The Web and SKOS

Alistair Miles

ISKO

London, July 2008

Outline

- SKOS in 5 minutes
- The Web and SKOS
- Design issues

SKOS IN 5 MINUTES

Term: Economic cooperation

Used For:
Economic co-operation

Broader terms:
Economic policy

Narrower terms:
Economic integration
European economic cooperation
European industrial cooperation
Industrial cooperation

Related terms:
Interdependence

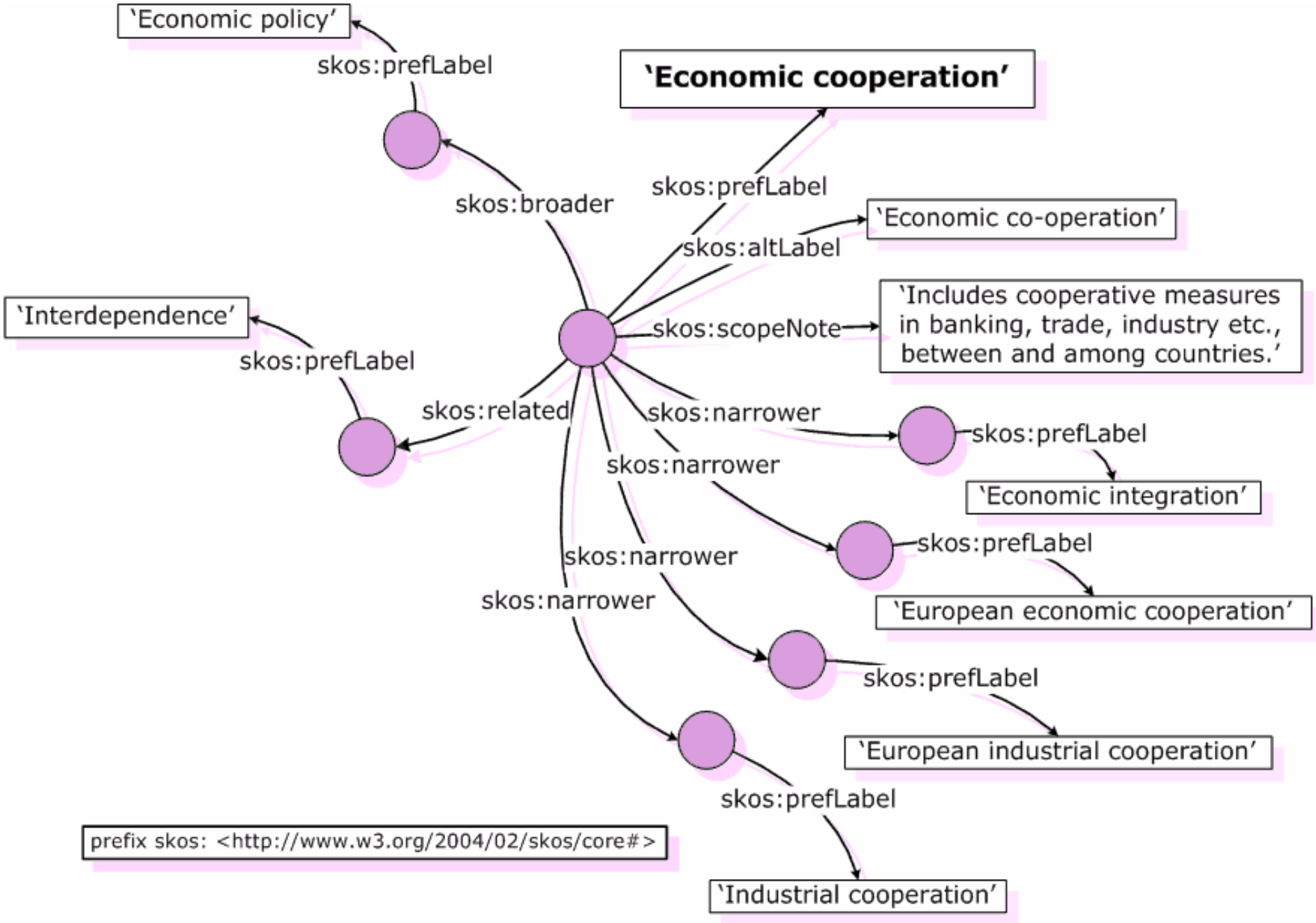
Scope Note:
Includes cooperative measures in banking, trade, industry etc., between and among countries.

Source: http://www.ukat.org.uk

- SKOS is a data model
- Use SKOS to represent a KOS as data
- Those data can be serialised...
- Those data can also be rendered as a graph...

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:skos="http://www.w3.org/2004/02/skos/core#">

  <skos:Concept rdf:about="http://www.ukat.org.uk/thesaurus/concept/1750">
    <skos:prefLabel>Economic cooperation</skos:prefLabel>
    <skos:altLabel>Economic co-operation</skos:altLabel>
    <skos:scopeNote>Includes cooperative measures in banking, trade, industry etc., between
and among countries.</skos:scopeNote>
    <skos:broader rdf:resource="http://www.ukat.org.uk/thesaurus/concept/4382"/>
    <skos:narrower rdf:resource="http://www.ukat.org.uk/thesaurus/concept/2108"/>
    <skos:narrower rdf:resource="http://www.ukat.org.uk/thesaurus/concept/9505"/>
    <skos:narrower rdf:resource="http://www.ukat.org.uk/thesaurus/concept/15053"/>
    <skos:narrower rdf:resource="http://www.ukat.org.uk/thesaurus/concept/18987"/>
    <skos:related rdf:resource="http://www.ukat.org.uk/thesaurus/concept/3250"/>
    <skos:inScheme rdf:resource="http://www.ukat.org.uk/thesaurus"/>
  </skos:Concept>

