Web and Semantic Web

MO826/MC936 - Information Systems Topics

André Santanchè
Laboratory of Information Systems – LIS
Institute of Computing – UNICAMP
February 2015

Web Science

Web Science
INCT

WEB SCIENCE BRASIL

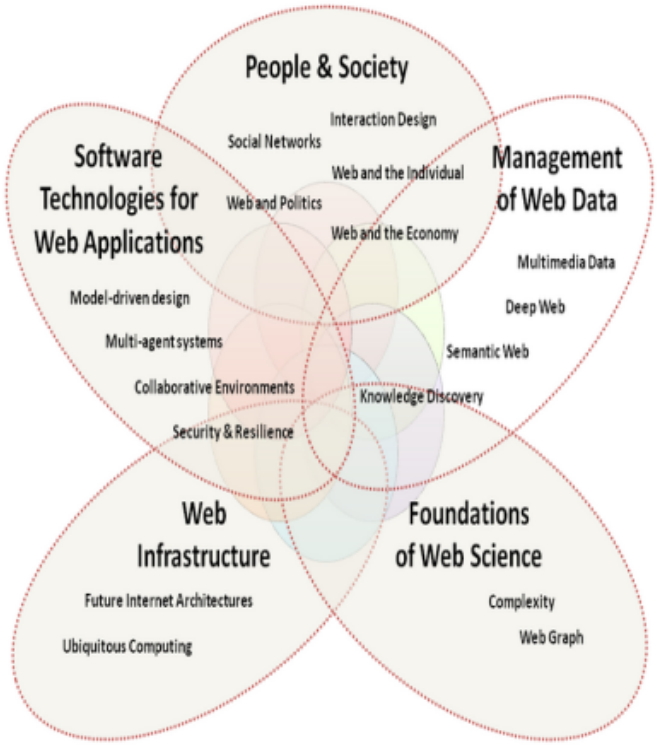
Brazilian Institute for Web Science Research

Brazilian Institute for Web Science Researc

Coordination: Prof. Carlos José Pereira de Lucena, Departamento de Informática,

The Institute's program addresses five dimensions of Web Science Research:

- People and Society
- Software Technologies for Web Applications
- Management of Web Data
- Web Infrastructure
- Foundations of Web Science



The Institute involves over 100 researchers. The following institutions are part of the Institute:

- PUC-Rio - Departamento de Informática
- UFRJ - Programa de Engenharia de Sistemas e Computação
- UNICAMP - Instituto de Computação
- Rede Nacional de Pesquisa
- UERJ
- UENF
- UFC
- UFRN
- UNIRio

Homework

- Individual
- Two slides (two and only two)
- First slide
  - Topics (bullets)
  - Challenges for the Web Science
- Second slide
  - Subject of our first presentation
- Upload the slides before the presentation at Moodle

Homework
References

- Hendler, J., Shadbolt, N., Hall, W., Berners-Lee, T., & Weitzner, D. (2008). **Web science: an interdisciplinary approach to understanding the web**. Communications of the ACM, 51(7), 60-69.

- Shneiderman, B. (2007). **Web science: a provocative invitation to computer science**. Communications of the ACM, 50(6), 25-27.

Web Science =

Web Engineering

+

Data Science

+

Web Social Science

+

Network Science

Web Engineering

Web Foundations

- Architecture
- Models, standards and languages
  - URL, URN, URI and IRI
  - HTML and XML
  - XPath and XLink
- Database perspective
  - Querying and XQuery

Web platform and applications

Foto “Family on Bike” por Mikael Colville-Andersen.

Web as Platform

Web as Platform


Semantics


Offline & Storage


Device Access


Connectivity

Multimedia


3D, Graphics & Effects


Performance & Integration


CSS3


Web platform and applications

- Web for mobiles
- Web services
- Web of Things

Web platform and applications

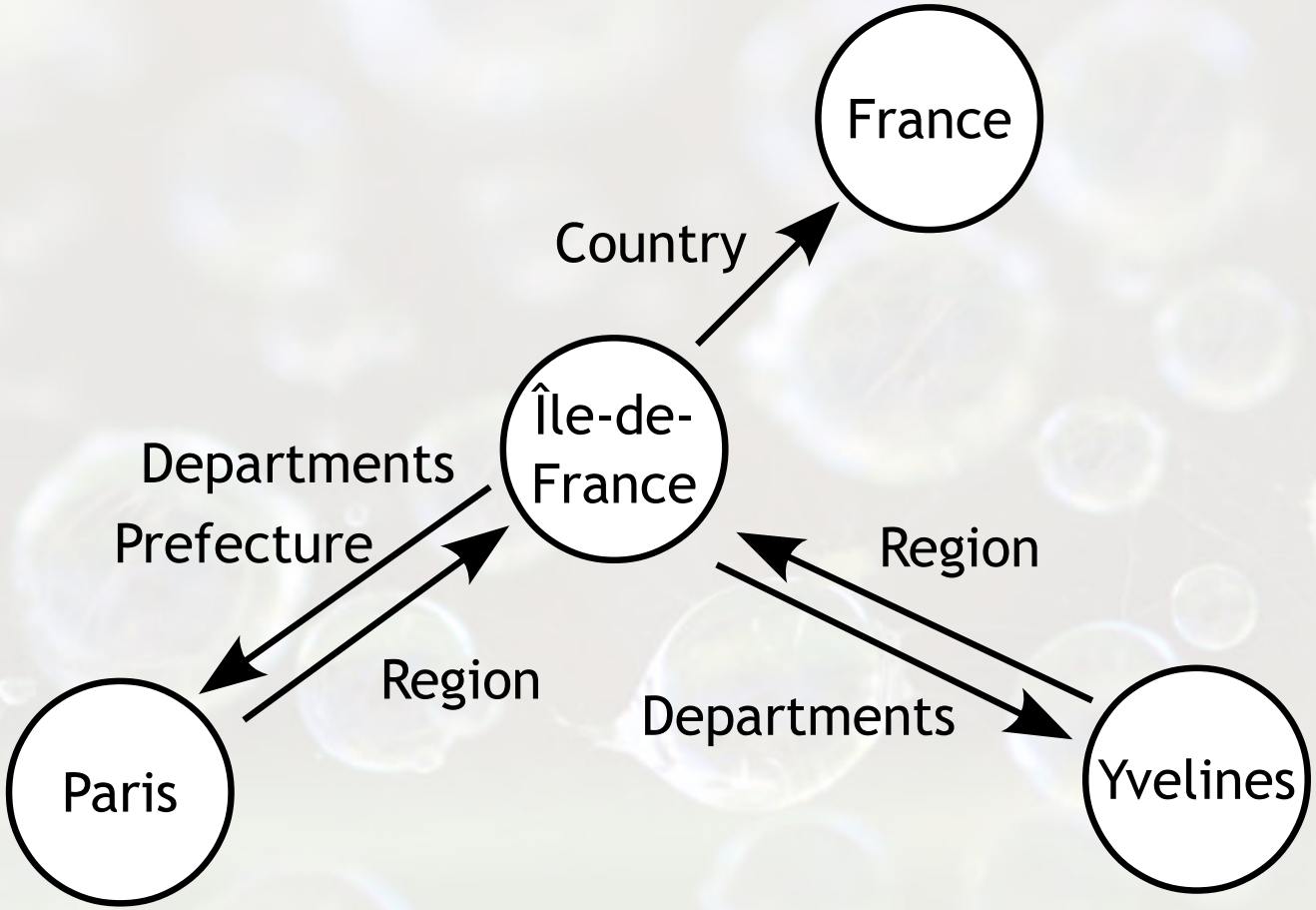
Web for Mobiles

Linked Data

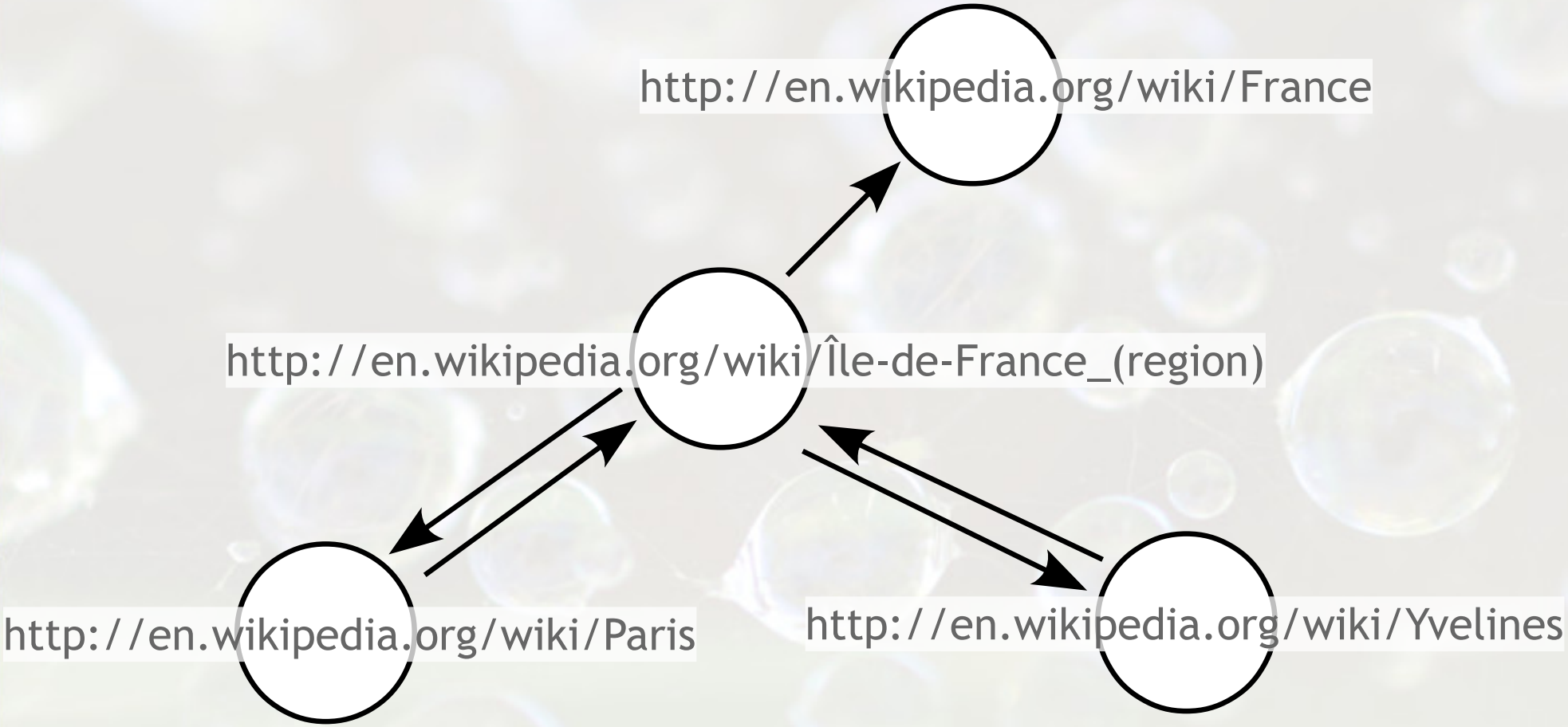
Wikipedia

Infobox

DBPedia


France
Country
Île-de-
France
Departments
Prefecture
Region
Region
Departments
Paris
Yvelines

DBPedia (URIs)



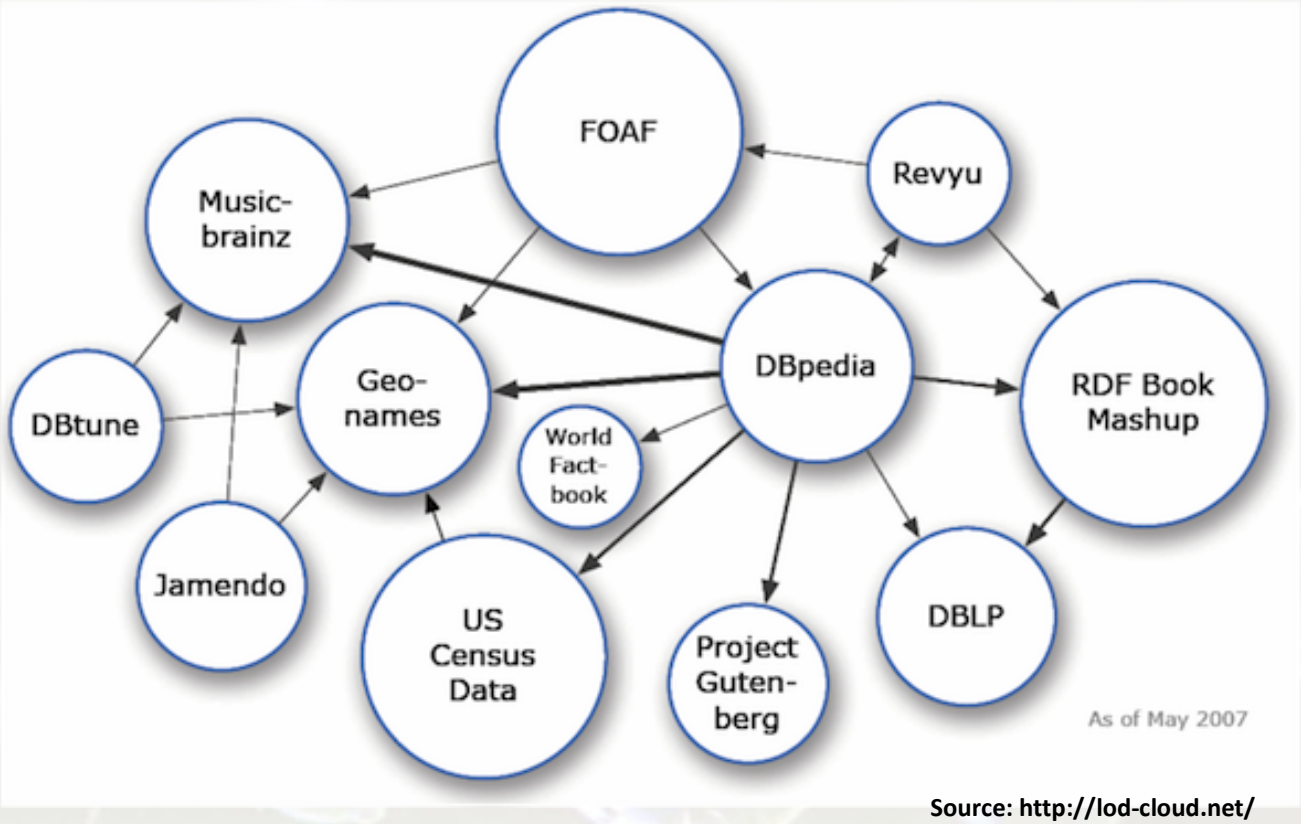
DBPedia – English

- **4 million things**
- 3.22 million classified in a consistent ontology
  - 832,000 persons
  - 639,000 places (427,000 populated)
  - 372,000 creative works
    - 116,000 music albums; 78,000 films; 18,500 video games
  - 209,000 organizations
  - 226,000 species
  - 5,600 diseases.

