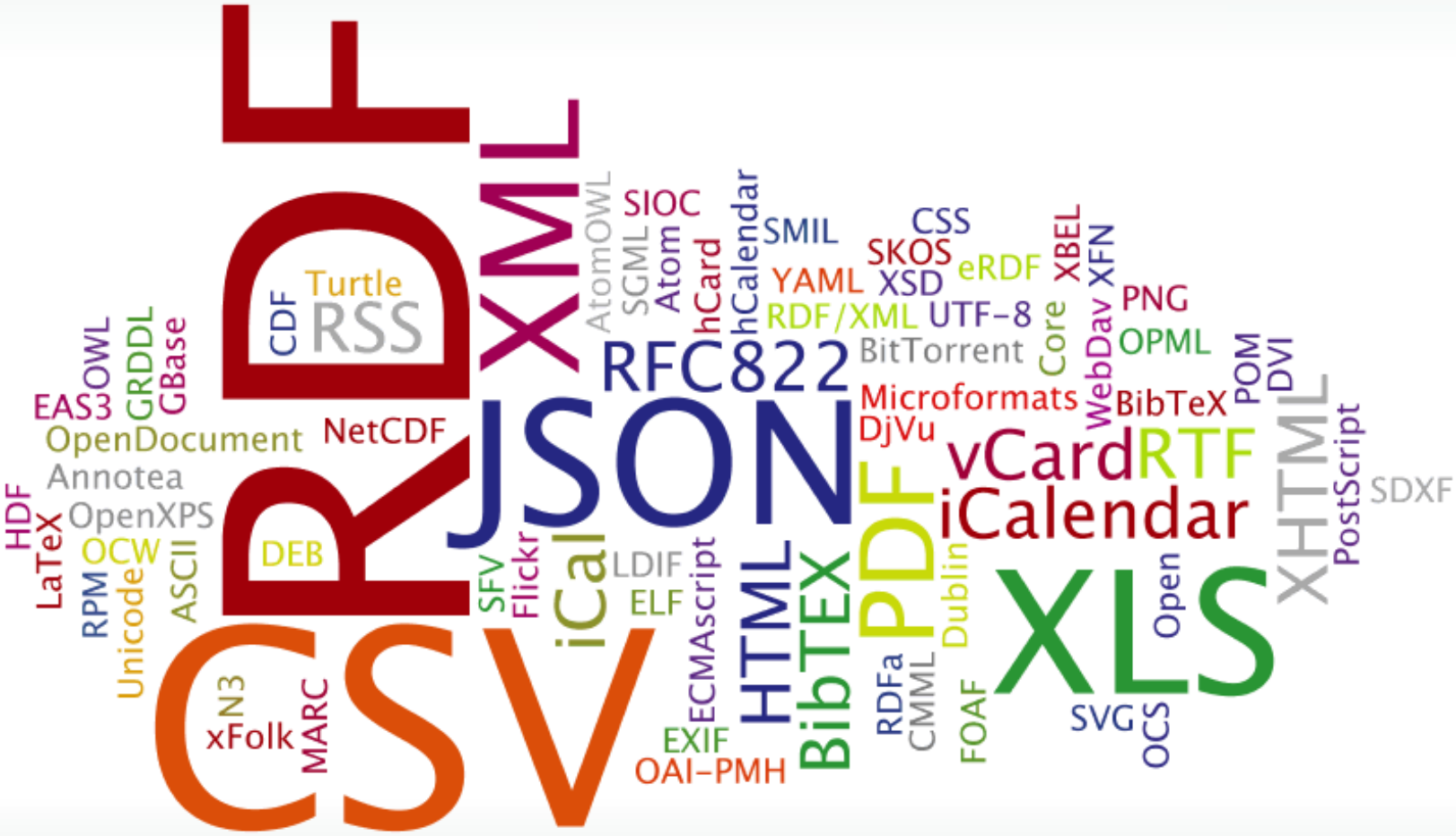
# Three Structured Data Web Assertions

<Heterogeneity> <isA> <Reality>

<LinkedData> <isOnly> <TinyContributor>

<Semantics> <isThe> <MissingLink>