</rdf:RDF>
```

Source: http://www.ukat.org.uk/downloads/data.php


'Economic policy'
skos:prefLabel
'Economic cooperation'
skos:broader
skos:prefLabel
'Economic co-operation'
skos:altLabel
'Interdependence'
skos:prefLabel
skos:scopeNote
'Includes cooperative measures in banking, trade, industry etc., between and among countries.'
skos:related
skos:narrower
skos:prefLabel
'Economic integration'
skos:narrower
skos:narrower
skos:prefLabel
'European economic cooperation'
skos:narrower
skos:prefLabel
'European industrial cooperation'
skos:prefLabel
'Industrial cooperation'
prefix skos: <http://www.w3.org/2004/02/skos/core#>

Those data can be published in the Web...

...linked with other data in the Web...

...shared between software applications...

FlyWeb Research Tools version 20080701-alpha

Search FlyTED and BDGP Images By Gene

CG10151 Go

Enter a gene name, e.g. "CG10151", "CG10253" or "Aos1" ... (what gene names can I use?)

Image results from FlyTED (www.fly-ted.org)

Found 2 results for gene "CG10151".

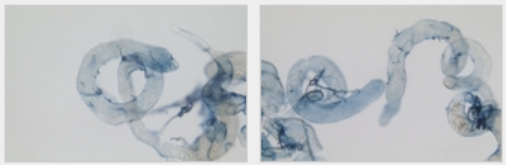
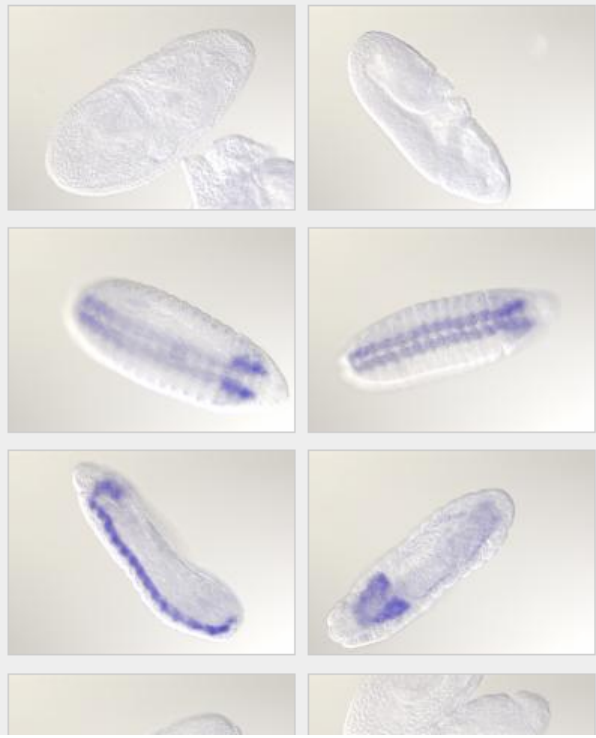


Image results from BDGP (www.fruitfly.org)

Found 10 results for gene "CG10151".



Source: http://rodos.zoo.ox.ac.uk/flyweb/search/flyted-bdgp/

- **SKOS Primer**
  - http://www.w3.org/TR/skos-primer
- **SKOS Reference**
  - http://www.w3.org/TR/skos-reference
- About to publish **Last Call Working Draft**
- **If there are issues, raise them now!**
- **public-swd-wg@w3.org – comments**
- **public-esw-thes@w3.org – informal discussion**

THE WEB AND SKOS

- SKOS is not trying to re-invent knowledge organisation...
- ...rather, provide a framework for porting KOS to a shared space: **the Web**
- SKOS born of...
  - Knowledge Organisation
  - The Web
- SKOS is being developed as a **Web Standard** (i.e. W3C Recommendation)

- Testimony of **Sir Timothy Berners-Lee**
  CSAIL Decentralized Information Group
  Massachusetts Institute of Technology
- Before the **United States House of Representatives** Committee on Energy and Commerce Subcommittee on Telecommunications and the Internet
- http://dig.csail.mit.edu/2007/03/01-ushouse-future-of-the-web.html

The Web and SKOS

FOUNDATIONS OF THE WEB

I. Foundations of the Web