DBPedia – International

- 119 languages
- 24.9 million things
- 16.8 million interlinked with English
- 12.6 million unique things

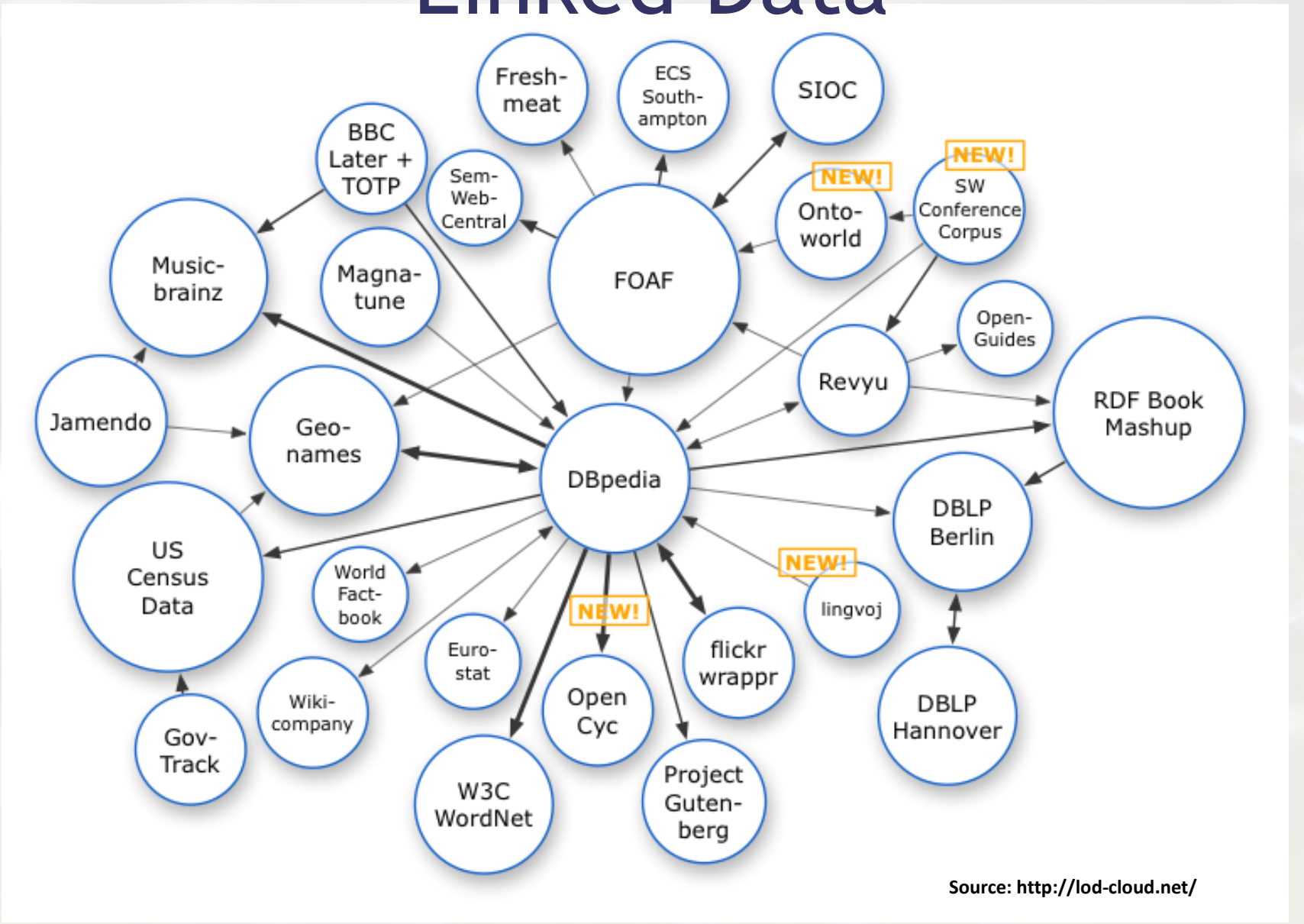
Linked Data



Source: http://lod-cloud.net/

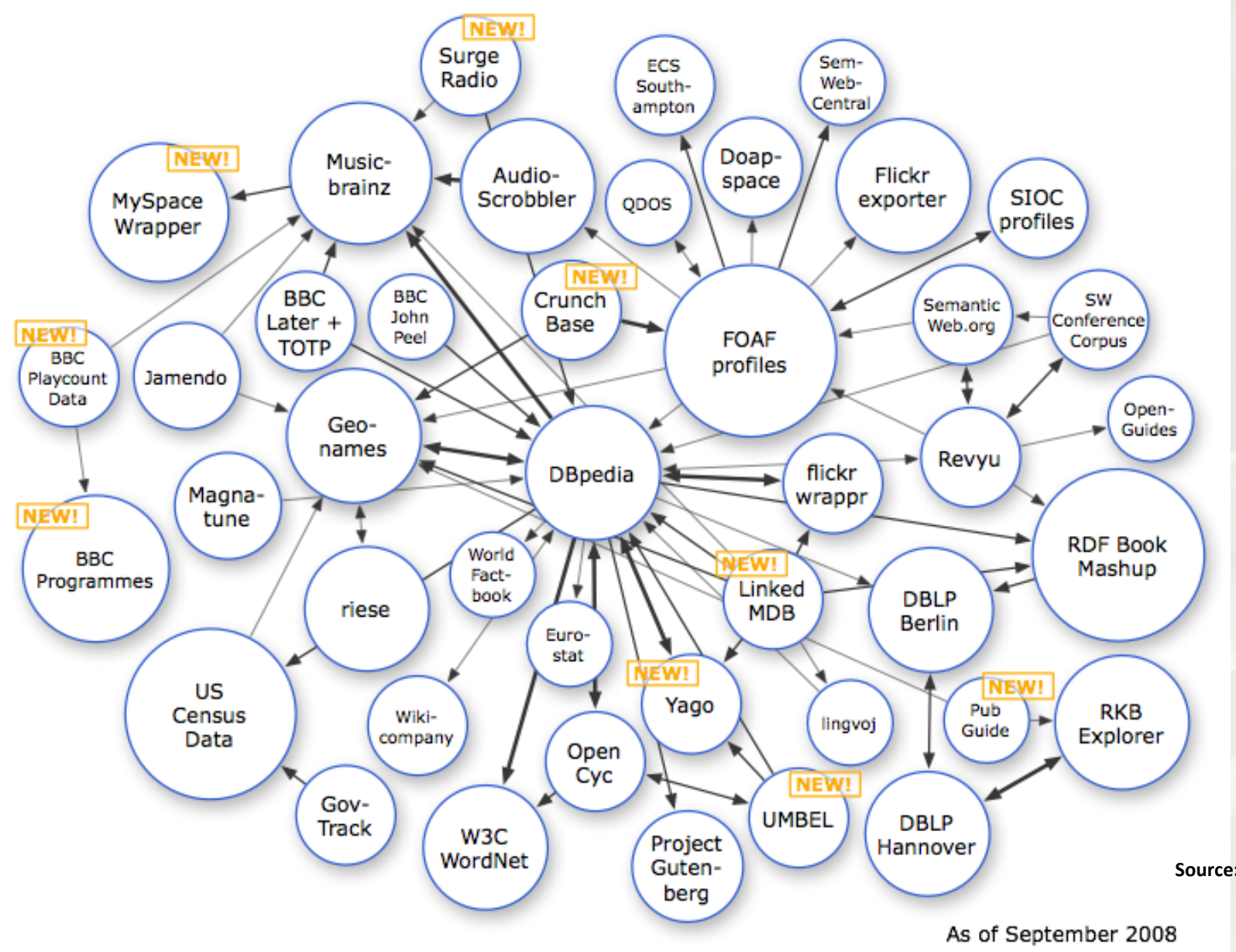
Datasets published following Linked Data 'format': **05/2007**

Linked Data



Datasets published following Linked Data 'format': **11/2007**

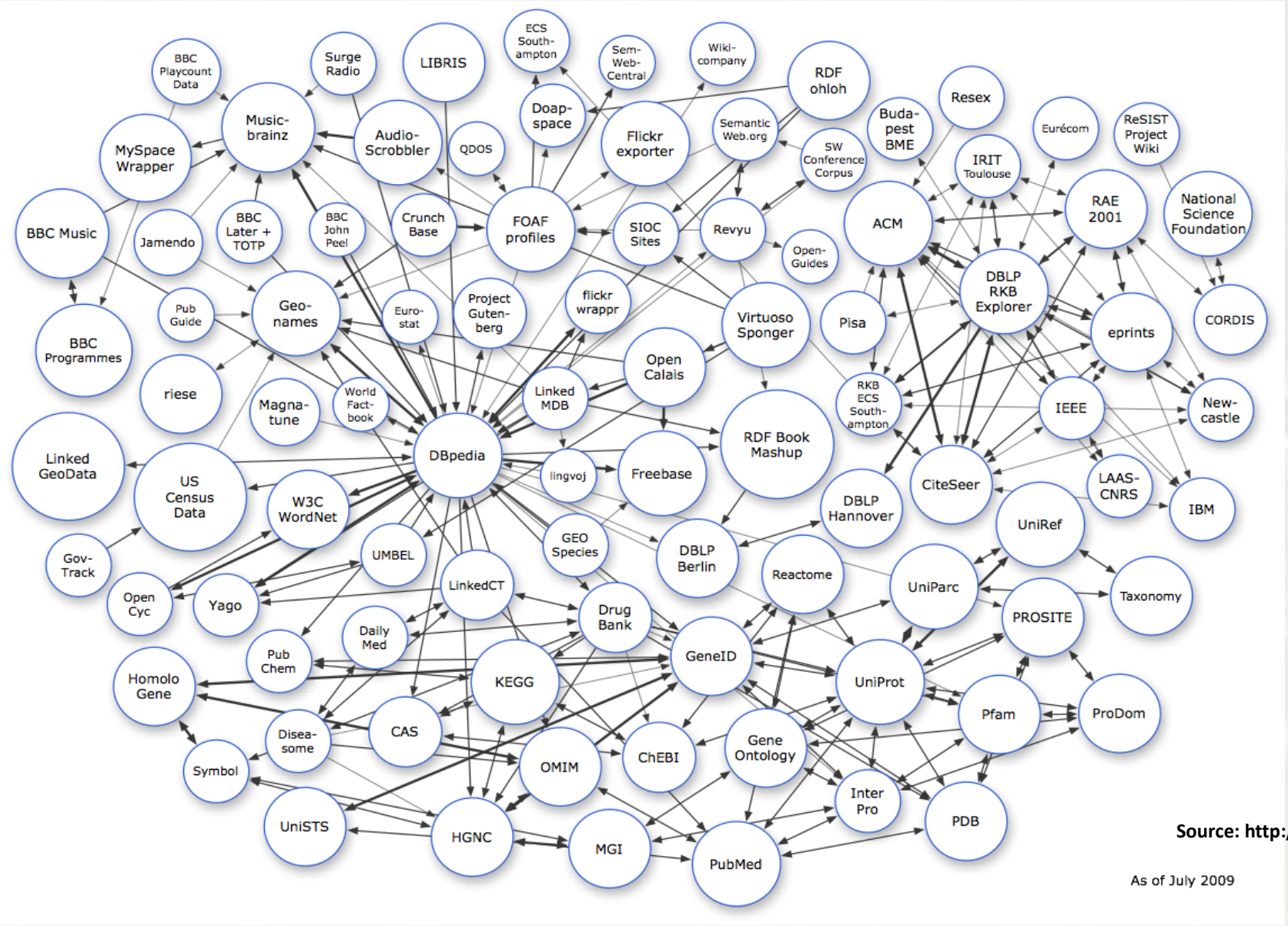
Linked Data



Source: http://lod-cloud.net/

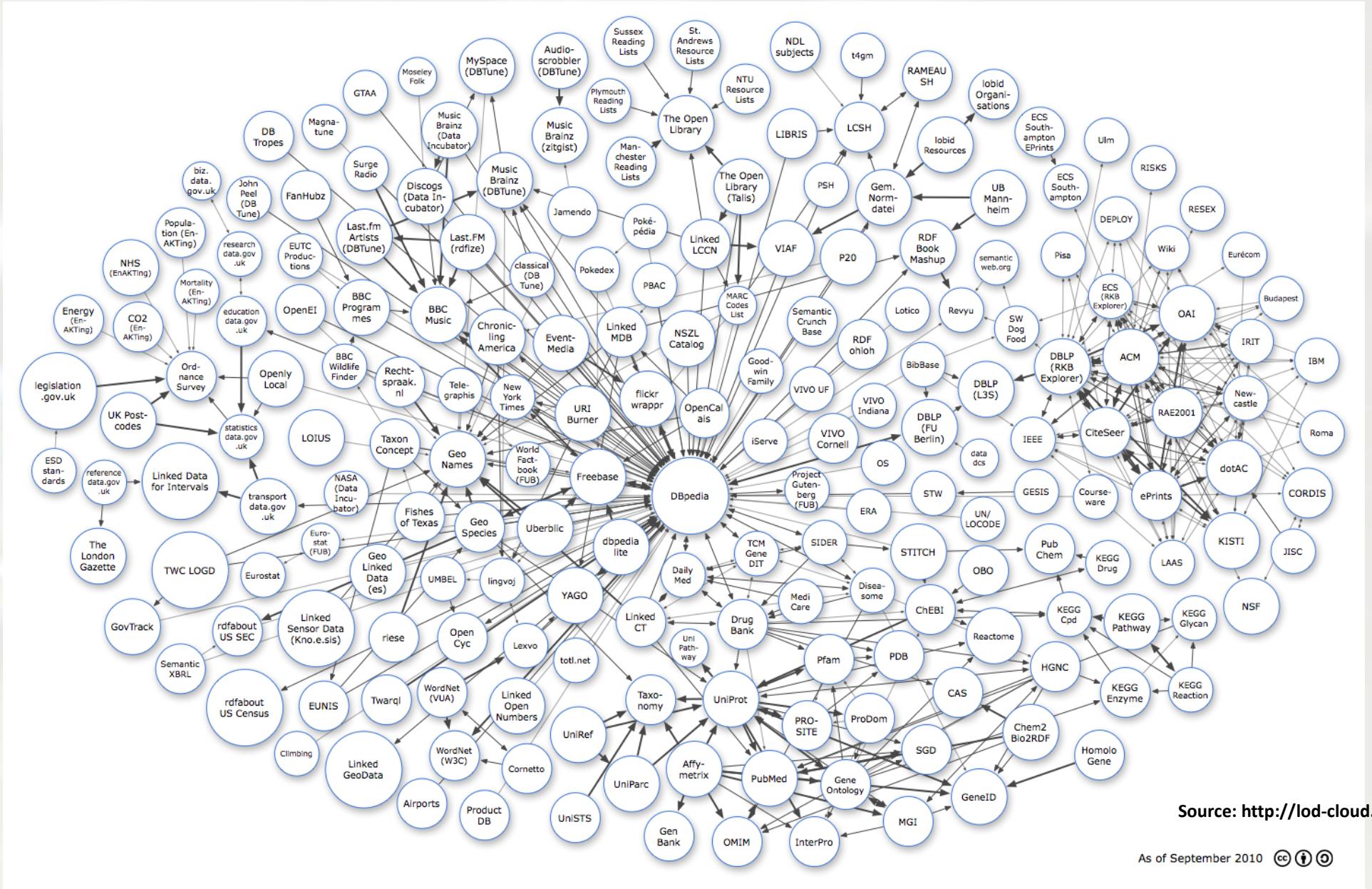
Datasets published following Linked Data 'format': **2008**