# Hundreds of Formats in the Wild



# How to Aliquot the Firehose ?

# Three Semantics Assertions (+ Axiom)

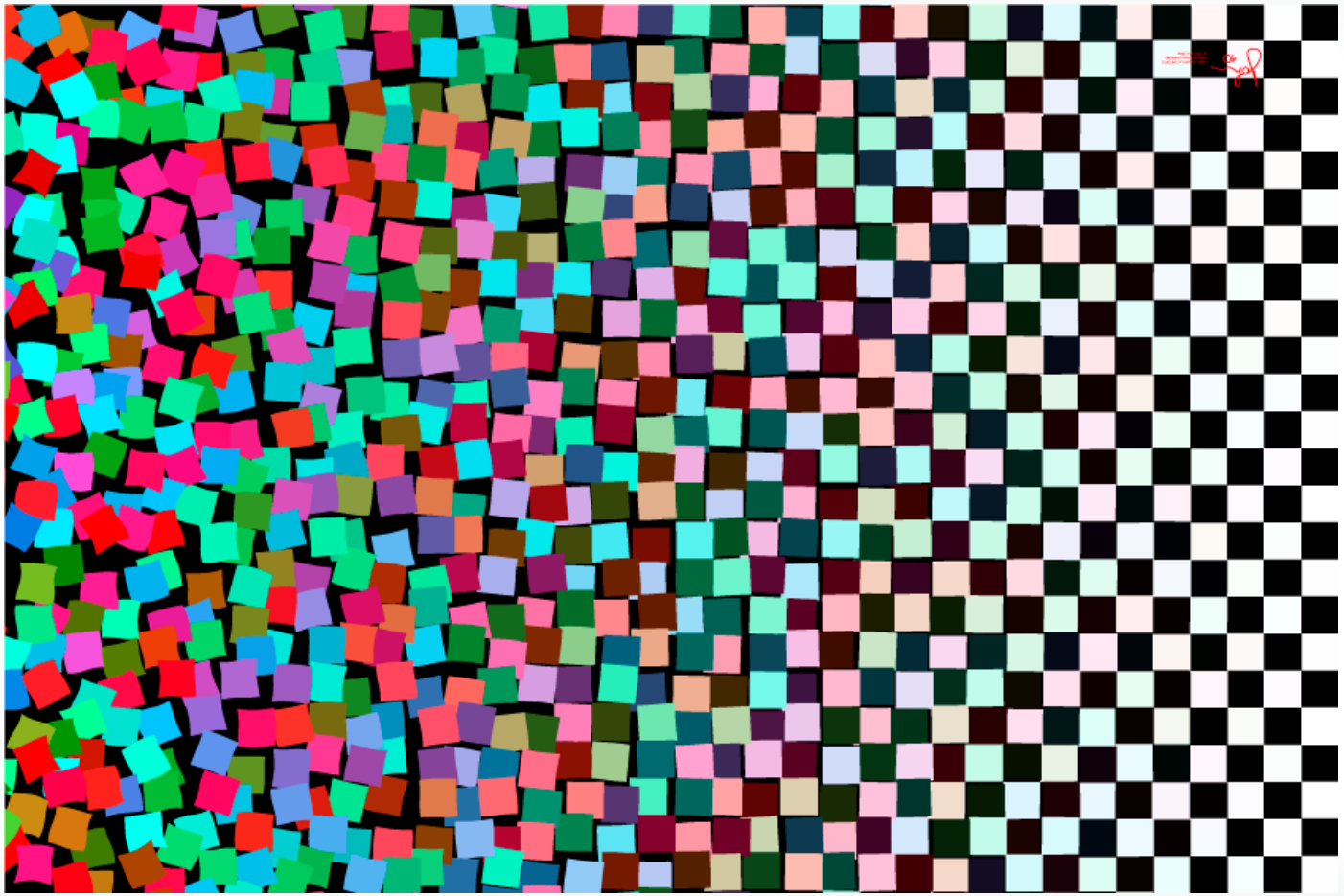
<ReferenceVocabs> <organize> <MassiveContent>