- *"The success of the World Wide Web, itself built on the open Internet, has depended on three critical factors:*
  1. ***unlimited links** from any part of the Web to any other;*
  2. ***open technical standards** as the basis for continued growth of innovation applications; and*
  3. ***separation of network layers**, enabling independent innovation for network transport, routing and information applications."*

I.A. Universal Linking: Anyone can connect to anyone...

- *"In simple terms, the Web has grown because it's **easy to write** a Web page and **easy to link** to other pages."*
- *"What makes it easy to create links ... is that there is **no limit** to the number of pages or number of links possible on the Web."*
- *"Adding a Web page requires **no coordination with any central authority**, and has an extremely low, often zero, additional cost."*

- *“Adding a page provides content, but* ***adding a link provide the organization, structure and endorsement*** *to information on the Web which* ***turn the content as a whole into something of great value****.”*

- *"The universality and flexibility of the Web's linking architecture has a unique capacity to* ***break down boundaries of distance, language, and domains of knowledge****."*
- *"These traditional barriers fall away because the cost and complexity of a link is unaffected by most boundaries that divide other media."*

- *“The Web's ability to allow people to forge links is why we refer to it as an **abstract information space**, rather than simply a network.”*

The Web and SKOS

THE FUTURE OF THE WEB

II. Looking Forward

- *"First, the Web will get better and better at helping us to **manage, integrate, and analyze data**."*
- *"Today, the Web is quite effective at helping us to publish and discover documents, but the individual information elements within those documents ... **cannot be handled directly as data**."*

- *“Today you can see the data with your browser, but **can't get other computer programs to manipulate or analyze it** without going through a lot of manual effort yourself.”*
- *“As this problem is solved, we can expect that Web as a whole to **look more like a large database or spreadsheet**, rather than just a set of linked documents.”*