Linked Data



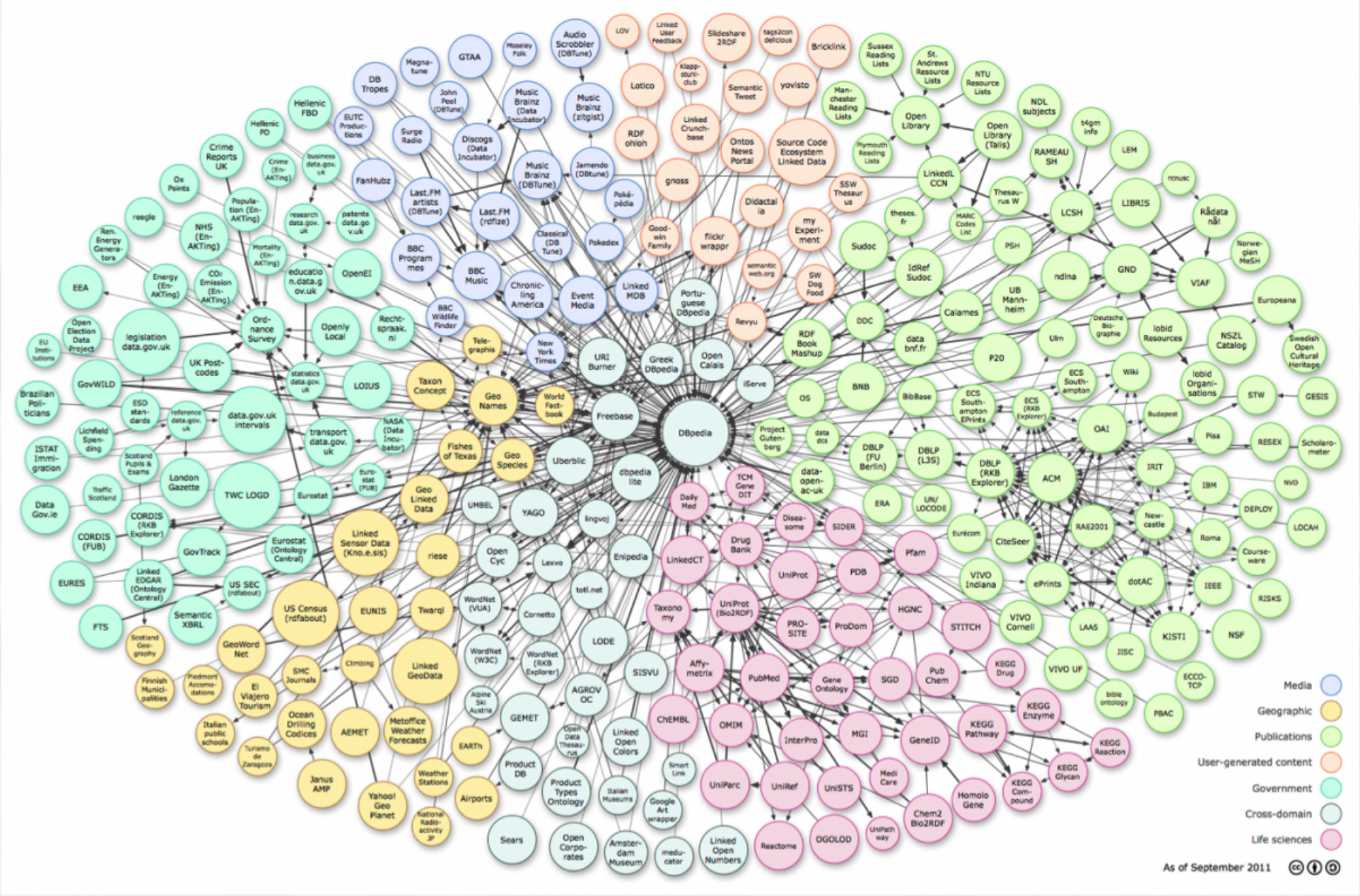
Datasets published following Linked Data ‘format’: **2009**

Linked Data



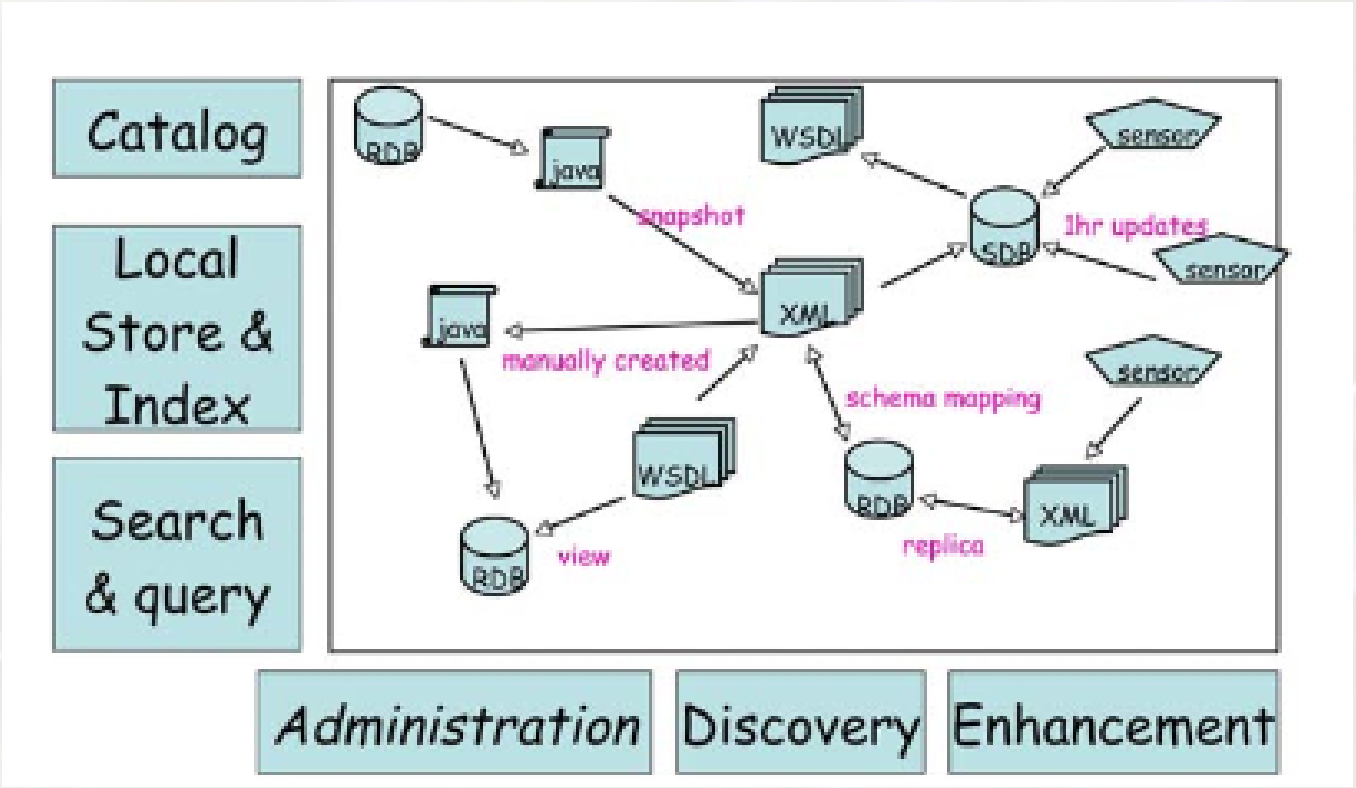
Datasets published following Linked Data 'format': **2010**

Linked Data



Datasets published following Linked Data ‘format’: **2011**

DataSpaces
Pay-as-you-go Integration

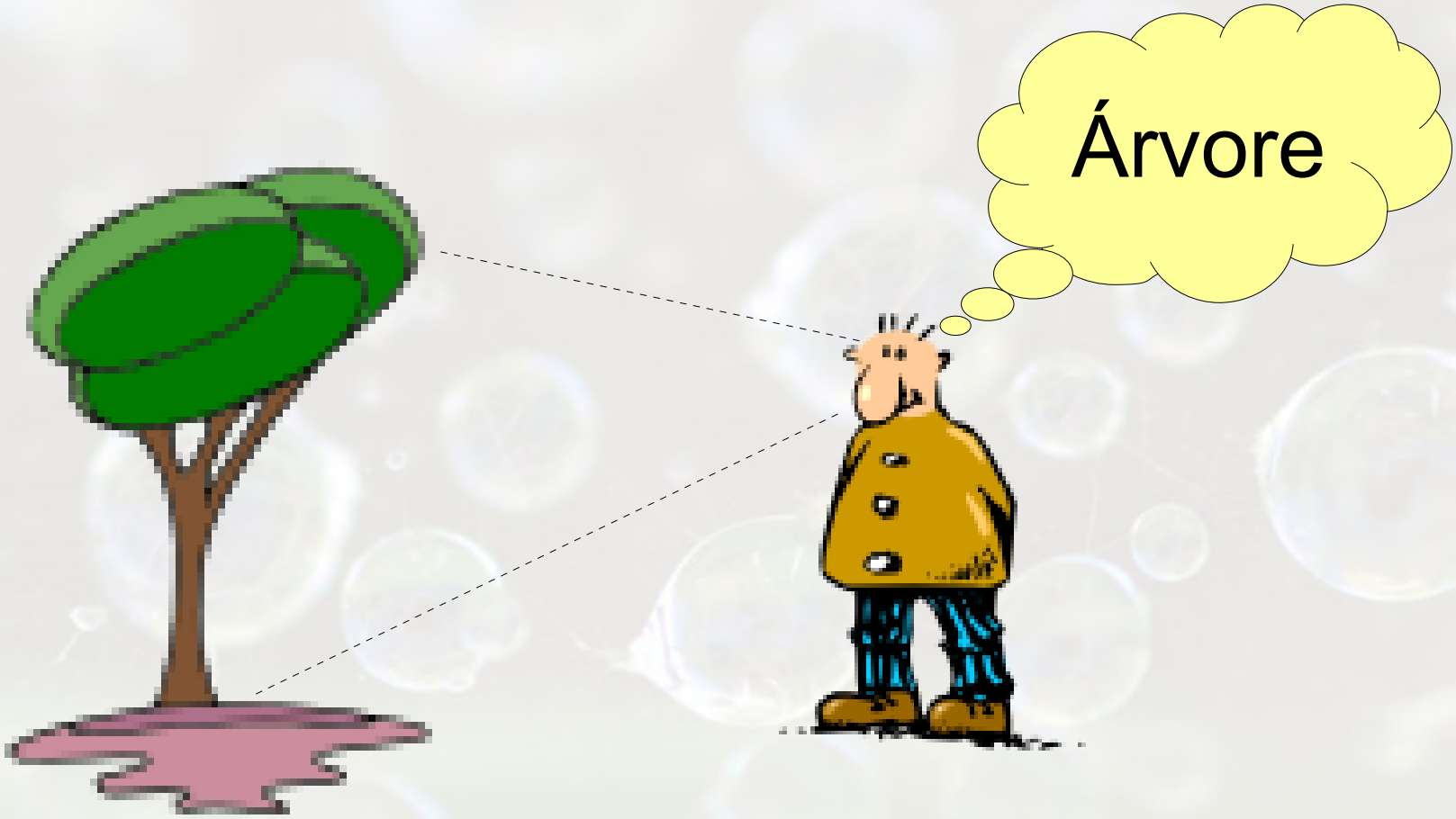


Franklin, M., Halevy, A., & Maier, D. (2005). From databases to dataspaces: a new abstraction for information management. SIGMOD Rec., 34(4), 27–33.

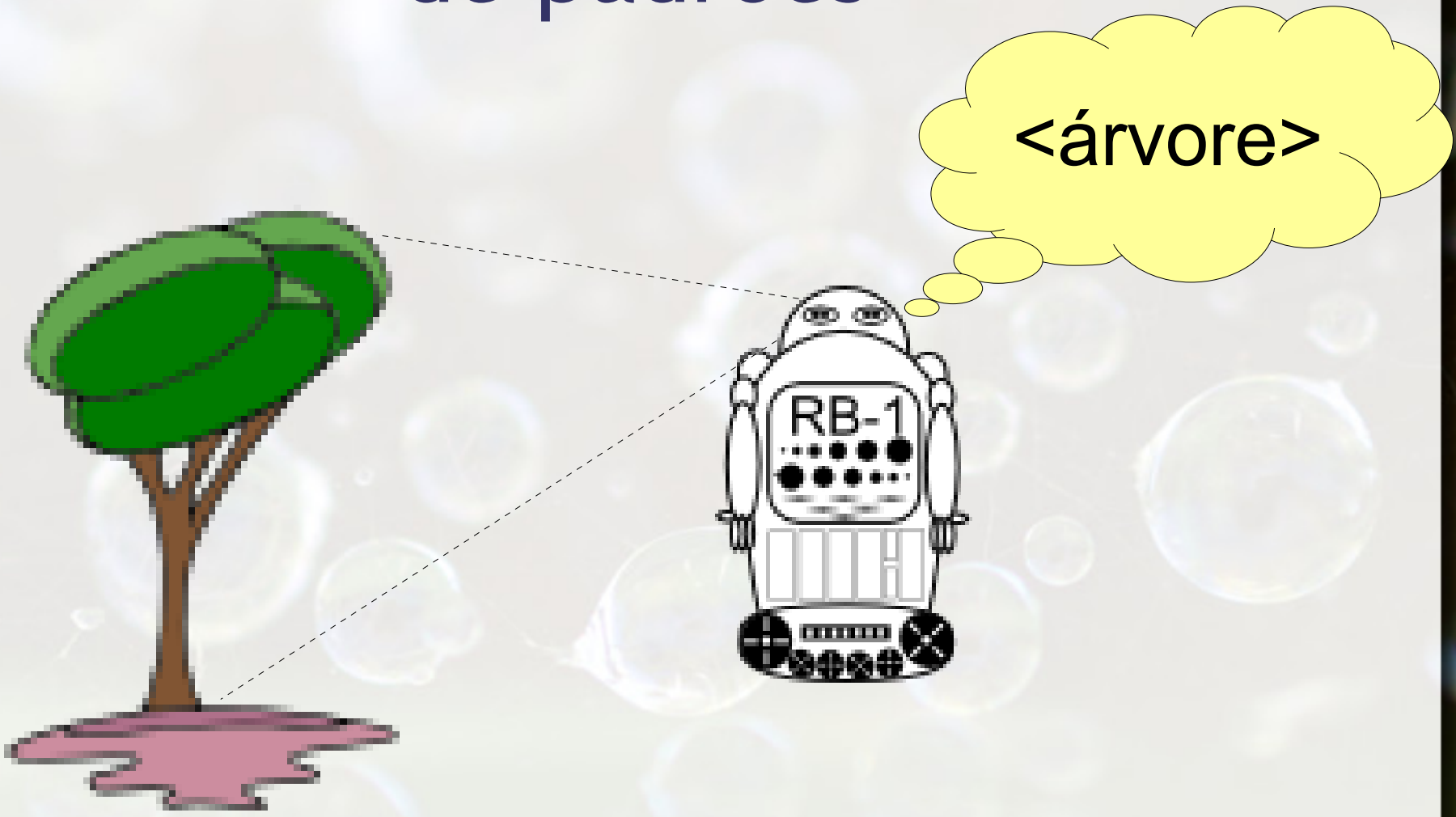
Semantic Web

- Architecture
- Models, standards and languages
  - RDF and OWL
- Web of Data and Metadata
- Querying and SPARQL
- Rules, reasoning and SWRL