<LinkingPredicates> <gather> <RelatedContent>

<intersectionOf>

<SemanticContent> <enables> <MeaningfulWork>

# Reference Vocabs Help Organize Chaos



# Fixed References Help Orient

# Concepts are the Fixed References

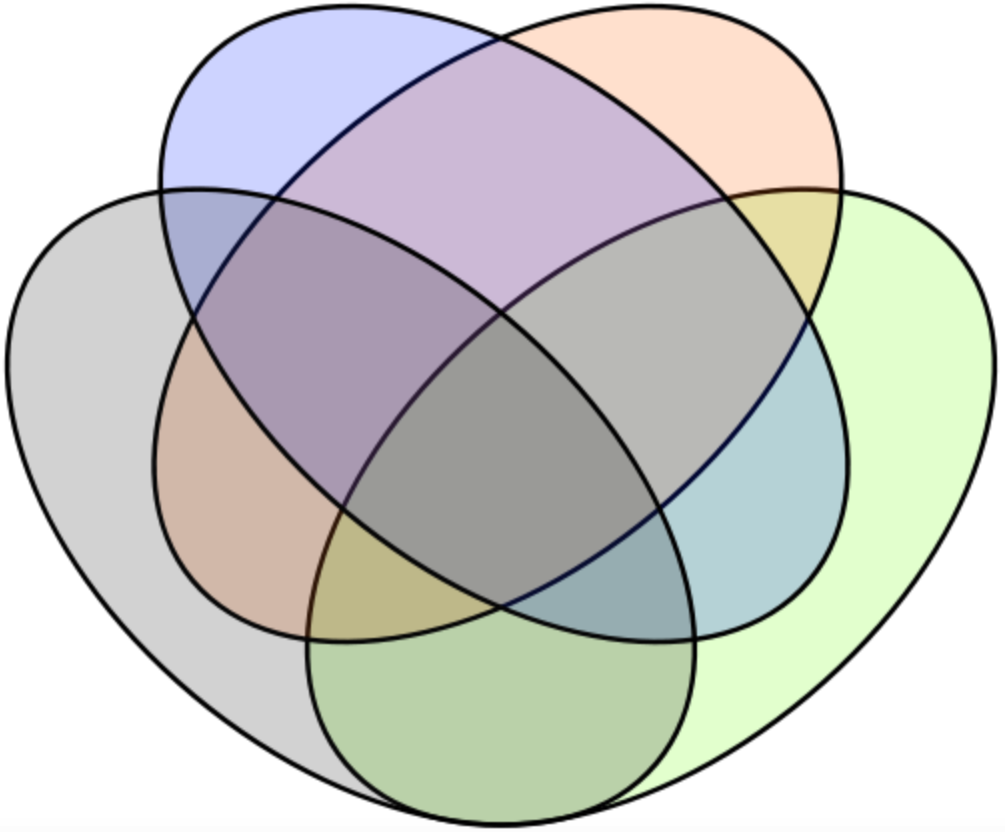
# Design Aspects of Reference Concepts

- **Truly are *concepts*, the idea of a thing**
- **Labels are language independent** (à la SKOS)**:**
  - ✓ Preferred, human-readable label (prefLabel)
  - ✓ Many, alternate synonyms, jargon, etc. (altLabel)
  - ✓ Misspellings (hiddenLabel)
  - → all combined for tagging, IE purposes
- ***MUST* have definition: what does this concept *mean* ?**
- **Organized into coherent structures** (graphs)
  - ✓ Inferencing
  - ✓ Discovery and navigation
- **Act as both classes and instances** (RDF / OWL-speak)
- ***MUST* have persistent URIs**

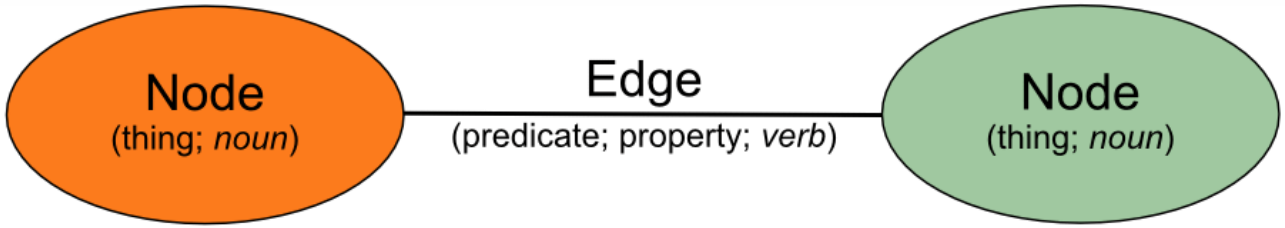
# Mappings Get Stuff into the Right Room