II.A. Data Integration

- *"Locked within all of this data is **the key to knowledge** about how to cure diseases, create business value, and govern our world more effectively."*
- *"The good news is that a number of technical innovations...*
- *... **RDF which is to data what HTML is to documents**, and the Web Ontology Language (OWL) which allows us to express how data sources connect together ...*
- *... along with more openness in information sharing practices are **moving the World Wide Web toward what we call the Semantic Web**."*

- *“Progress toward better data integration will happen through use of the key piece of technology that made the World Wide Web so successful:* ***the link****.”*
- *“The power of the Web today, including the ability to* ***find the pages we're looking for****, derives from the fact that documents are put on the Web in standard form, and then* ***linked together****.”*

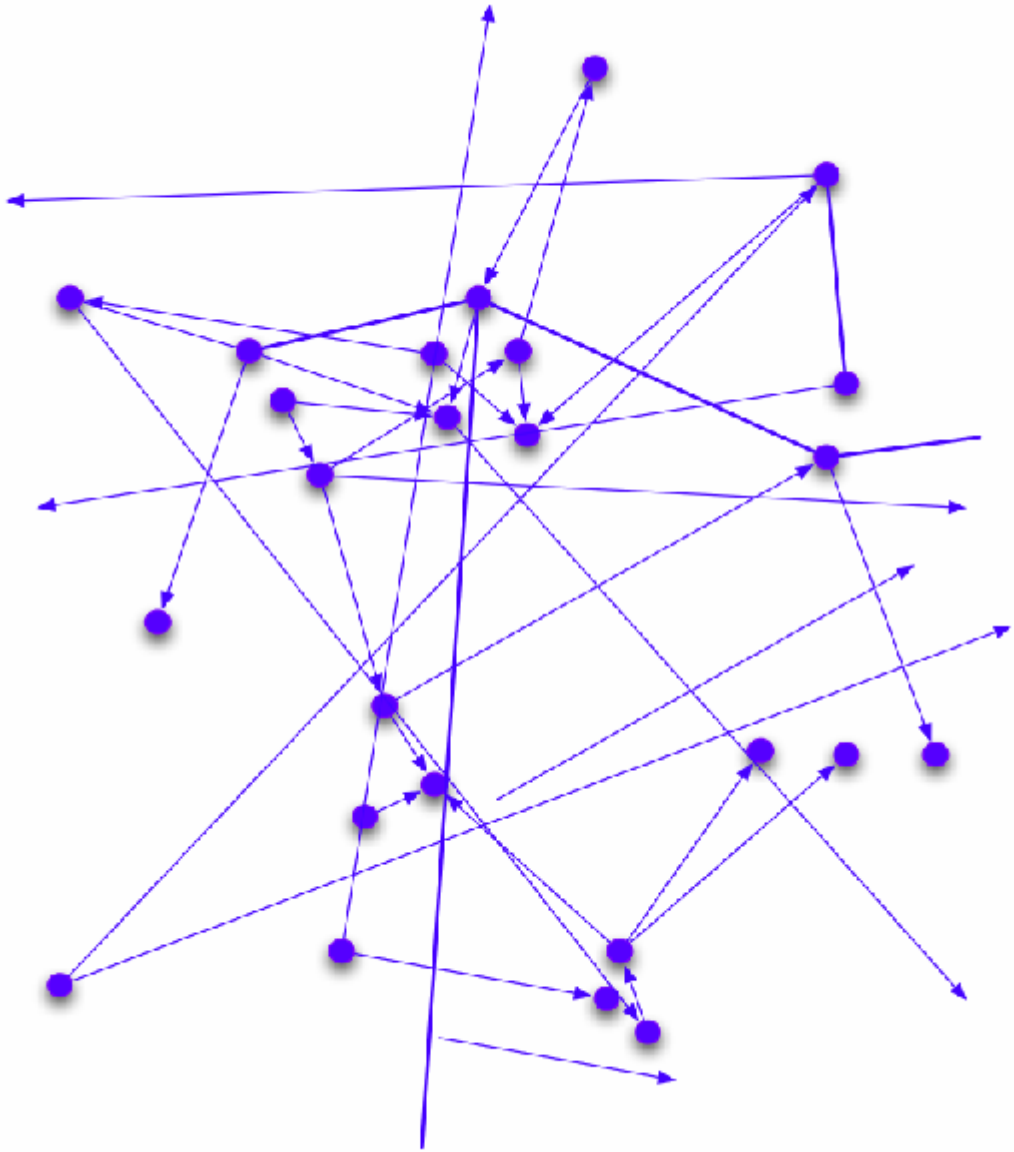
- *“The Semantic Web will enable better data integration by allowing everyone who puts individual items of data on the Web to* ***link them with other pieces of data*** *using standard formats.”*

The Web and SKOS

FROM THE WEB TO SKOS

- *"The success of the World Wide Web ... has depended on ... **separation of network layers**, enabling independent innovation for network transport, routing and information applications."*
- *"The Web's ability to allow people to forge links is why we refer to it as an **abstract information space**, rather than simply a network."*

http://www.w3.org/2007/Talks/1211-whit-tbl/



Layers in the Web

- http://www.w3.org/2007/Talks/1211-whit-tbl/#(23)
- Third layer is network (graph) of links between...
- ... people, organisations, diseases, genes, proteins, concepts ...
- ... capture these data **in the Web**
- **I.e. The Web as a platform for Linked Data**

KOS as Linked Data

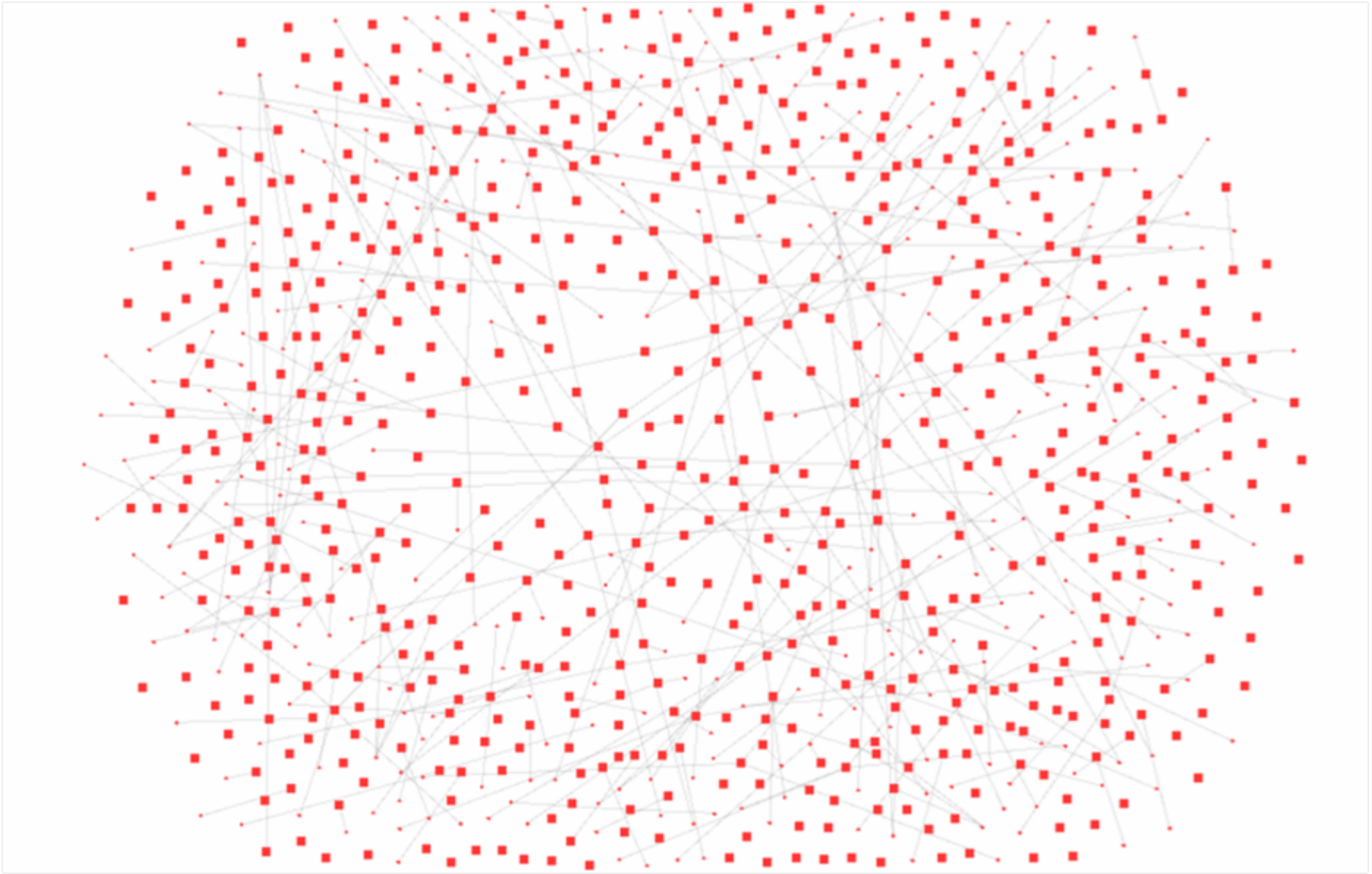
- Knowledge Organisation Systems can be viewed as a **network of linked concepts**
- e.g. Library of Congress Subject Headings

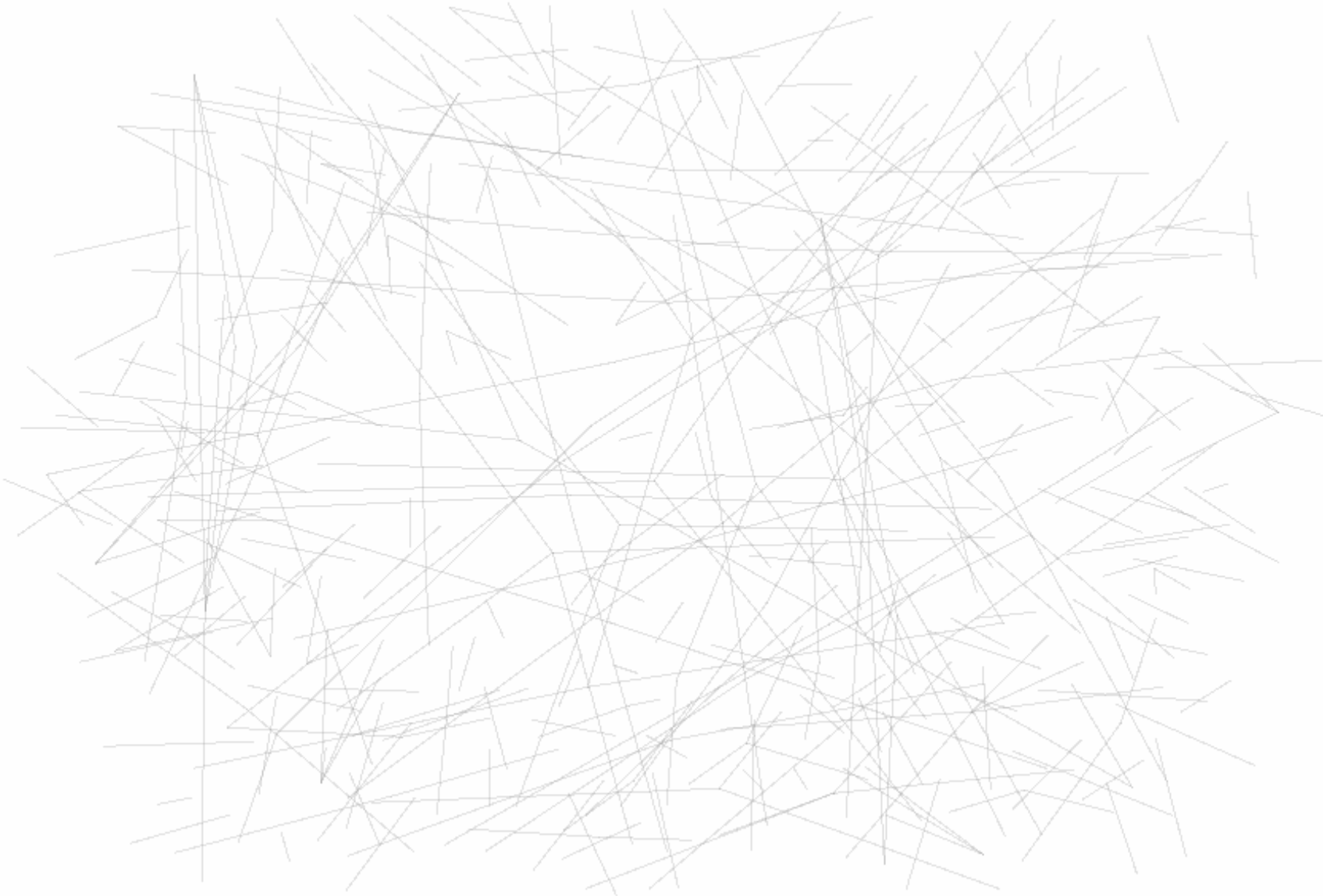
NT International trusteeships
Mandates

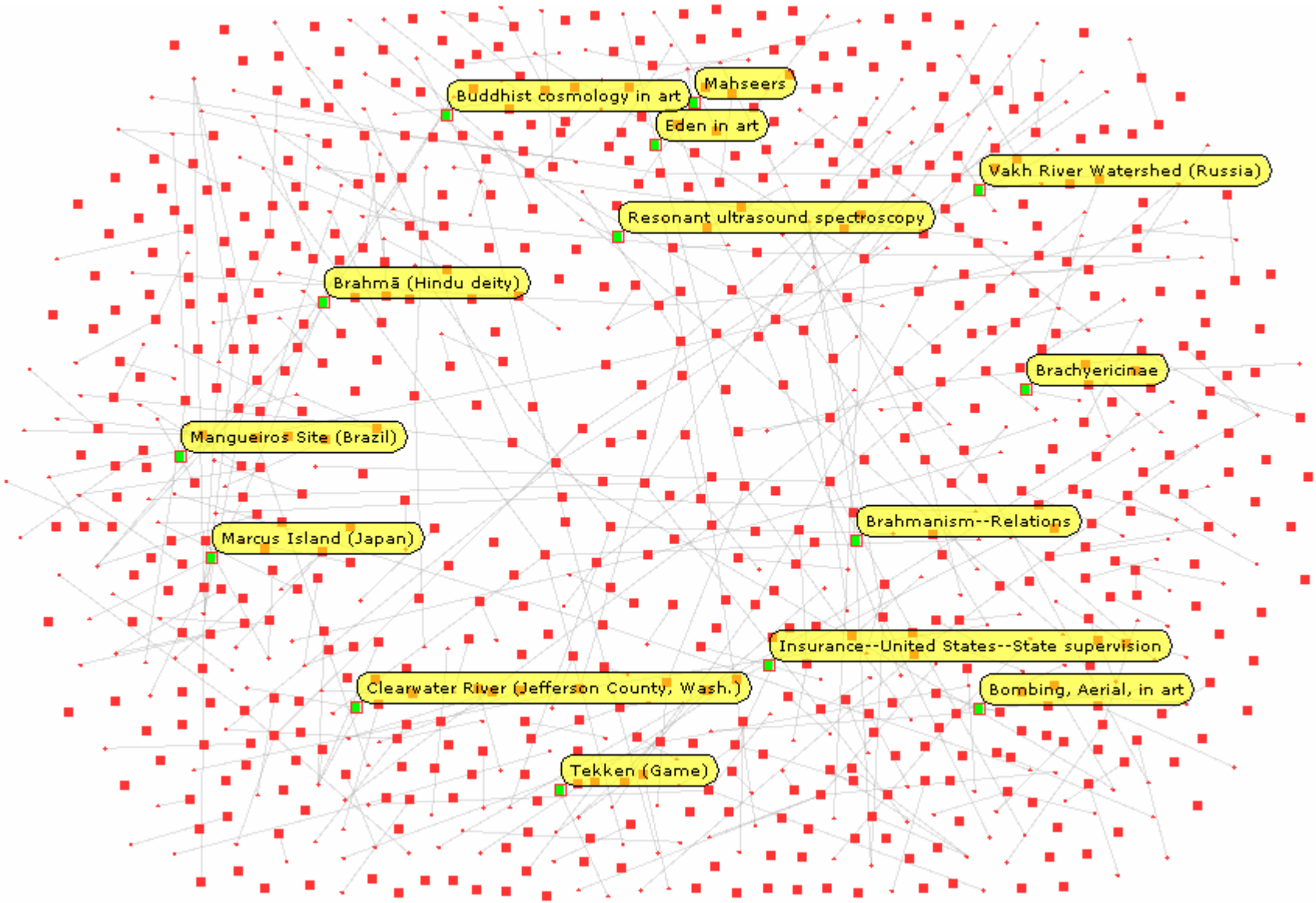
Internet (Computer network)
[TK5105.875.I57]
UF DARPA Internet (Computer network)