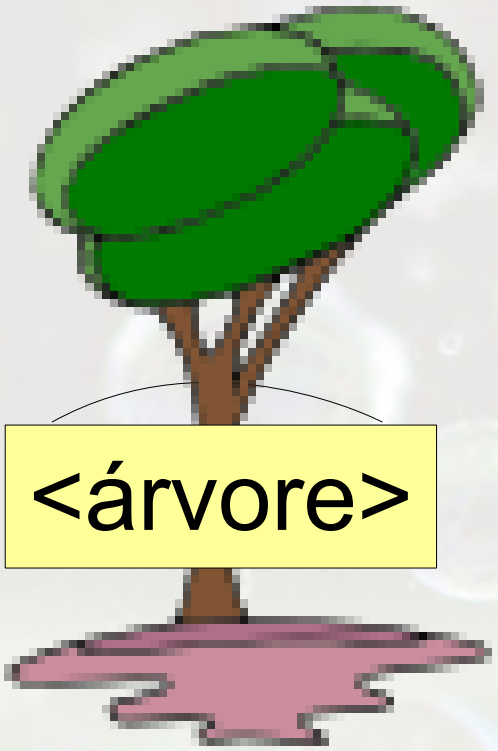
A Web na ótica humana



A Web na ótica de reconhecimento de padrões



A Web Semântica



A Web Semântica

http://www.edissauros.com.br

http://www.x.org/razao_social

Editora Edissauros

http://www.x.org/contratado

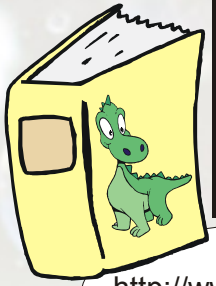
mailto:horacio@paleo.org

http://purl.org/dc/elements/1.1/publisher

http://www.x.org/nom

Horácio Montéquio

http://purl.org/dc/elements/1.1/creator



2ª

http://www.x.org/edicao

http://www.paleo.org/dinos.pdf

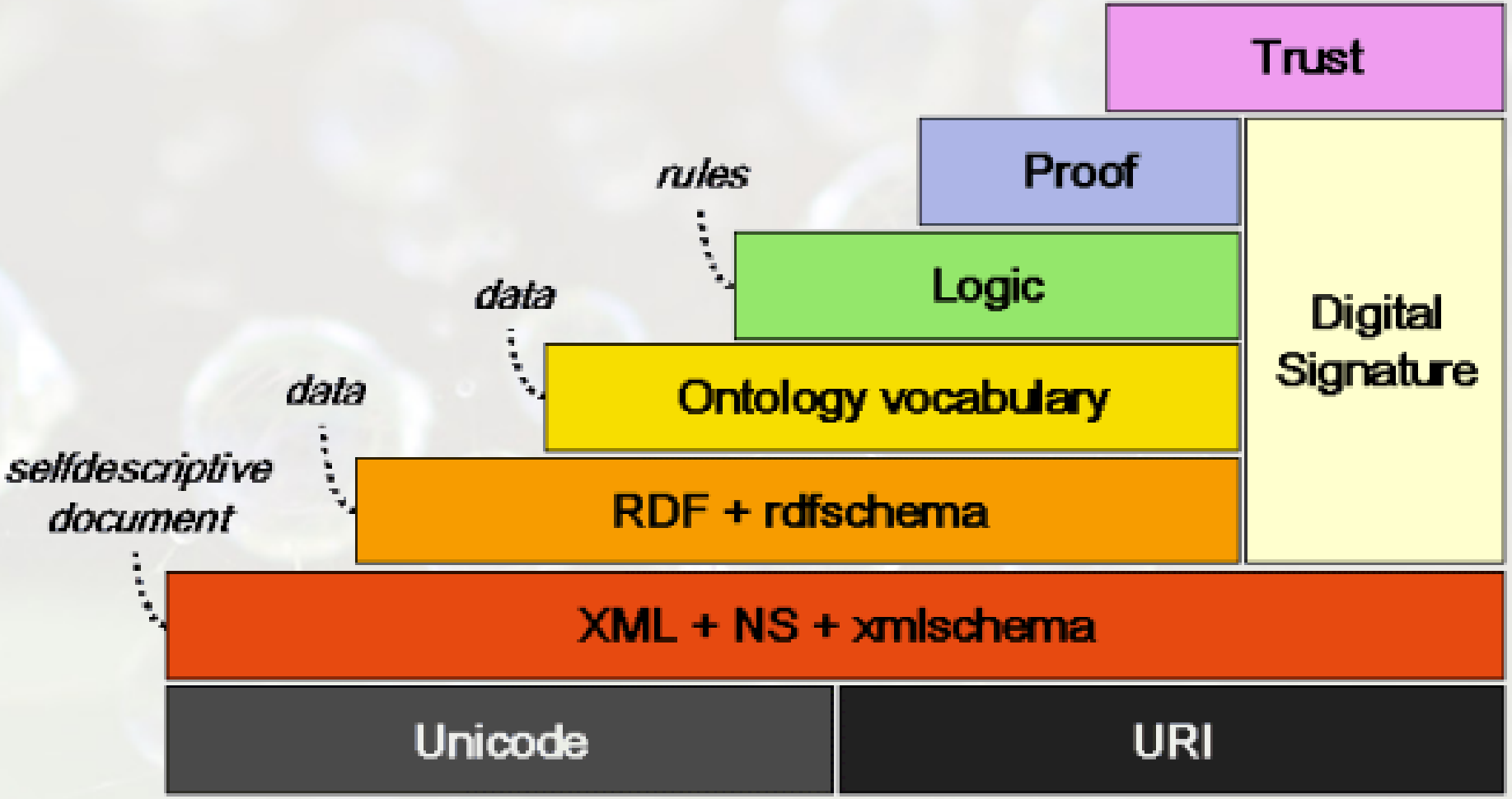
http://www.x.org/data_publicacao

17/05/2001

http://purl.org/dc/elements/1.1/title

Vida dos Dinossauros

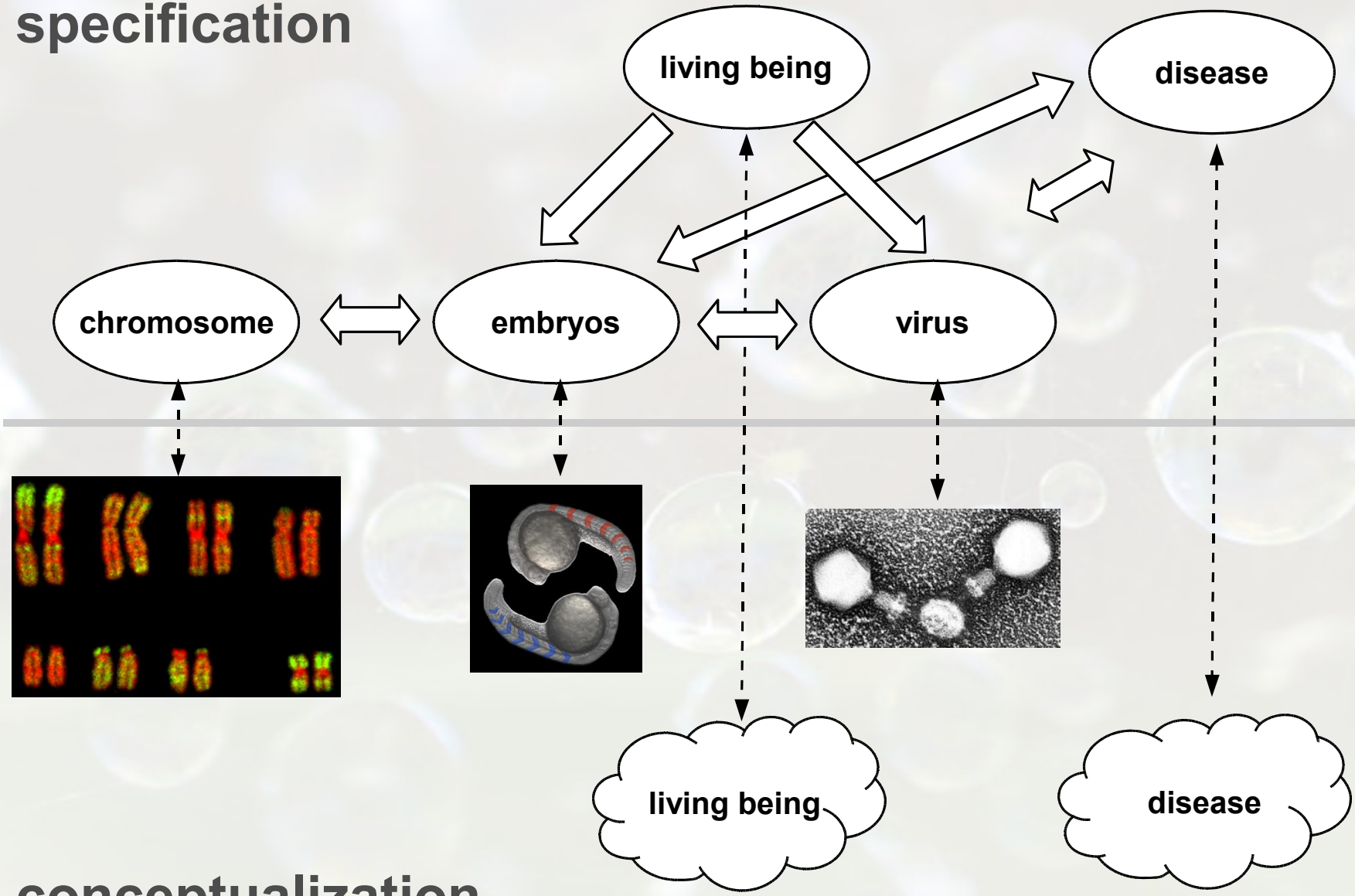
Semantic Web



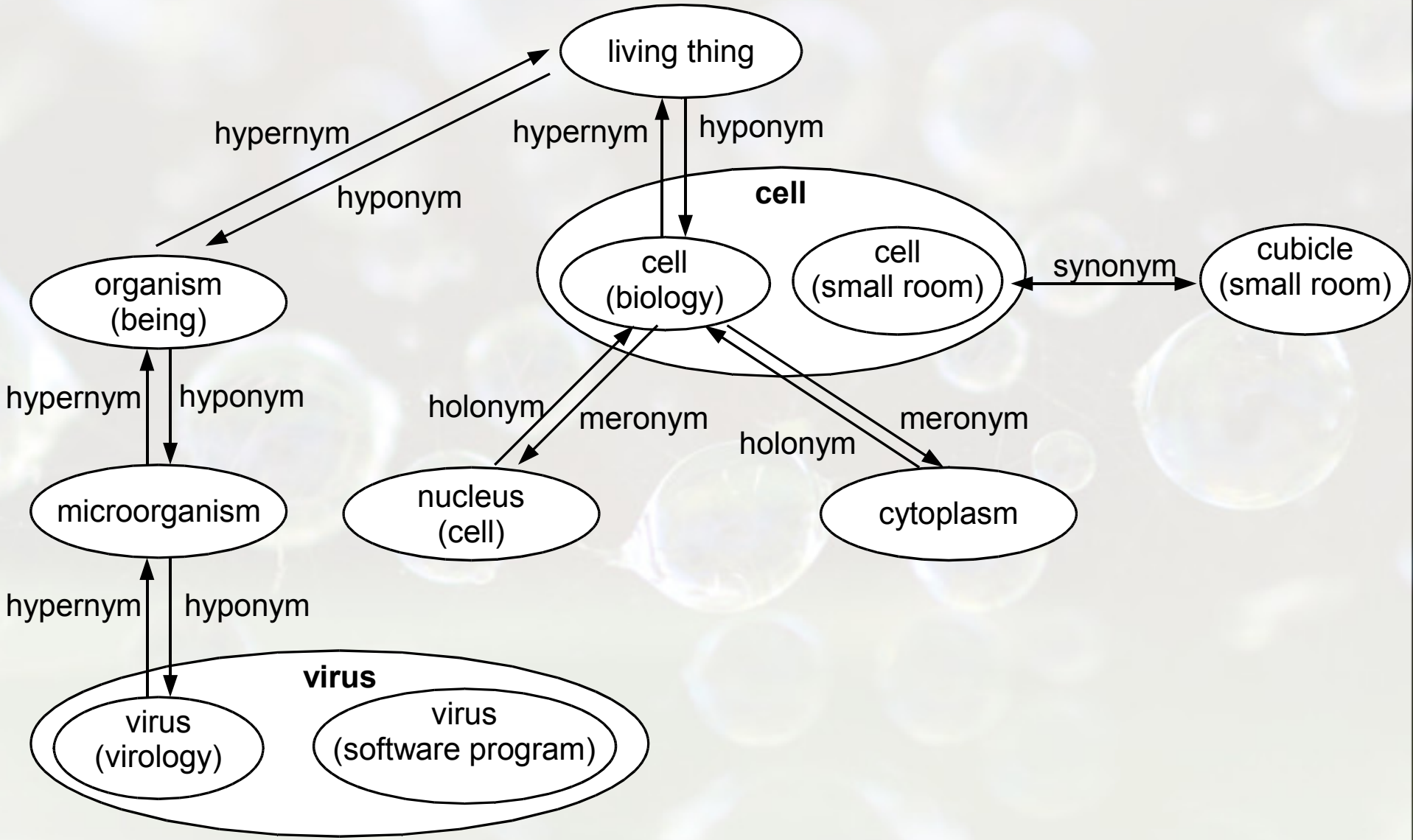
Ontologies

- Knowledge representation
- Ontology spectrum
  - Controlled vocabularies
  - Taxonomies
  - Thesaurus

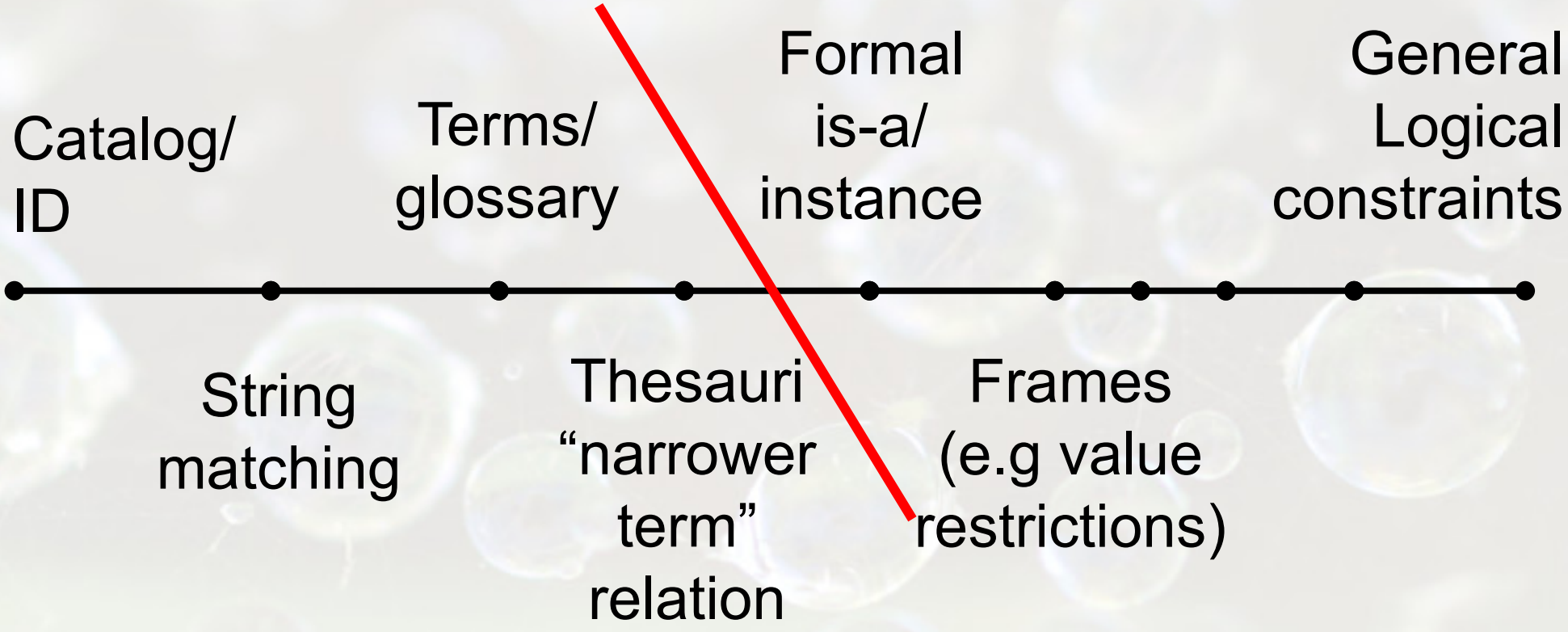
Ontologies


specification
living being
disease
chromosome
embryos
virus
living being
disease
conceptualization

WordNet


living thing
hypernym
hyponym
hypernym
hyponym
cell
cell
(biology)
cell
(small room)
synonym
cubicle
(small room)
organism
(being)
hypernym
hyponym
holonym
meronym
holonym
meronym
microorganism
nucleus
(cell)
cytoplasm
hypernym
hyponym
virus
virus
(virology)
virus
(software program)

Ontology Spectrum



(Welty et al., 1999)

Gene Ontology

Go to path: [1][2][3][4][5][6][7][8][9][10][11][12][13][14][back to top]

Term		C	T
cellular_component		19054	192748
① cell part		13	160330
① intracellular part		64	90599
① cytoplasmic part		32	36757
		767	1836
		90	124

][13][14][back to top]

Annotation

Molecular Function

Note

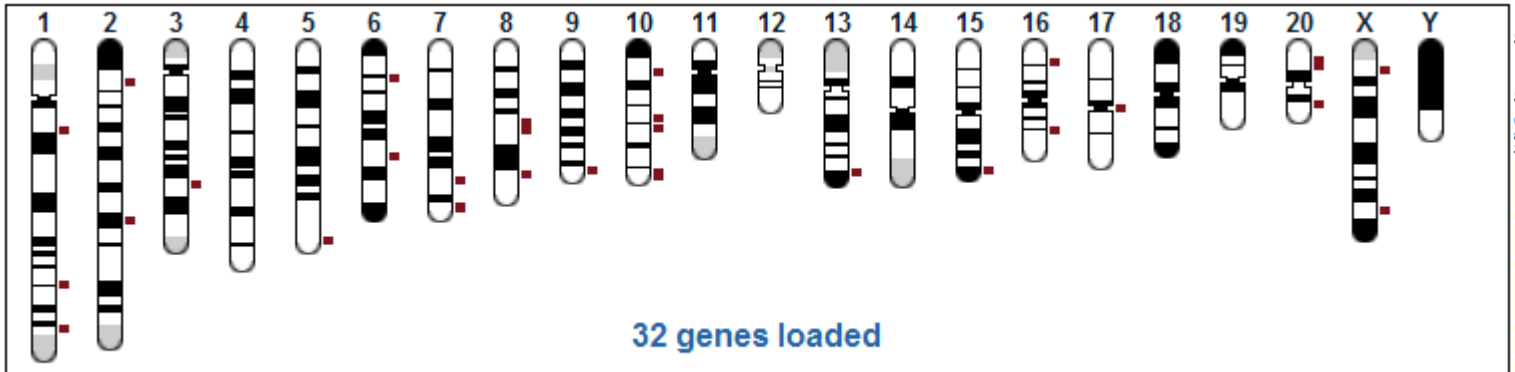
cytokine

Gene Ontology

promoter bindi

Biological

Note



List All Objects | CSV Export | Add Objects | Clear

Gene Ontology

osteoclast differentiation, positive regulation of neuron apoptosis, lation of interleukin-6 production ... (details...)