# Many Mappings Should be Approximate

- skos:broadMatch
- skos:related
- ore:similarTo
- umbel:isAbout
- vmf:isInVocabulary
- skos:closeMatch
- lvont:nearlySameAs
- umbel:isLike
- umbel:hasCharacteristic
- lvont:somewhatSameAs
- rdfs:seeAlso
- ore:describes
- map:narrowerThan
- skos:narrower
- map:broaderThan
- skos:broader
- dc:subject
- link:uri
- foaf:isPrimaryTopicOf



# Some Conditions for Interoperability




<Interoperability> <needsMapping> <Predicates>

<Interoperability> <needsReference> <Nouns>

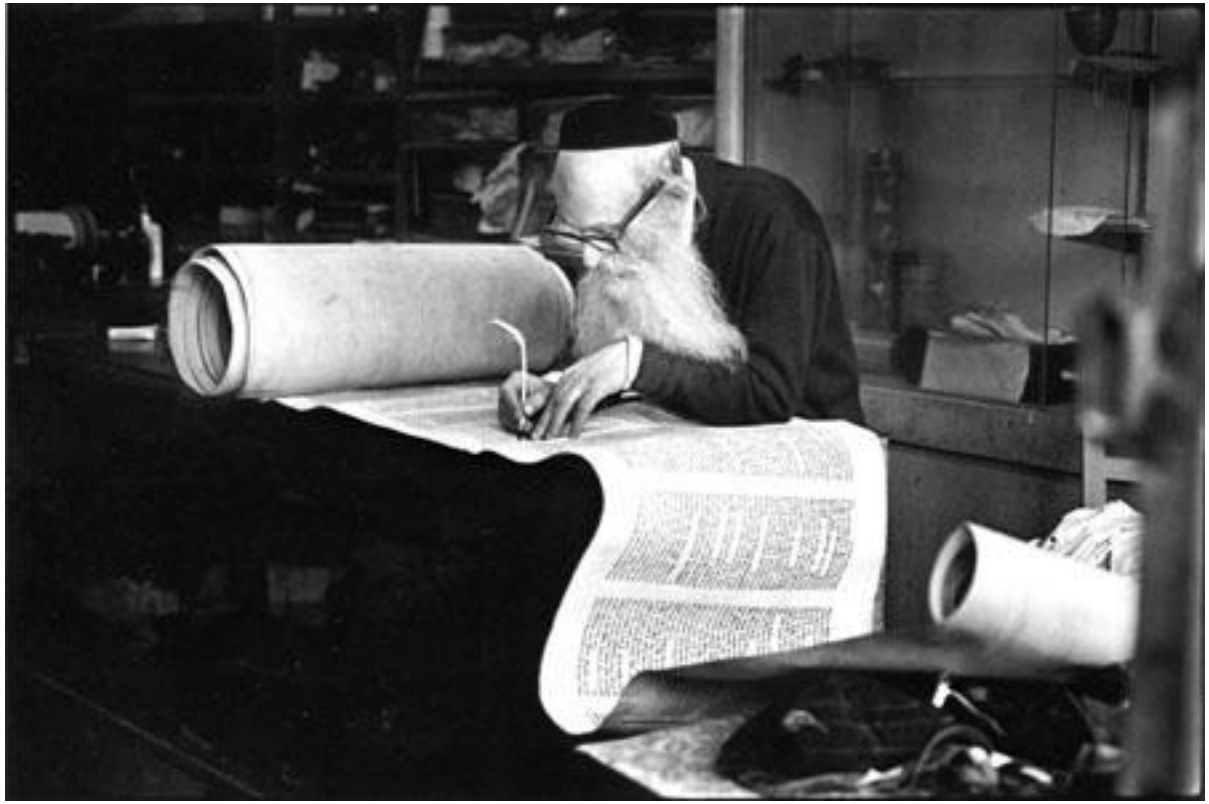
# Three Major Players

# World Role

<World> <hasRole> <ContentAndStructure>

# W3C Role

<W3C> <hasRole> <Standards>



# DCMI Role

<DCMI> <hasRole> <ReferenceMetadata>

# Three Going Forward Assertions

<LinkedData> <hasNeedOf> <MapPredicates>

<DataWeb> <hasNeedOf> <ReferenceConcepts>

<DCMI> <hasUniqueRole> <BothRequirements>

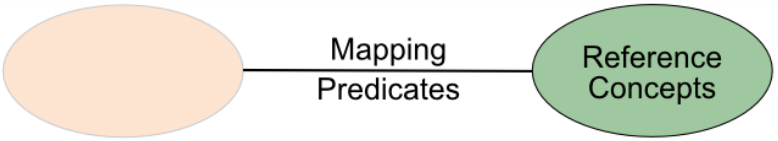
# DCMI: the Unique Franchise

- **DCMI already has unique *authority* in:**
  1. dc:subject
  2. dc:subject qualifiers
  3. initial Open Registry effort
  4. core foundational properties
- **DCMI has unique *experience* in:**
  1. diverse vocabularies
  2. cataloging and classification
  3. semantics