BT Wide area networks (Computer networks)

Internet advertising *(May Subd Geog)*
[HF6146.I58]
BT Advertising

Internetworking (Telecommunication)
(May Subd Geog)
UF Inter-networking (Telecommunication)
Interoperability in computer networks






Buddhist cosmology in art
Mahseers
Eden in art
Vakh River Watershed (Russia)
Resonant ultrasound spectroscopy
Brahmā (Hindu deity)
Brachyericinae
Mangueiros Site (Brazil)
Marcus Island (Japan)
Brahmanism--Relations
Insurance--United States--State supervision
Clearwater River (Jefferson County, Wash.)
Bombing, Aerial, in art
Tekken (Game)


Infants in art
Asians in art
Death in art
Days in art
Falconry in art
Beguinages in art
Buddhist cosmology in art
Bullfights in art
Dinosaurs in art
Architecture, Roman, in art
Dwellings in art
Art patrons in art
Islamic calligraphy in art
Buddhism in art
Barbershops in art
Horoscopes in art
Ball games in art
Bacchantes in art
Feminism in art
Artificial life in art
Asceticism in art
British in art
Imaginary places in art
Ark of the Covenant in art
Architecture, Hindu, in art
Braid in art
Baseball player
Astrology, Chinese, in art
Architecture, Victorian, in art
Chimpanzees in art
Dahlias in art
Baader-Meinhof gang in art
Dancers in art
Gold mines and mining in art
Brooms and brushes in art
Boars in art
Excavations (Archaeology) in art
Architecture, Classical, in art
Baseball in art
Holmes, Sherlock (Fictitious character) in art
Clearwater River (Jefferson County, Wash.)
Horns in art
Audiences in art
Boundaries in art
Hotels in art
Daggers in art
Hounds in art
Tekken (Game)
Bottlenose dolphin in art
Ballet dancing in art
Jewelry in art
Bars (Drinking establishments) in art
Crowds in art
Cycling in art
Axes in art
Bill (Anatomy) in art
Interior architecture
Brahmanism--Relations
Crowns in art
Boathouses in art
Holy Year in art
Germans in art
Balconies in art
Diesel locomotives in art
Astronomy in art
Insurance--United States--State sup
Crime in art
Coal mines
Crusoe, Robinson (Fictitious
Columbines in art
Desert animals in art
Lotus
Buttocks in art
Bronze
Canidae in art
Cranes (Birds) in art
Vakh River Watershed (Russia)
Field crops in art
Armed Forces in art
Goddesses in art
Basques in art
Ark of the Covenant in art