Cellular Component

Gene Ontology

recycling endosome, integral to plasma membrane, phagocytic cup ... (details...)

Disease

Note	Reference
cosegregates with various cardiovascular phenotypes and body mass in GH x BN F2 hybrids	1357163
may be involved in insulin resistance; coadministration with insulin mimics human type II diabetes	1358946

Ontology

Pulmonary Heart Disease, Nasal Polyps, Pre-Eclampsia ... (details...)

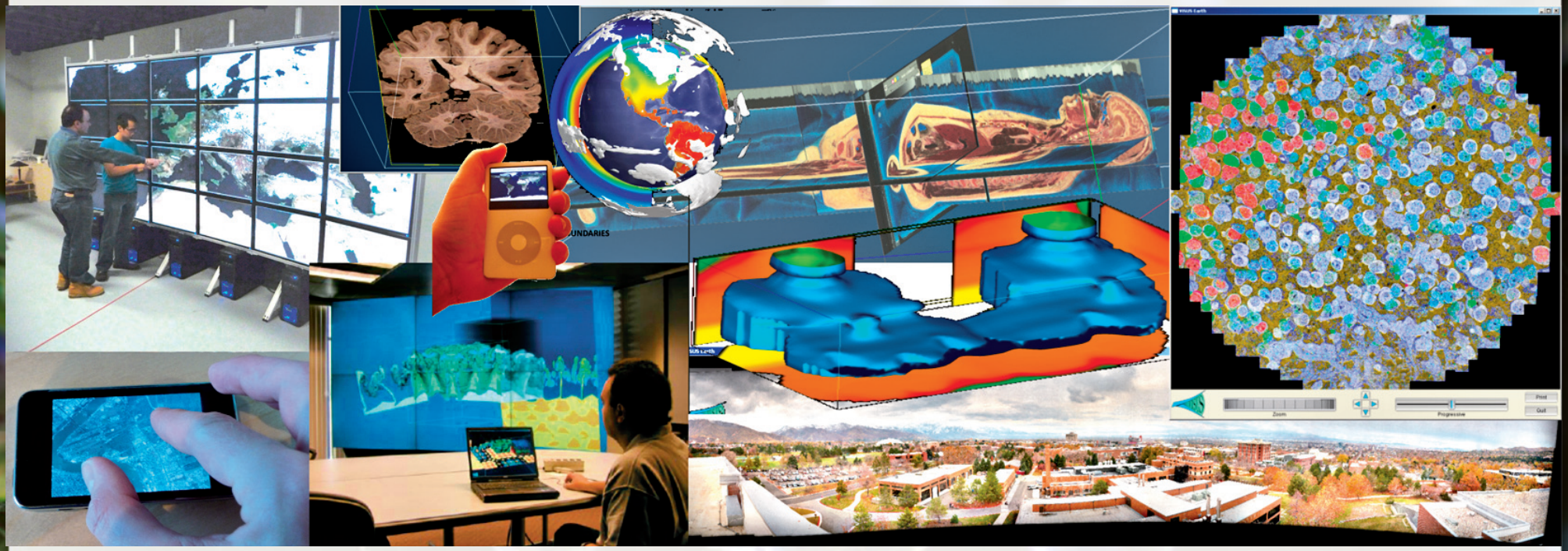
Semantic Web Services

Data Science

Digging the Web

- Information extraction
- Mining
- Searching
- Matching
- Entity resolution
- Deep Web

Fourth Paradigm



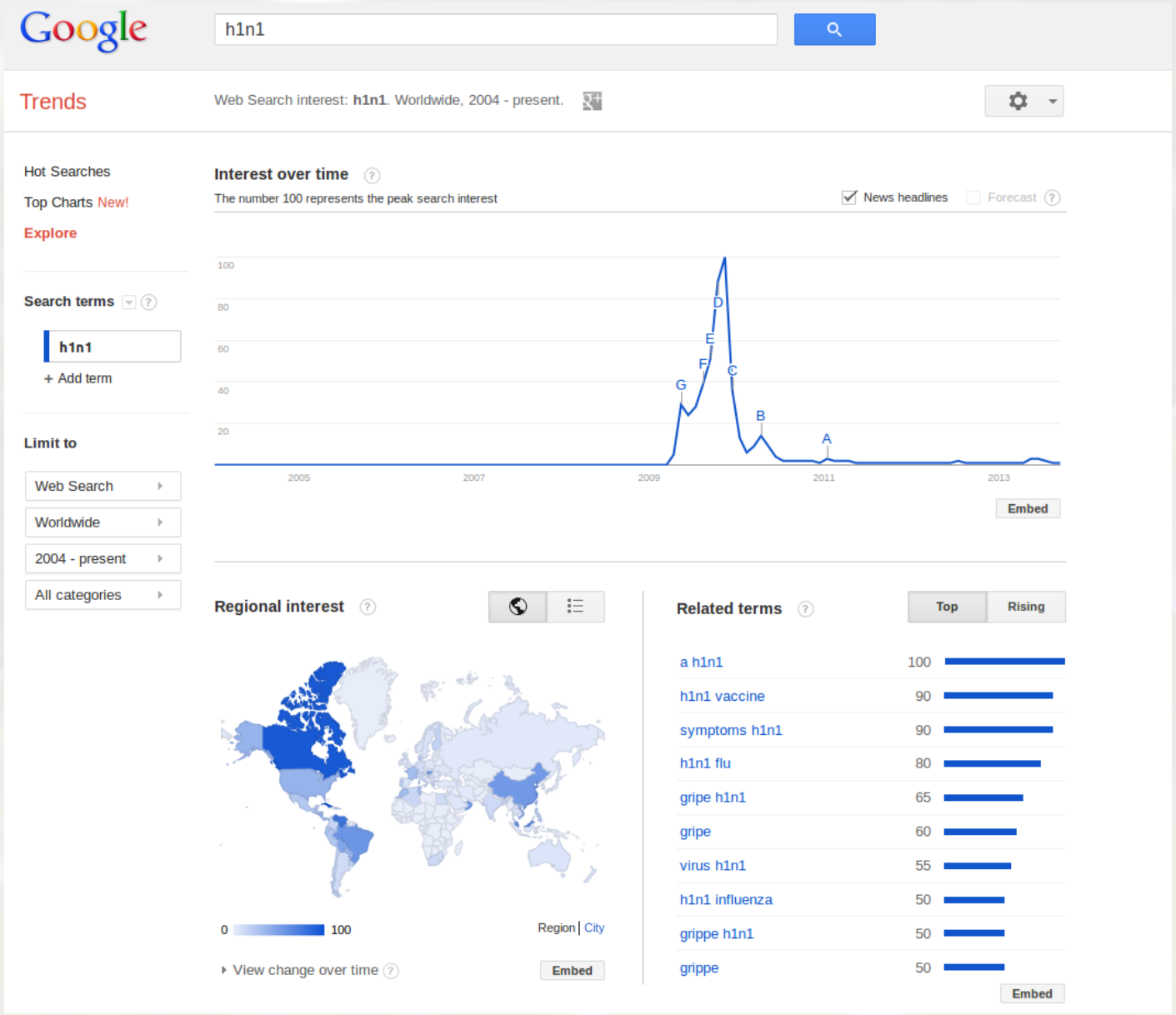
The Fourth Paradigm: Data-Intensive Scientific Discovery
Editado por Tony Hey, Stewart Tansley, and Krist in Tolle
Microsoft Research
Redmond, 2009

Fourth Paradigm

- **Thousand years ago:**
  - science was empirical; describing natural phenomena
- **Last few hundred years:**
  - theoretical branch; using models, generalizations
- **Last few decades:**
  - a computational branch; simulating complex phenomena
- **Today: data exploration (eScience)**
  - unify theory, experiment, and simulation

(Jim Gray, 2007)

Google Trends



What Wal-Mart Knows About Customers' Habits

Constance L. Hays

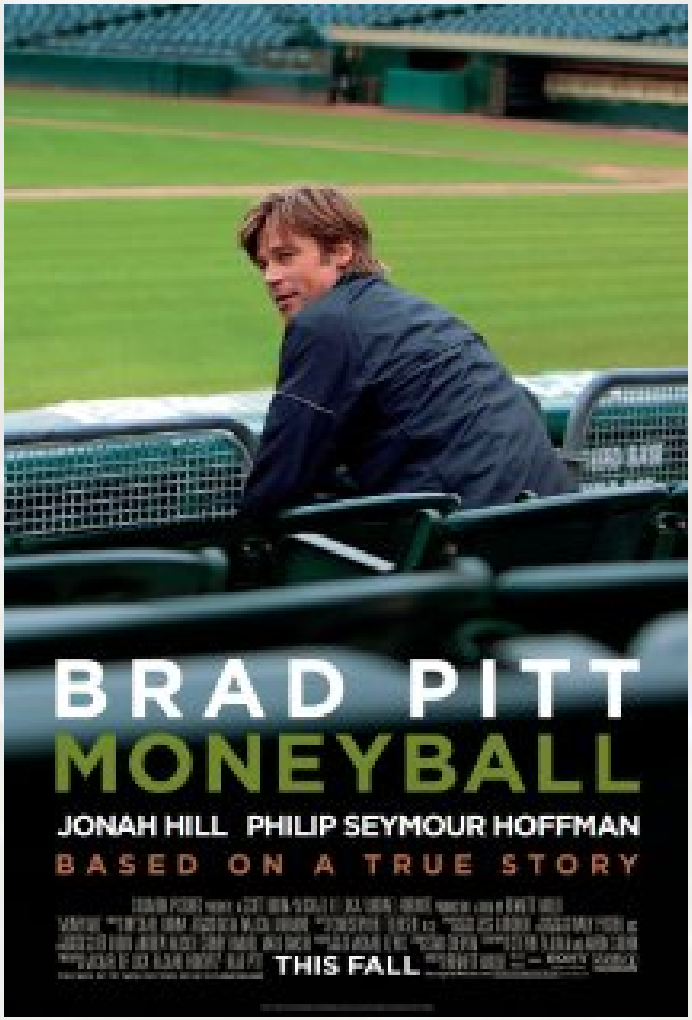
The New York Times, 2004

What Wal-Mart Knows About Customers' Habits

- "start predicting what's going to happen, instead of waiting for it to happen"
- “We didn't know in the past that strawberry Pop-Tarts increase in sales, like seven times their normal sales rate, ahead of a hurricane”
- “And the pre-hurricane top-selling item was beer”

Linda M. Dillman – Wal Mart

Data and Strategy



German and Big Data

SAP and Germany Make a Big Data Team at the World Cup

July 8, 2014 By Ben Hammonds

Sporttechie

http://www.sporttechie.com/2014/07/08/sap-and-germany-make-smart-big-data-choices-at-world-cup/

German and Big Data

- SAP is using Big Data to help the German coaching staff **make smart decisions** on tactics, player fitness, scouting, preparation as well as in game management. SAP has introduced a new concept called SAP **Match Insights** that assists players and coaches to prepare themselves for upcoming matches by dissecting key situations that may present themselves throughout the course of the match.

Our Brand Is Crisis
by Rachel Boynton

- Documentary of the 2002 Bolivian presidential election
- Gonzalo Sánchez de Lozada x Evo Morales
- Tacts by the Greenberg Carville Shrum (GCS) company

Web Observatory

Área Restrita | English Português

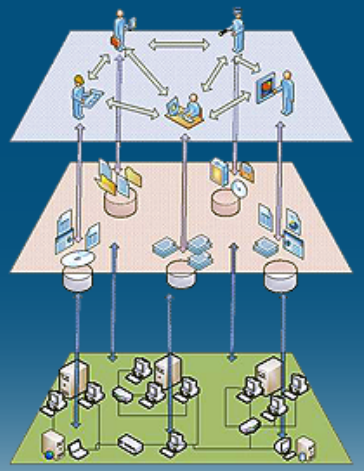
Instituto Nacional de Ciência e Tecnologia para a Web

O InWeb | Linhas de Pesquisa | Projetos | Publicações | Equipe | Eventos | Contato | blog

Web em 3 Camadas

O InWeb vê a Web como um sistema composto de múltiplas camadas de redes complexas dinâmicas e interdependentes, pelas quais a informação flui e é disseminada. A pesquisa do InWeb está foca nas camadas de interação, serviços e infra-estrutura.