# Reference Authority - Needed DCMI Role

<RefMetadata> <notSameAs> <OneRingRulesAll>

# Reference Metadata is Not a Third Rail




# The Web is Parched for Semantics

- Reference vocabularies
- Persistent URIs
- Re-use of vocabs
- Vetting ranking
- Alignment services (?)
- Annotation services (?)
- RDFa injection (?)
- Open source frameworks

# We're also Ready to Help

ontotext + STRUCTURED DYNAMICS

umbel + proton ontology +  + ???

# A First Exemplar: FactForge

- **A "*reason-able*" view to linked open data**
  - ✓ Pre-loaded semantic repository: reasoning, querying, exploration
- **Ontologies**
  - ✓ Dublin Core, SKOS, RSS, FOAF
- **Datasets**
  - ✓ DBpedia, Freebase, Geonames, UMBEL, MusicBrainz, Wordnet, CIA World Factbook, Lingvoj
- **Very large scale**
  - ✓ 1.2B explicit + 0.9B inferred → 10B retrievable statements
  - ✓ Managed by BigOWLIM
- **Free public service with many features:**
  - ✓ Auto-suggest
  - ✓ Query and explore through Forest, RelFinder and Tabulator
  - ✓ RDF search
  - ✓ SPARQL end-point

# Next Step, RENDER

- **New EU project**
- **Large-scale LOD interoperability, methods**
- **Players:**
  - ✓ Ontotext
  - ✓ Google (EU)
  - ✓ Wikimedia (DE)
  - ✓ Structured Dynamics
- **Generalizing the FactForge approach**
- **Basis for possible follow-ons ??**

# Possible Ontotext + SD Contributions

1. **Mapping services to all comers** (“vocabulary neutrality”)
2. **Tagging services**
3. **Software + systems for other tagging services**
4. **Possible technical support for Metadata Registry**
5. **Lead / support for possible EU grant-seeking efforts**

↓↓↓

***If DCMI willing to partner, Ontotext + SD willing to contribute in a neutral, open source manner***

# Ontotext + SD Links

- **FactForge**

  http://www.factforge.net

- **PROTON**

  http://proton.semanticweb.com

- **Ontotext**

  http://www.ontotext.com

- **UMBEL**

  http://www.umbel.org

- **Structured Dynamics**

  http://structureddynamics.com

# Conclusion

# Main Assertions Re-visited

- **Interoperability on the Web *not* working:**
  1. Not (generally) fulfilled by linked data in current state
  2. Predicates for approximate mappings lacking
  3. Reference vocabularies essential as connecting nodes
- **DCMI is the best** (only?) **player to plug these gaps**
- **We are willing to help find the resources + right process to help plug the interoperability gap**

# DCMI Interoperability Services ?

# Q & A

STRUCTURED
DYNAMICS

# Contacts & Information

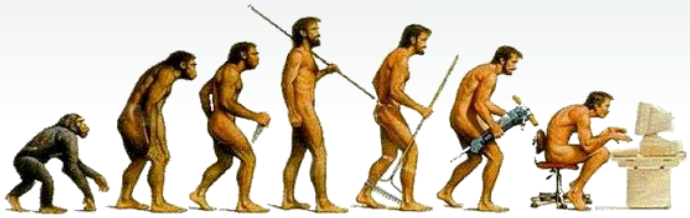

**Michael K. Bergman**

**CEO**

319.621.5225

mike@structureddynamics.com

blog: www.mkbergman.com

## Web Sites

structureddynamics.com

citizen-dan.org (community indicator systems)

openstructs.org (open source software)

techwiki.openstructs.org (open license technical documentation)

umbel.org

umbel.structureddynamics.com (UMBEL Web services)


STRUCTURED
DYNAMICS