Brachycericinae
Coffeehouses in art
Desserts in art
Chimera (Greek mythology) in art
Geographical myths in art
Buddhas in art
Falcons in art
Crosses in art
Desire in art
Childbirth in art
Beds in art
Hell in art
Aboriginal
Bombin
Brahmanism in art
Authority in art
Architecture, Greek, in art
Folklore in art
Difference (Philosophy) in art
Homemakers in art
Architecture, Romanesque, in a
Cowgirls in art
Architecture, Renaissance, in
Beards in art
Architecture, Islam
Islands in art
Church vestments i
Batavi (Germanic people) in ar

- Publish KOS as linked data in the Web
  - Make their concepts and their interconnections part of the Web of data
- Why?
- How? (SKOS...)

The Web and SKOS

WHY KOS IN THE WEB?

The Value of Links

- The Web showed, links between documents are really useful
- Google's pagerank showed, link structure of network means something (and is worth something!)
- Social networking Web sites showed, how much we value other kinds of links
- **KOS link ideas ... If those links are in the Web, can be exploited by anyone, anywhere ...**

Quality (and Openness)

- The Web values **quality** and **openness**
  - E.g. Wikipedia, very high search rank
- Most KOS are high-quality resources
  - Both concepts and links
- In emerging Web of data, KOS are **natural hubs** ... **attractors** ... **high gravity** ... attract links
- KOS act as firm foundations for growth of Web of data ...

Value Proposition

- Links are paths to discovery (of documents, data, ...)
- Links can be exploited in useful and surprising ways (serendipity)
- Well-established KOS like LCSH can be **hubs** in the Web of linked data

The Web Gives You...

- A globally unambiguous point of reference (URI) for each concept
- A linking architecture that...
  - ...makes linking virtually free...
  - ...requires no central coordination...
  - ...has unlimited scalability...

Why KOS in the Web?

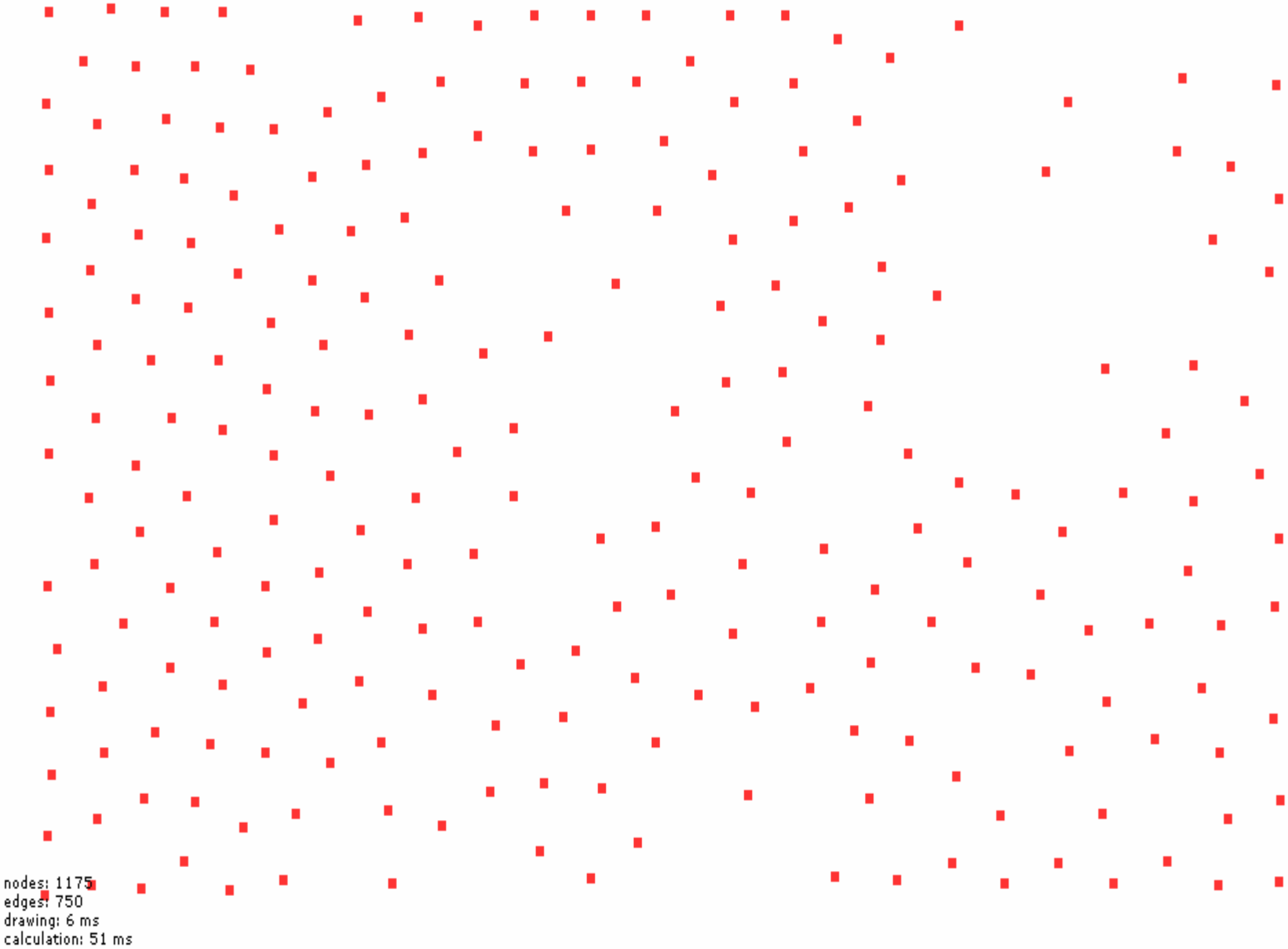
- The Web and KOS are deeply complementary

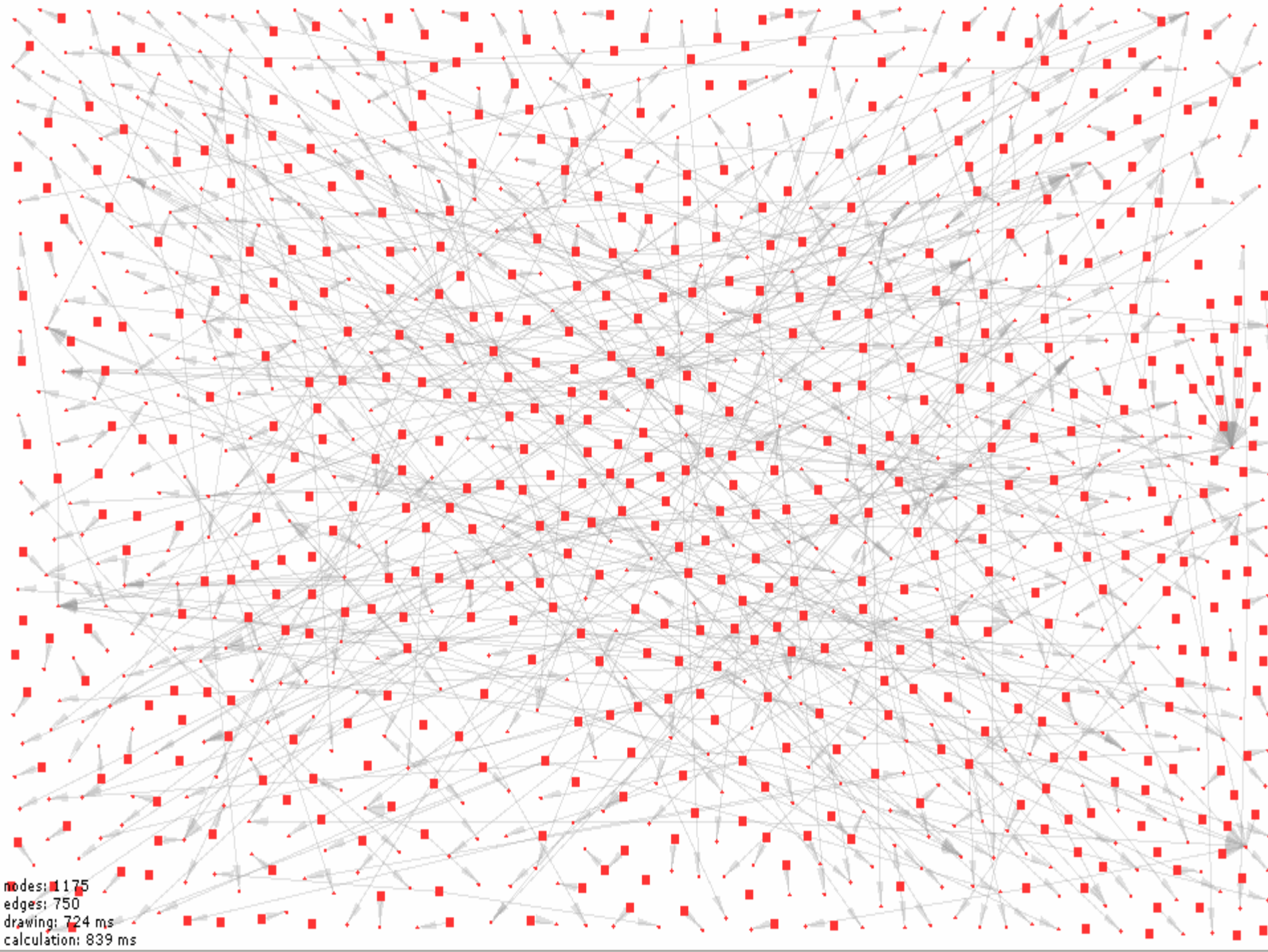
Demo

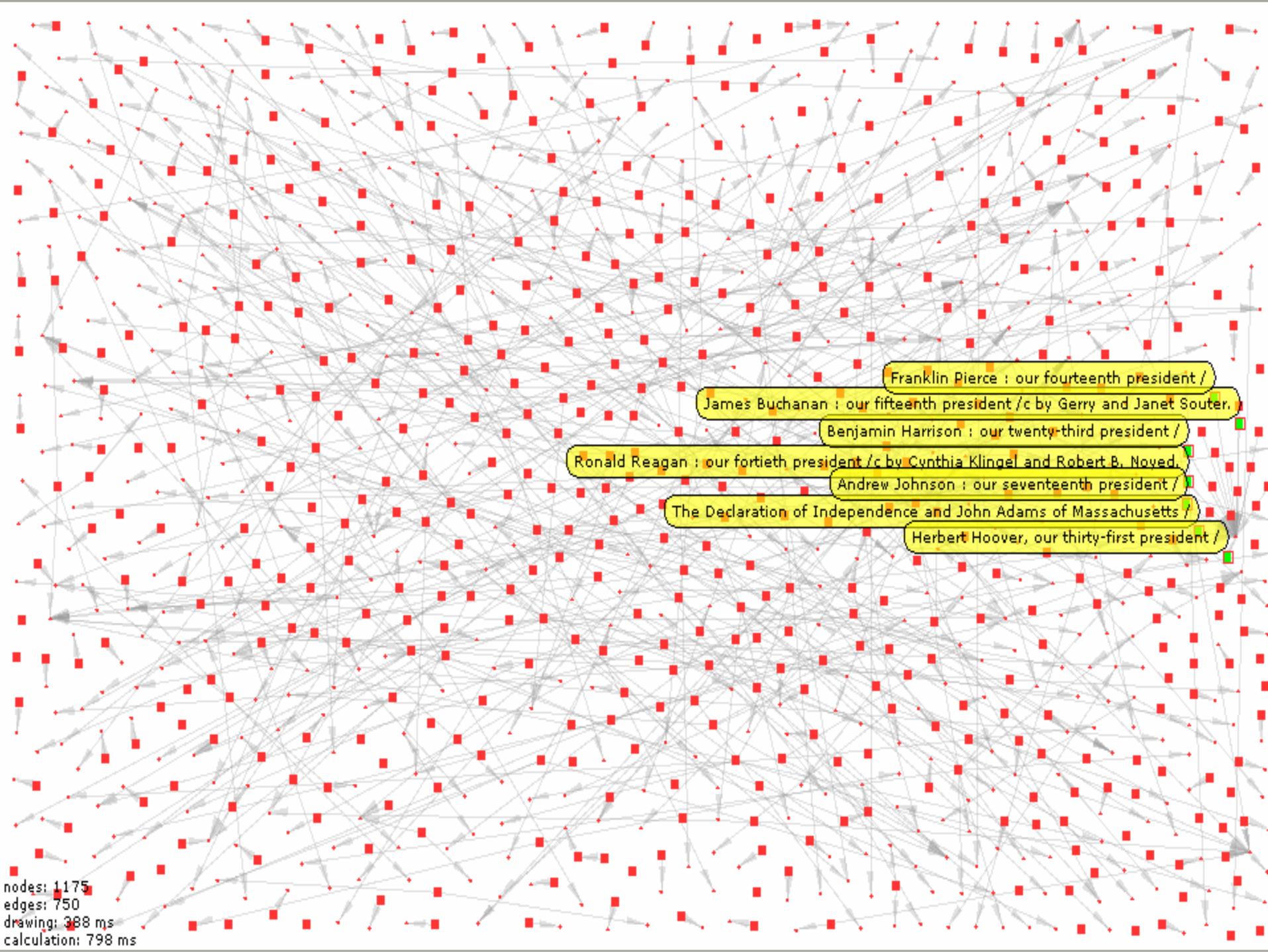
- Linked library metadata...
- http://inkdroid.org/bzr/lcsh/docs/slides/ -- Ed Summers

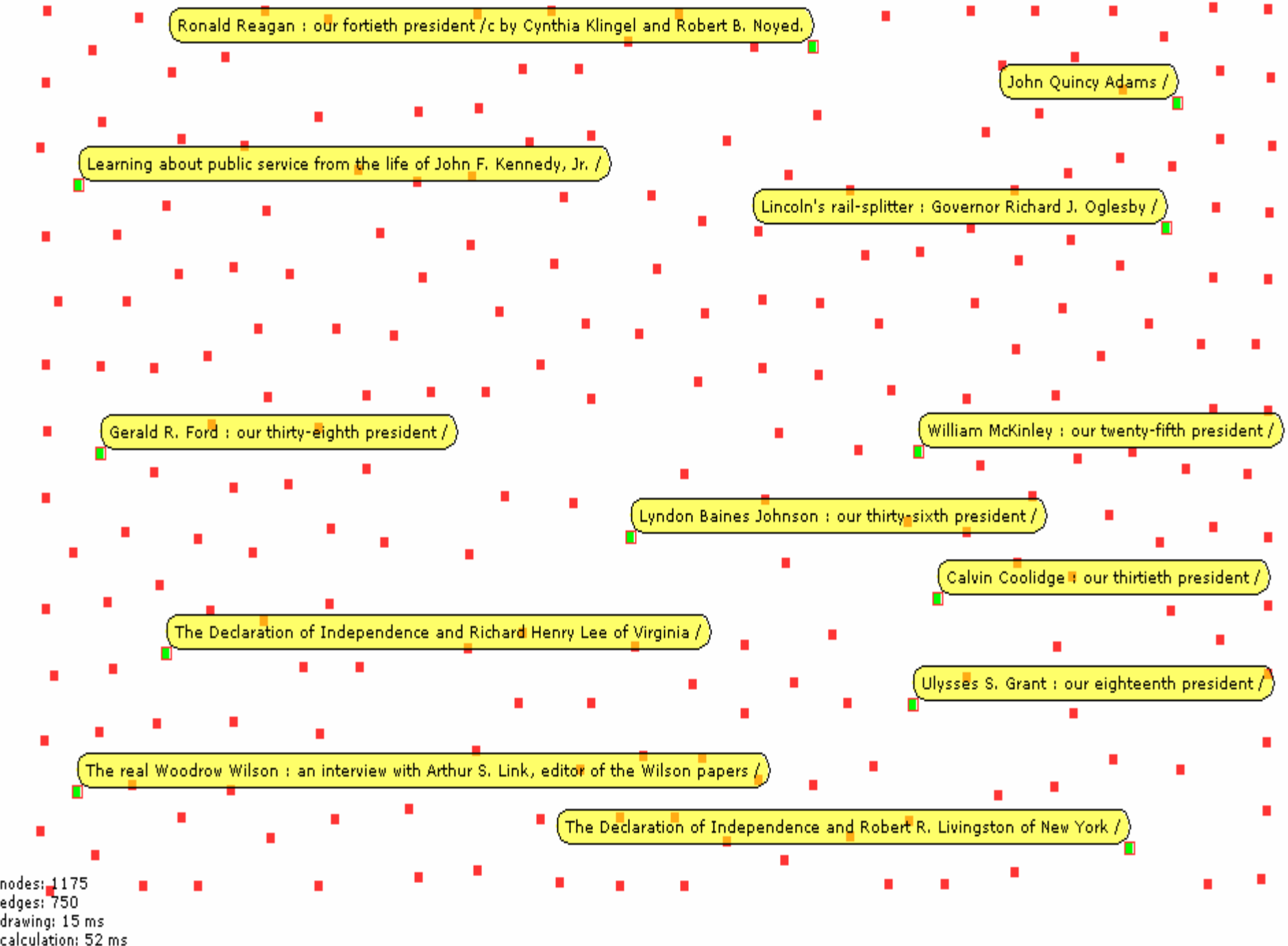
```
@prefix dc: <http://purl.org/dc/terms/> .