[+]



1 2 3

» Home

Observatório da Dengue na Mídia

28/07/2011 | Notícias | Sem comentários

Parceria do InWeb com o Instituto Nacional da Dengue, nos últimos dias o

Enquete

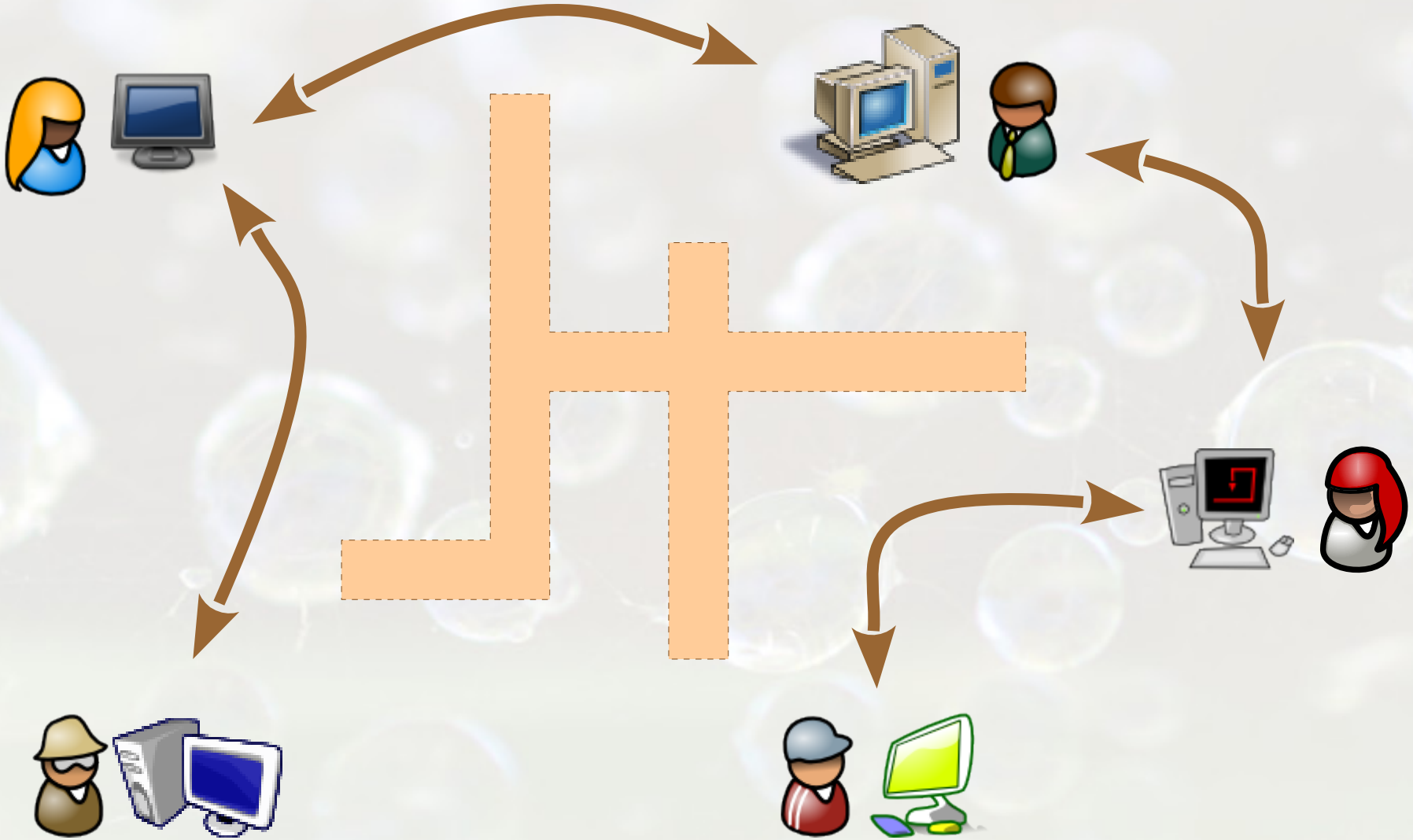
Nenhuma enquete aberta no momento.

Web Observatory

- INCT INWeb
  - http://observatorio.inweb.org.br/
  - Elections Observatory
  - Brasileirão Observatory
  - Dengue Observatory

Physical World

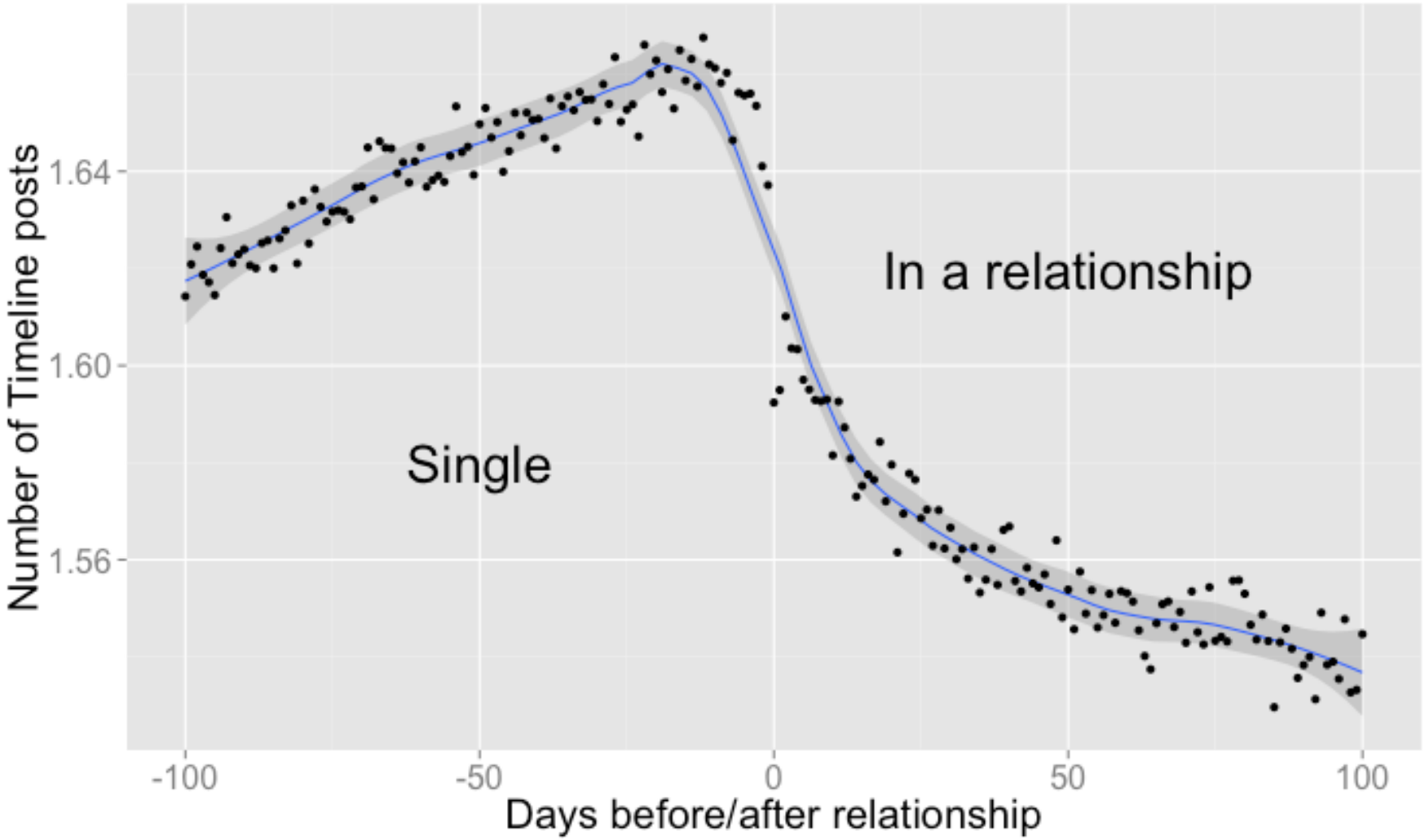
Virtual World



Facebook

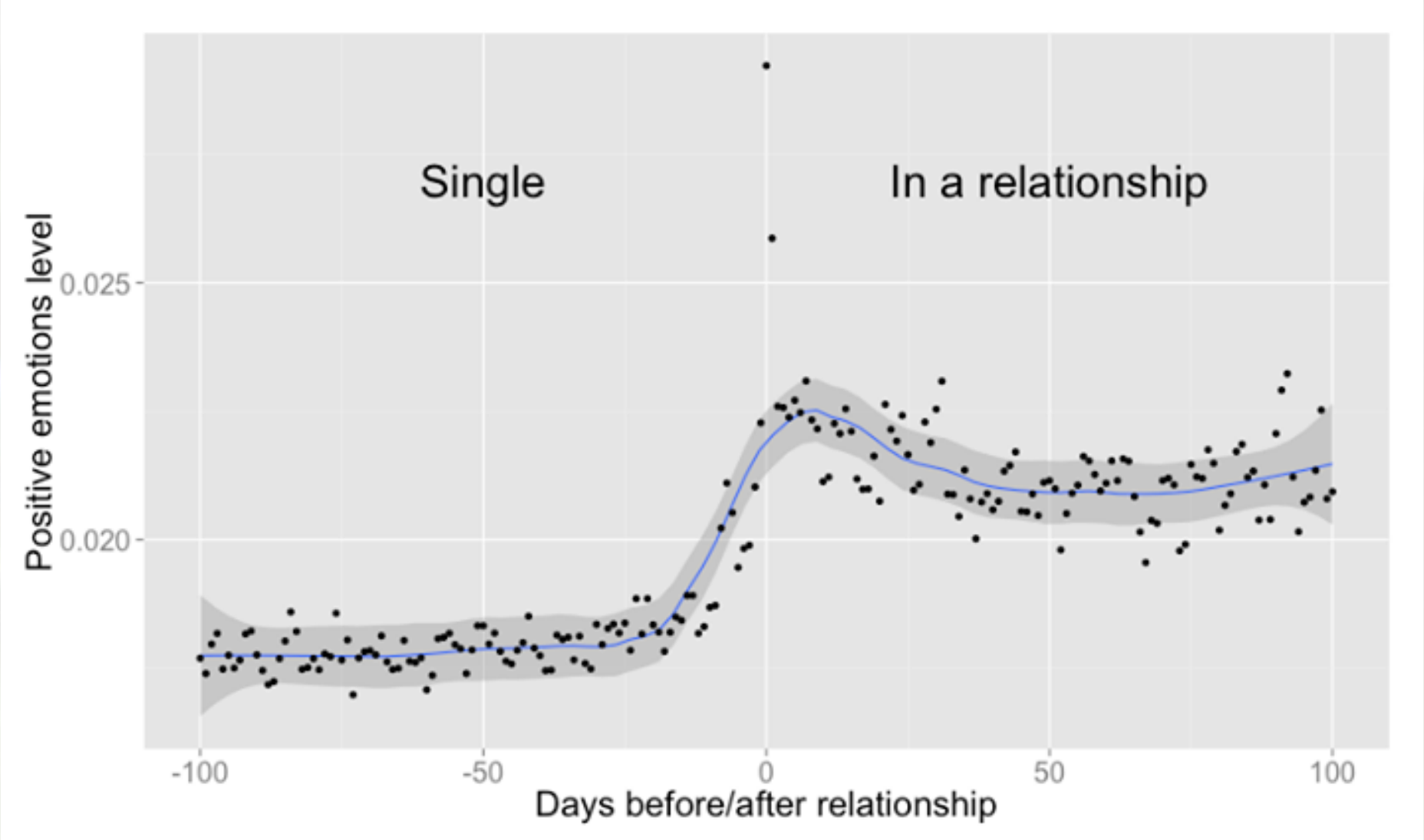
May 2013 - 1.11 billion people

Facebook



(Diuk, 2014)

Facebook



(Diuk, 2014)

The Formation of Love

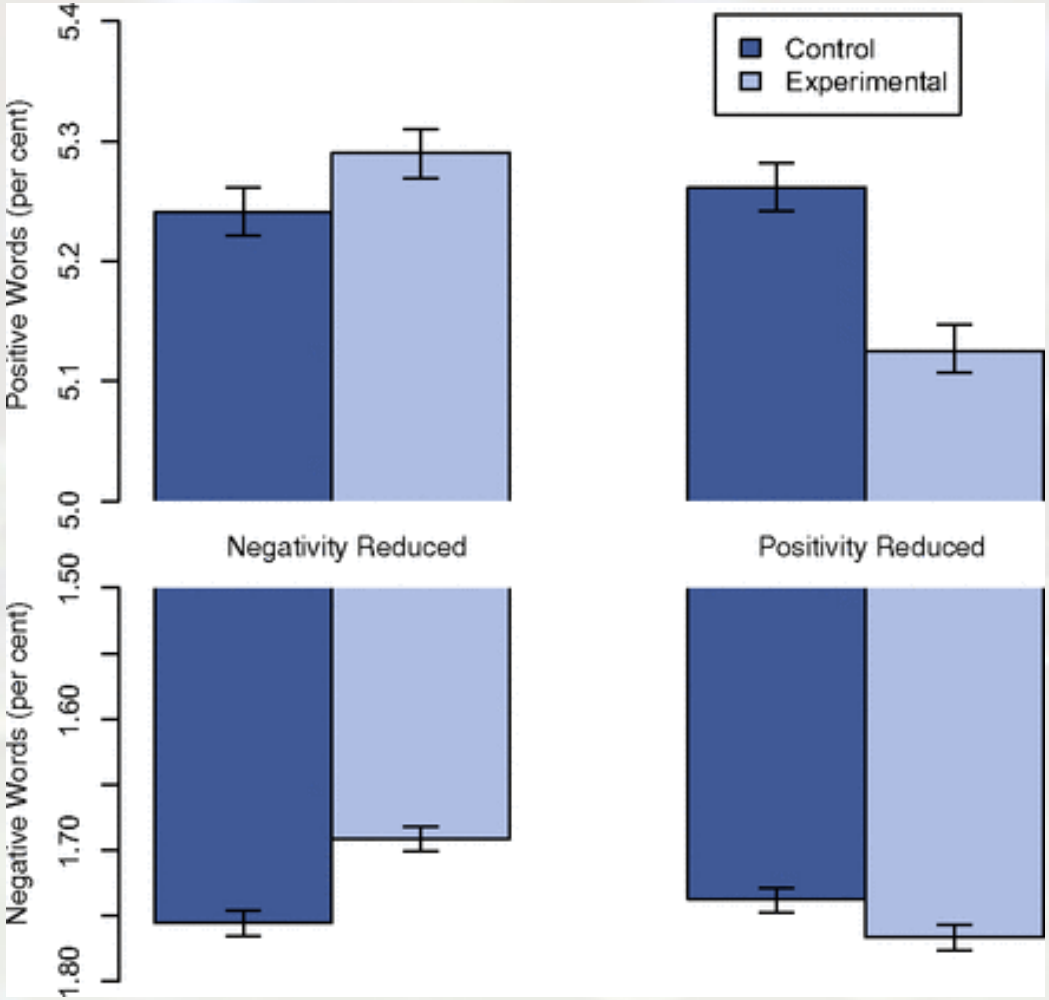
The Formation of Love

By Carlos Greg Diuk on Friday, February 14, 2014 at 3:59pm

by Carlos Diuk, Facebook Data Science

https://www.facebook.com/notes/facebook-data-science/the-formation-of-love/10152064609253859

Massive-scale Emotional Contagion



(Adam et al., 2014)

The Formation of Love

Experimental evidence of massive-scale emotional contagion through social networks

By Adam D. I. Kramera (Facebook), Jamie E. Guillory (Cornell), and Jeffrey T. Hancock (Cornell)

Proceedings of the National Academy of Sciences of the United States of America (PNAS)