<http://lccn.loc.gov/99027665>
  dc:title "Weaving the Web : the original design
    and ultimate destiny of the World Wide Web by
    its inventor /" ;
  dc:creator "Berners-Lee, Tim." ;
  dc:creator "Fischetti, Mark." ;
  dc:type "text" ;
  dc:publisher "San Francisco : HarperSanFrancisco" ;
  dc:date "1999" ;
  dc:language "eng" ;
  dc:description "Includes index." ;
  dc:subject "Berners-Lee, Tim." ;
  dc:subject "World Wide Web" ;
  dc:identifier "URN:ISBN:0062515861 (cloth)" ;
  dc:identifier "URN:ISBN:006251587X (paper)" .
```

```
@prefix dc: <http://purl.org/dc/terms/> .

<http://lccn.loc.gov/99027665>
  dc:title "Weaving the Web : the original design
    and ultimate destiny of the World Wide Web by
    its inventor /" ;
  dc:creator "Berners-Lee, Tim." ;
  dc:creator "Fischetti, Mark." ;
  dc:type "text" ;
  dc:publisher "San Francisco : HarperSanFrancisco" ;
  dc:date "1999" ;
  dc:language "eng" ;
  dc:description "Includes index." ;
  dc:subject "Berners-Lee, Tim." ;
  dc:subject <http://loc.gov/lcsh/sh95000542> ;
  dc:identifier <urn:isbn:0062515861> ;
  dc:identifier <urn:isbn:006251587X> .
```


nodes: 1175
edges: 750
drawing: 6 ms
calculation: 51 ms


nodes: 1175
edges: 750
drawing: 724 ms
calculation: 839 ms


Franklin Pierce : our fourteenth president /
James Buchanan : our fifteenth president /c by Gerry and Janet Souter.
Benjamin Harrison : our twenty-third president /
Ronald Reagan : our fortieth president /c by Cynthia Klingel and Robert B. Noyed.
Andrew Johnson : our seventeenth president /
The Declaration of Independence and John Adams of Massachusetts /
Herbert Hoover, our thirty-first president /
nodes: 1175
edges: 750
drawing: 388 ms
calculation: 798 ms

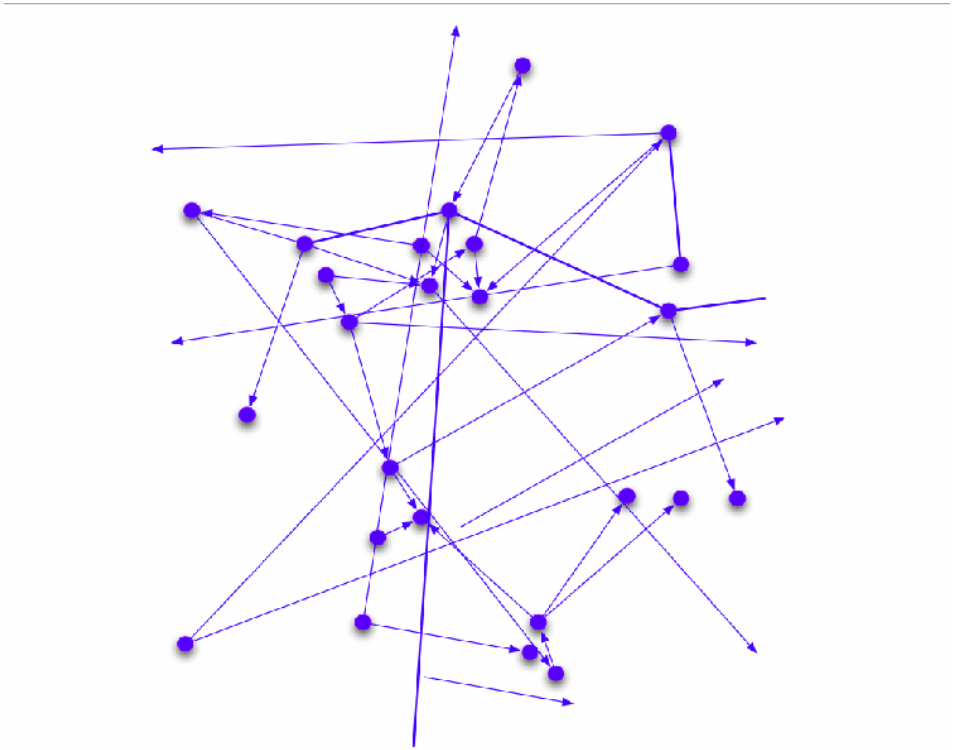

Ronald Reagan : our fortieth president /c by Cynthia Klingel and Robert B. Noyed.
John Quincy Adams /
Learning about public service from the life of John F. Kennedy, Jr. /
Lincoln's rail-splitter : Governor Richard J. Oglesby /
Gerald R. Ford : our thirty-eighth president /
William McKinley : our twenty-fifth president /
Lyndon Baines Johnson : our thirty-sixth president /
Calvin Coolidge : our thirtieth president /
The Declaration of Independence and Richard Henry Lee of Virginia /
Ulysses S. Grant : our eighteenth president /
The real Woodrow Wilson : an interview with Arthur S. Link, editor of the Wilson papers /
The Declaration of Independence and Robert R. Livingston of New York /
nodes: 1175
edges: 750
drawing: 15 ms
calculation: 52 ms

How SKOS?

HOW TO PUT KOS IN THE WEB

Publishing KOS in the Web?

- RDF
  - Basic data formalism for describing “graphs” of data



Publishing KOS in the Web?

- SKOS
- Standard set of...
  - Resource types (Classes)
  - Link types (Properties)
- ... For representing **KOS** as RDF data

SKOS Resource Types (Classes)

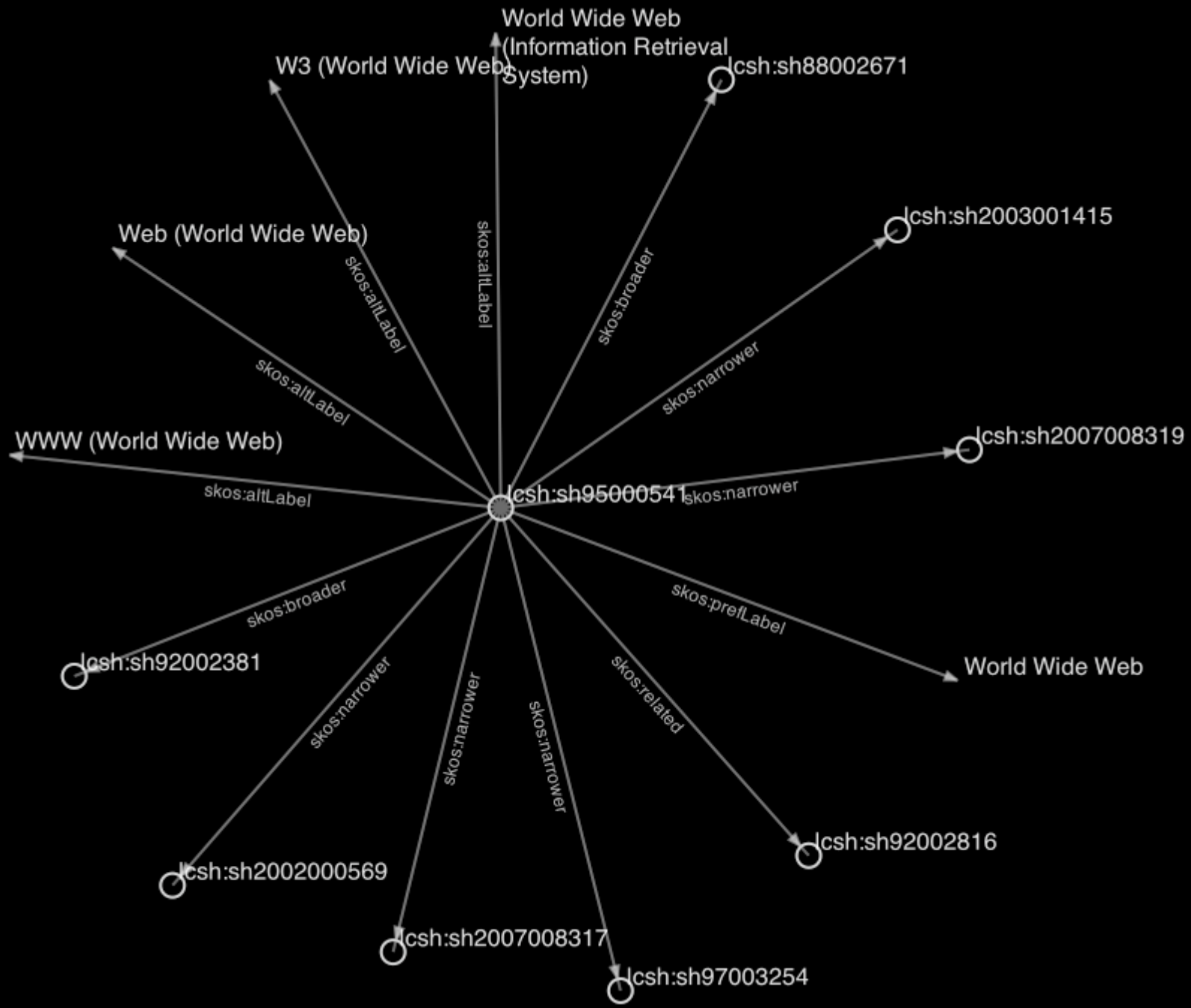
- skos:Concept
  - E.g. LCSH concept of US Presidents
- skos:ConceptScheme
  - E.g. LCSH itself

SKOS Link Types (Properties)

- For labeling concepts
  - skos:prefLabel, skos:altLabel, skos:hiddenLabel
- For documenting concepts
  - skos:note, skos:scopeNote, skos:definition, skos:editorialNote...
- For linking concepts
  - skos:broader, skos:narrower, skos:related

“SKOSIFY”

- Create a SKOS representation of a KOS e.g. LCSH


World Wide Web (Information Retrieval System)
W3 (World Wide Web)
lcsh:sh88002671
Web (World Wide Web)
lcsh:sh2003001415
WWW (World Wide Web)
lcsh:sh2007008319
lcsh:sh95000541
lcsh:sh92002381
World Wide Web
lcsh:sh92002816
lcsh:sh2002000569
lcsh:sh2007008317
lcsh:sh97003254
skos:altLabel
skos:altLabel
skos:altLabel
skos:altLabel
skos:broader
skos:narrower
skos:narrower
skos:broader
skos:prefLabel
skos:related
skos:narrower
skos:narrower
skos:narrower

Publishing SKOS Data in the Web

- Serve data file from a Web/file server
- As **linked data** in the Web
  - E.g. http://lcsh.info/sh85106459#concept
  - Content-negotiable