June 17, 2014 , vol. 111 no. 24

Human Genome

3.3 billions base-pairs

Genoma Humano

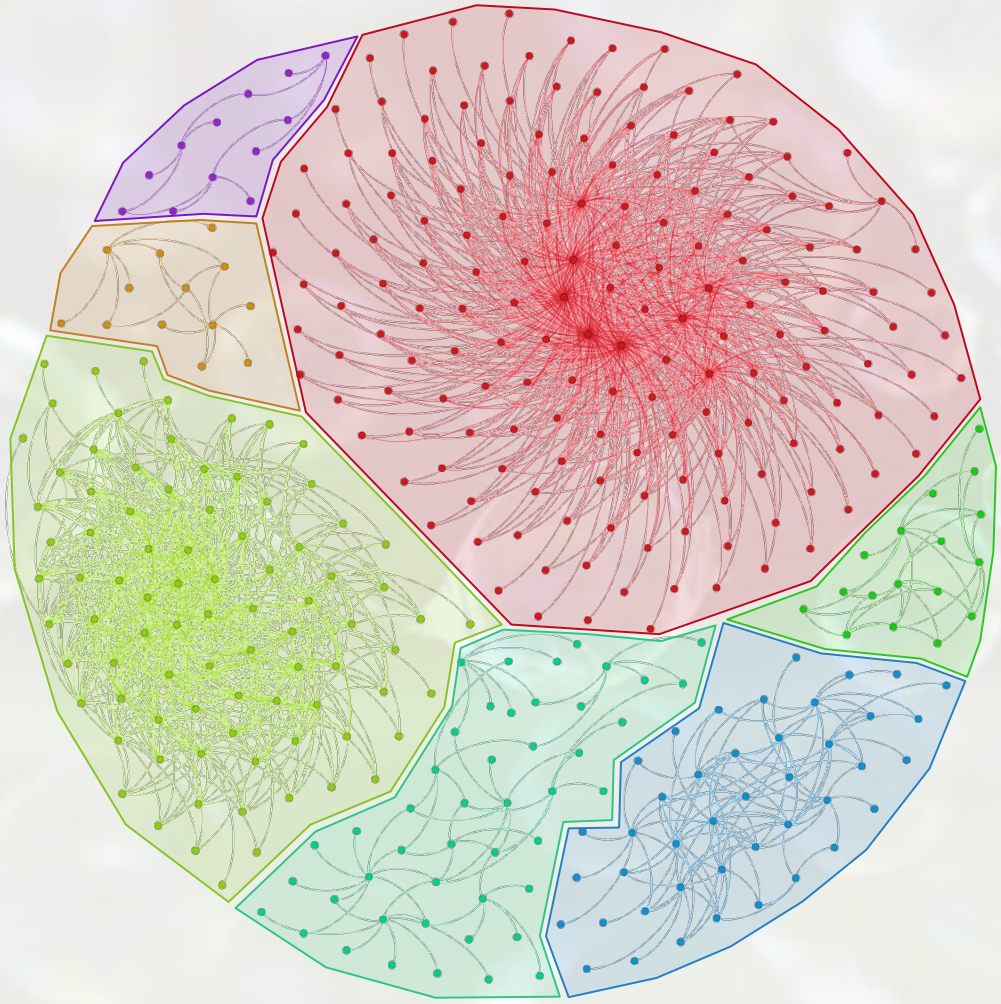
3.3 billions base-pairs

BioInformatics

Web Social Science

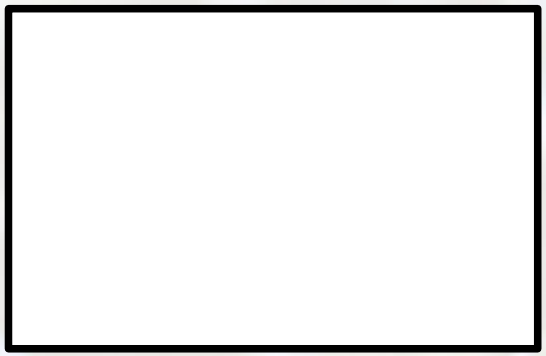
Social Networks and Social Content and Latent Semantics