  - HTML or RDF/XML or N3 or JSON
- Via a **SPARQL endpoint** in the Web
  - E.g. http://sparql.lcsh.info/
- Via other **Web API**

DESIGN ISSUES

Scope of SKOS?

- SKOS will be an all-encompassing standard for the lossless representation and exchange of all varieties of knowledge organisation system...?
- **http://lists.w3.org/Archives/Public/public-swd-wg/2008Feb/0116.html -- Antoine Isaac**

- *"Yet, it happens, **looking at the way these things are used** now and will be in the near future (with more and more links established between them), that (i) some standardisation has to take place, and that (ii) this standardisation can be actually grounded on some observed practical similarities* (http://www.w3.org/TR/skos-ucr/)"
- *"Our aim is not to replace the original objects in their initial context of use, but to allow to **port them to a shared space**, based on a **simplified model**, enabling wider re-use and better interoperability*."

Lessons from the Web

- The Web is a **platform for innovation**
- **Web standards promote innovation and diversity**
  - (unlocking new markets & opportunities)
- In the Web, you can never predict all uses
- SKOS should capture a small amount of common ground ... Just enough to enable KOS's valuable concepts and connections to be deployed in the Web and be linked to/from
- SKOS is infinitely extensible
  - Easy to mix & match
  - Easy to refine

Representation of Lexical Entities

- SKOS Reference includes a new, optional, extension called “SKOS eXtension for Labels (XL)”
- Provide a foundation for finer-grained extensions... BS8723?
- N.B. Read carefully, some subtleties!

Living in an Open World

- SKOS is an “open world” data model
- ... Why?
- ... What does it mean?
- ... What is different?

Inference & Integrity

- SKOS data model support some simple inferences
- SKOS data model also supports some simple integrity checks
- ...but, N.B. Open world!

Linking & Mapping

- SKOS Reference defines properties for mapping between concepts in different schemes...
  - skos:exactMatch, skos:closeMatch, skos:broadMatch, skos:narrowMatch, skos:relatedMatch
- In an open world, hard to draw a sharp distinction between **linking** two schemes and **mapping** two schemes

Linking & Mapping

- Where two schemes are complementary but have little overlap in scope ... Link them to create a single “virtual” scheme
  - Use skos:broader, skos:narrower, skos:related
- Where two schemes have significant overlap in scope... Map them to allow translation of queries and/or metadata
  - Use skos:exactMatch, skos:broadMatch etc.

Linking & Mapping

- In an open world, the “boundaries” between concept schemes become optional
- Freedom to observe or ignore boundaries
- What is “linking” and “mapping” depends on point of view...
- Hence SKOS Reference currently takes the view that mapping is a special type of linking (rather than that they are absolutely distinct)

SKOS & OWL

- Beginning to explore design patterns
- We've been careful to leave the options open...
  - Why some aspects of SKOS data model may appear counterintuitive

Challenges

- SKOS is pulled in many different directions...
- Sits between “semi-formal” knowledge organisation, informal & socially-mediated information organisation, formal knowledge representation, text-mining, the Web, ...
- We’ve tried to support as many of “core” functionalities of “traditional” KOS applications ...
- ... without closing the door on exploring new interactions, hybrid applications ...
- **SKOS as a platform for innovation and diversity**

Thank You

- All members of public-esw-thes@w3.org

Final Word

- KOS are natural hubs, through which many things are (and can be) connected
- The Web is a cheap, scalable, ubiquitous and open platform for linking things together
- The Web and KOS are natural partners...
- *"...a unique capacity to* ***break down boundaries of distance, language, and domains of knowledge****."*