Folksonomies and emergent social structures



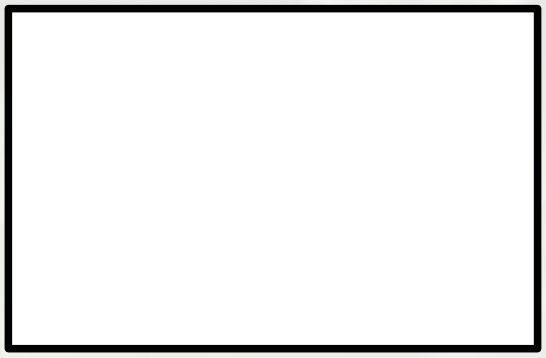
Searching **Pet**

cat, kittens, eyes, ears, pet, animal

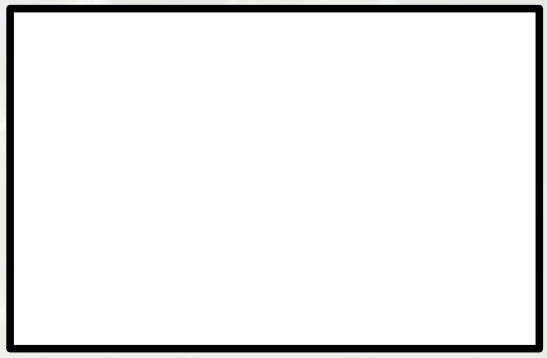


dog, pet, alaskan malamute

dog, pet, animal, funny, glasses



cat, kitty, eyes, pretty



cat, kitten, garden, pet

recycle, pet, plastic bottle,
polyethylene terephthalate

recycle, pet, plastic bottle,
polyethylene terephthalate

wine, pet, bottle

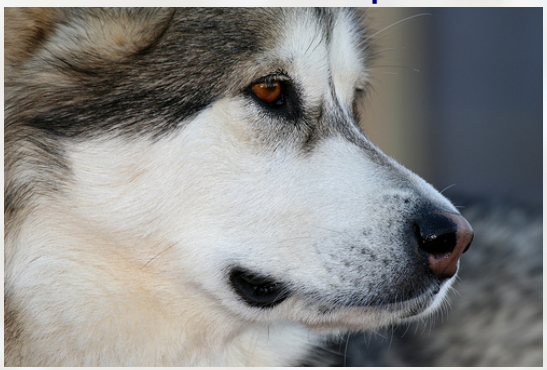
sfroehlich1121

cat, kittens, eyes, ears, pet, animal

shorty_nz_2000

dog, pet, animal, funny, glasses

Edward Corpuz



dog, pet, alaskan malamute

sfroehlich1121

cat, kitty, eyes, pretty

Jay Woodworth

cat, kitten, garden, pet

Nemo's great uncle

recycle, pet, plastic bottle, polyethylene terephthalate

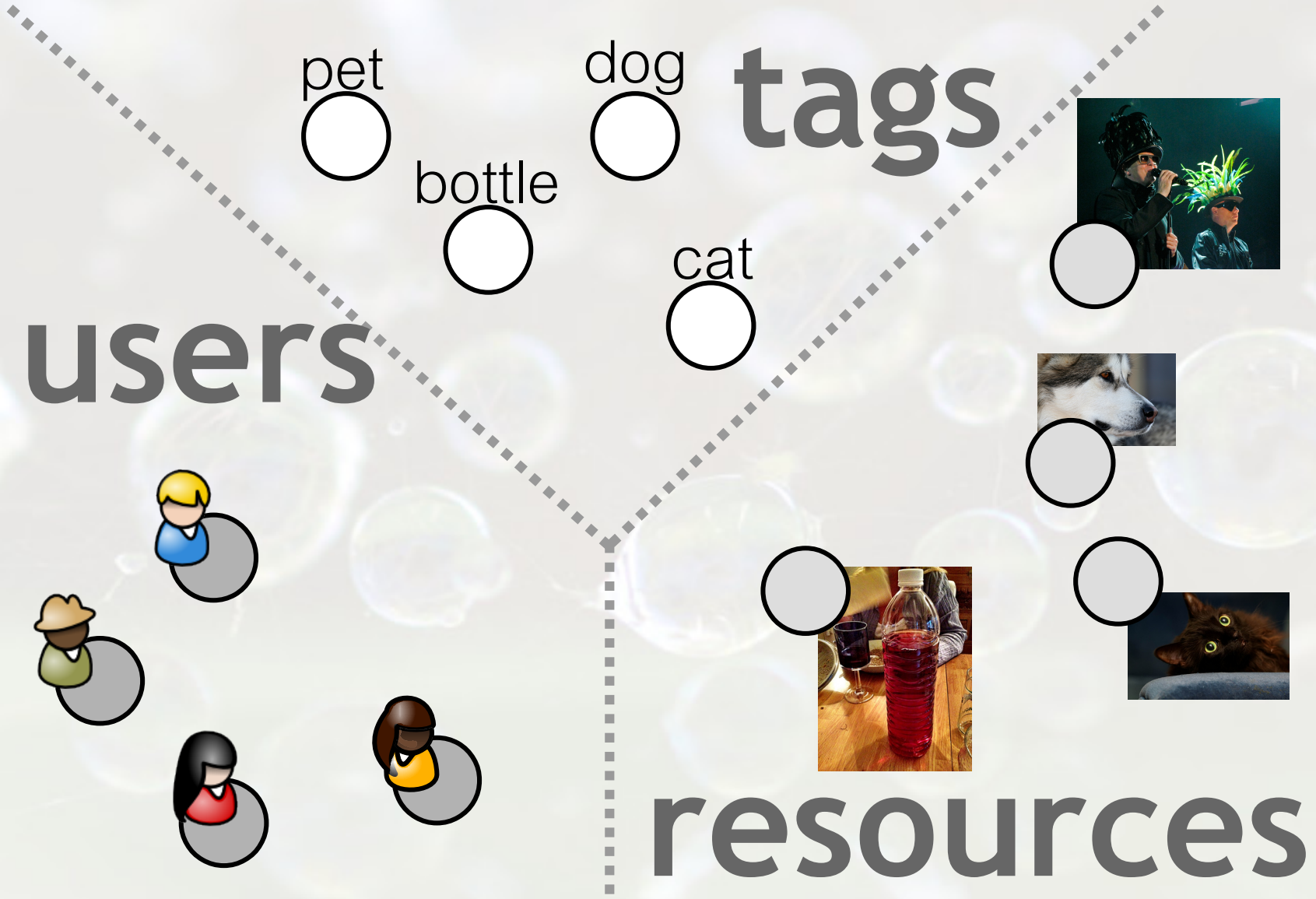
Karl Baron

wine, pet, bottle

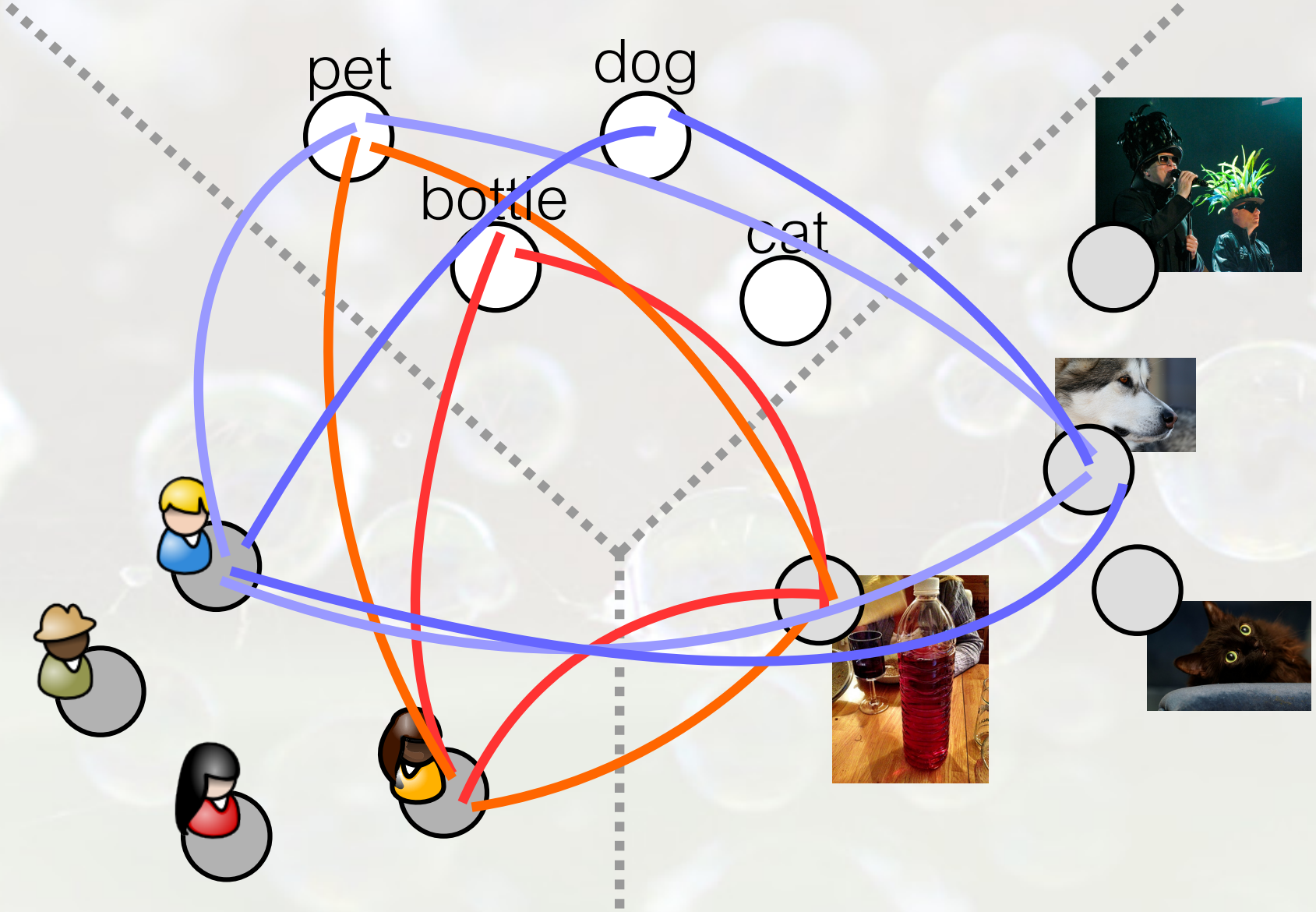
FaceMePLS

recycle, pet, plastic bottle, polyethylene terephthalate

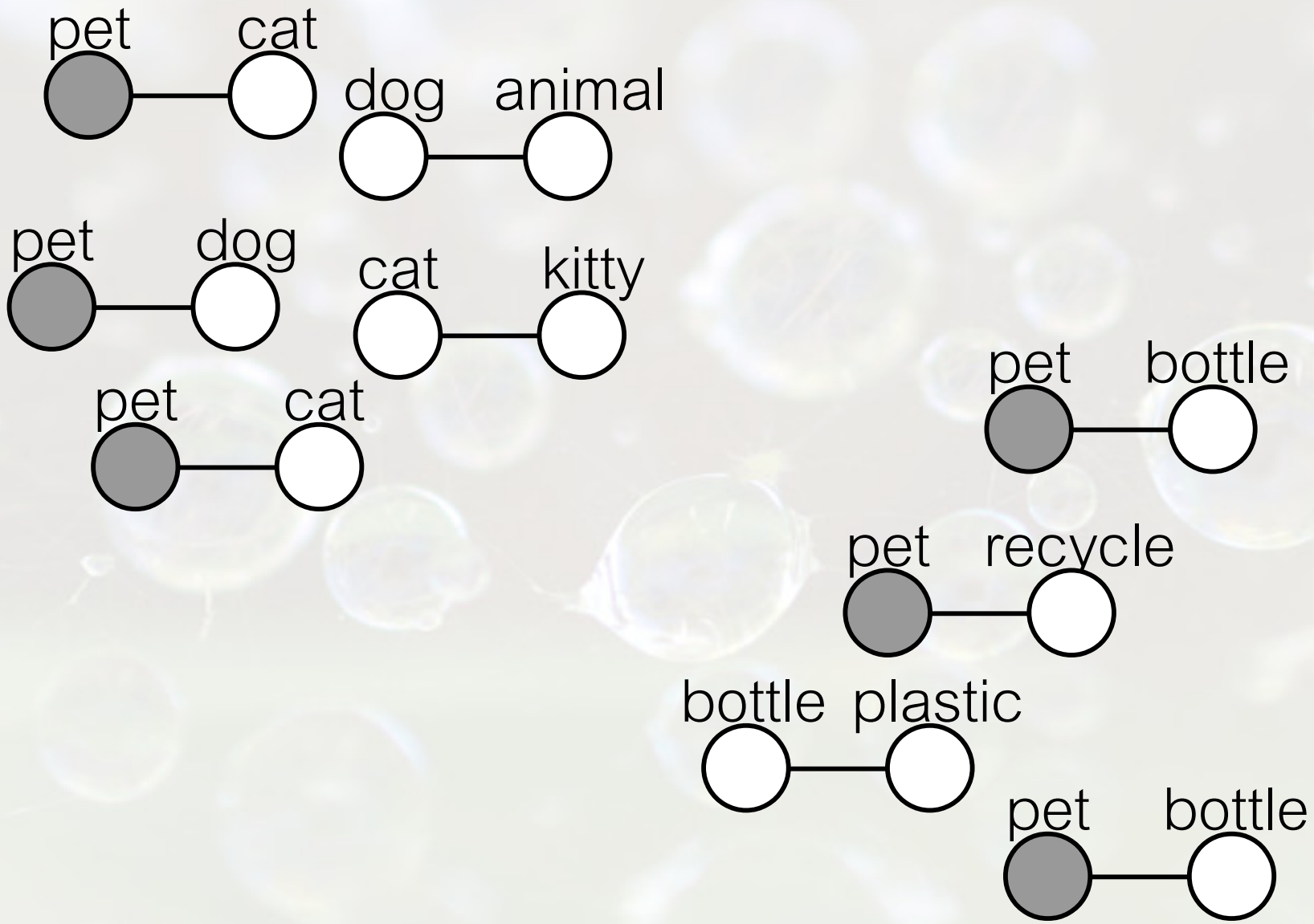
Graph users/tags/resources



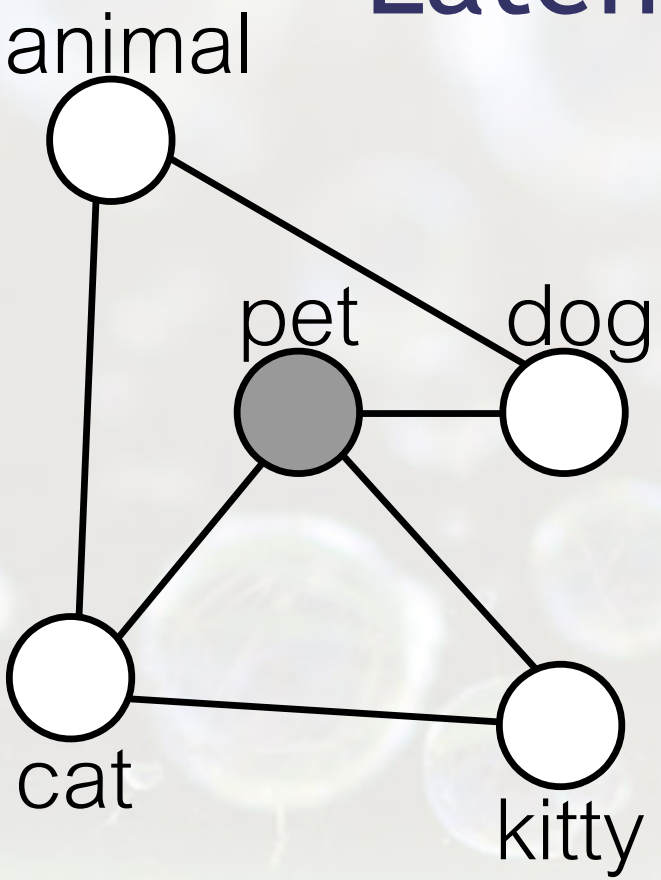
Graph users/tags/resources

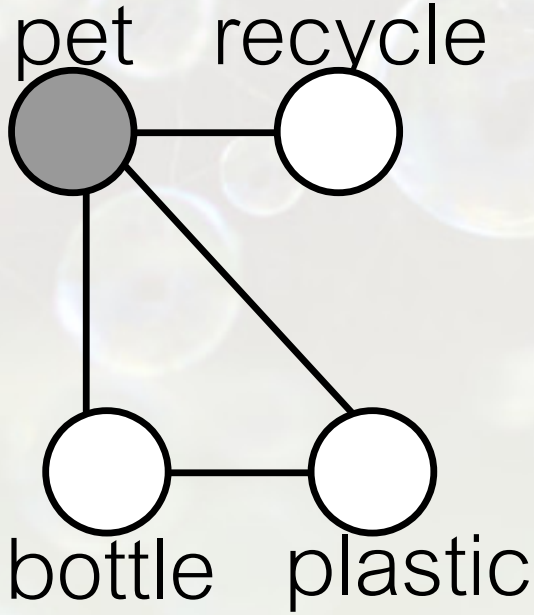


Co-ocurrences and Latent Semantics

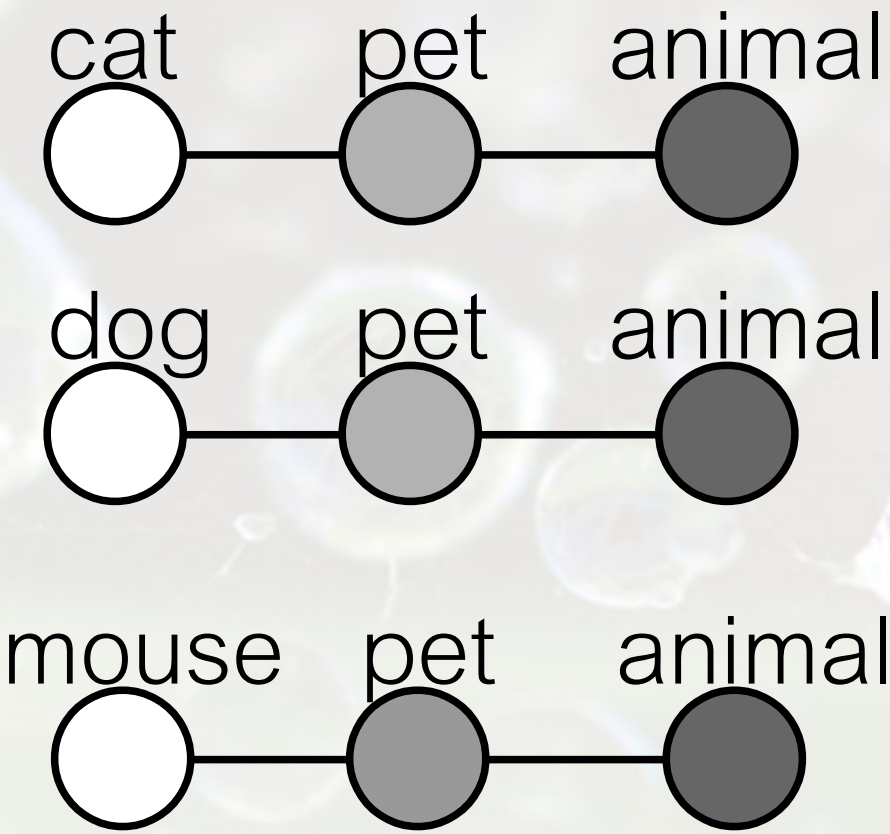


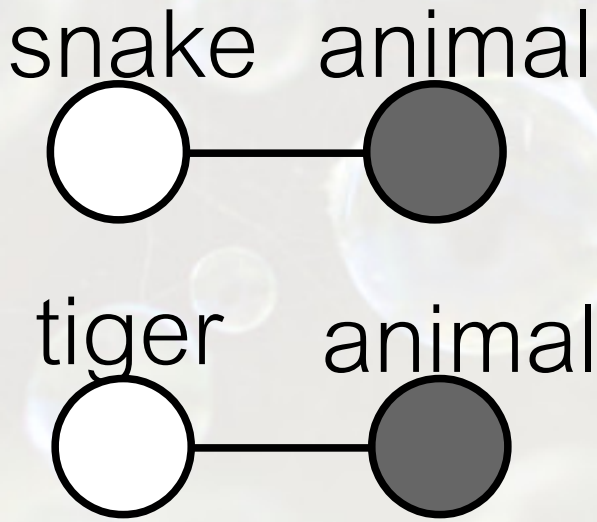
Co-ocurrences and Latent Semantics



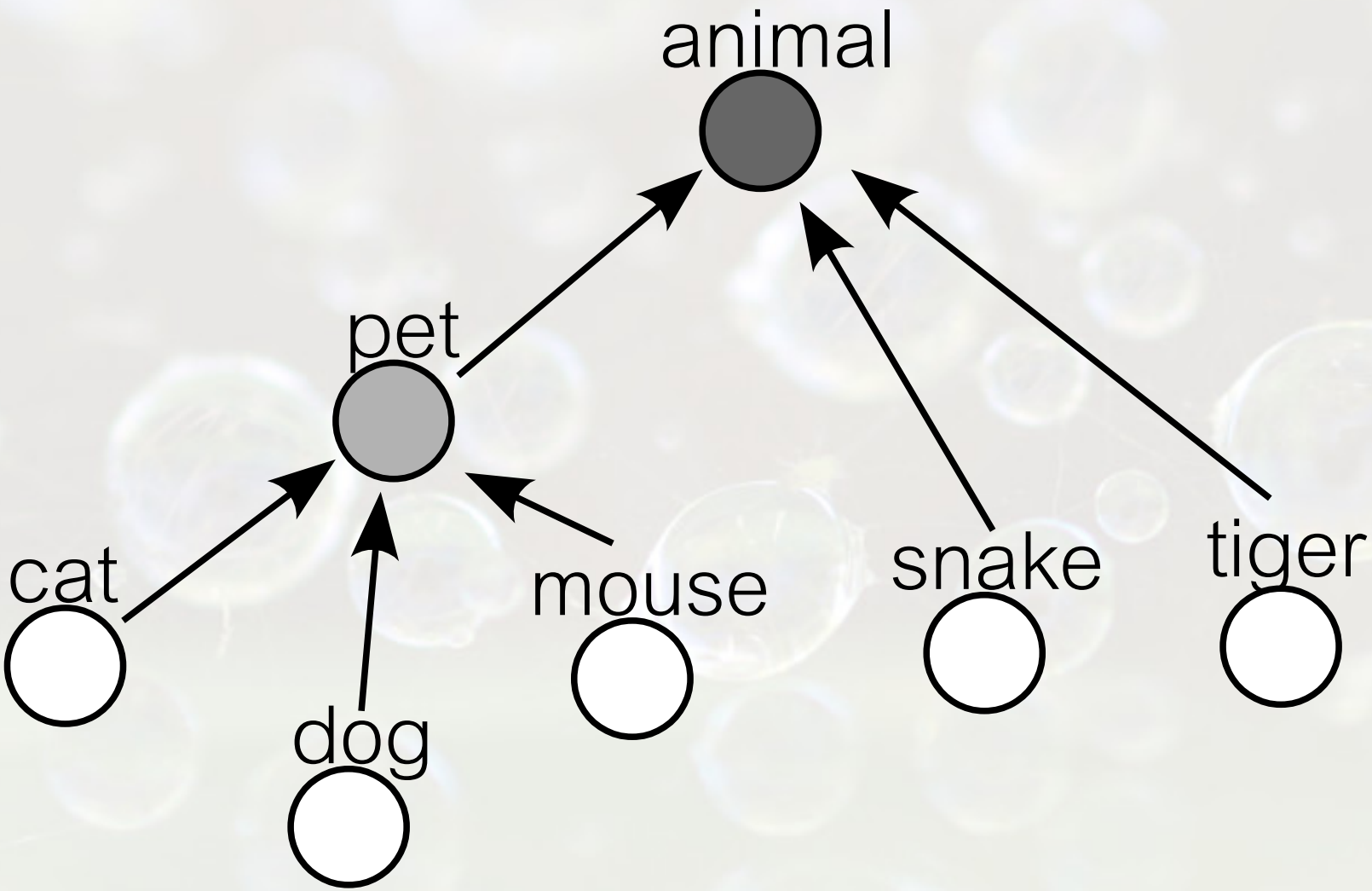


Co-ocurrences and Latent Semantics





Co-ocurrences and Latent Semantics



Social Effect
Suggested/Reinforced Tags

sfroehlich1121

black, cat, kitty, katze, long, hair, blue, eyes, pretty, canon, t1i, 500d, ef 100mm f/2.8 usm macro

Social Effect
Suggested/Reinforced Tags

sfroehlich1121

black, cat, kitty, katze, long,

hair, blue, eyes, pretty,

canon, t1i, 500d, ef 100mm f/2.8 usm macro

Limitações da Catalogação

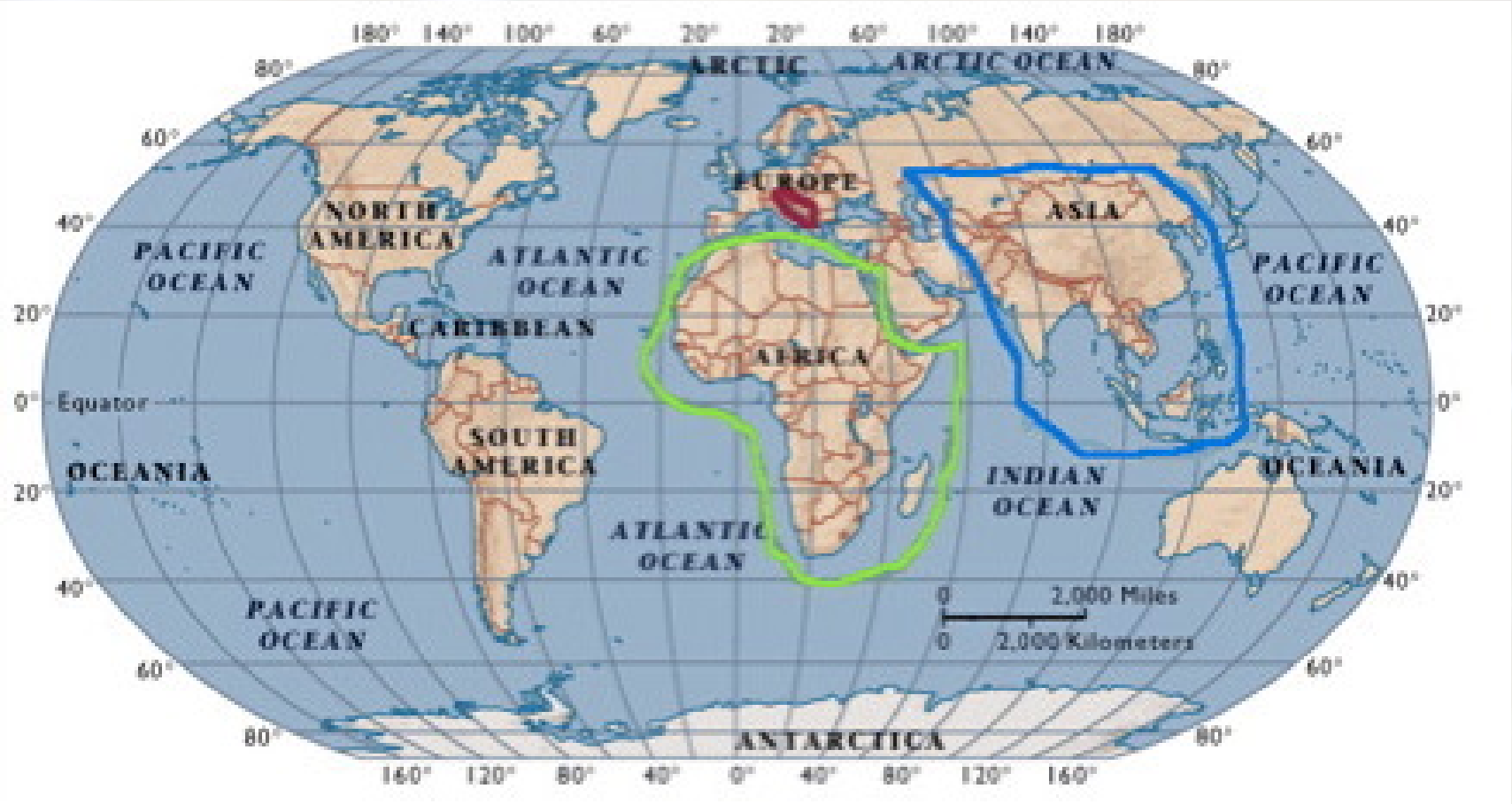
- Categorização para História na Biblioteca do Congresso Norte Americana:

D: History (general)

DA: Great Britain	DK: Former Soviet Union
DB: Austria	DL: Scandinavia
DC: France	DP: Iberian Peninsula
DD: Germany	DQ: Switzerland
DE: Mediterranean	**DR: Balkan Peninsula**
DF: Greece	**DS: Asia**
DG: Italy	**DT: Africa**
DH: Low Countries	DU: Oceania
DJ: Netherlands	DX: Gypsies

(Shirky, 2005)

Limitações da Catalogação



(Shirky, 2005)

“*there is no shelf*” (Shirky, 2005)

Em Direção ao Usuário

- Do Binário ao Probabilístico
- Abordagem Orgânica
- “Nos estamos nos afastando de uma categorização binária – livros são ou não são entretenimento – e em direção a este mundo probabilístico, em que N% dos usuários pensam que livros são entretenimento.” (Shirky, 2005)

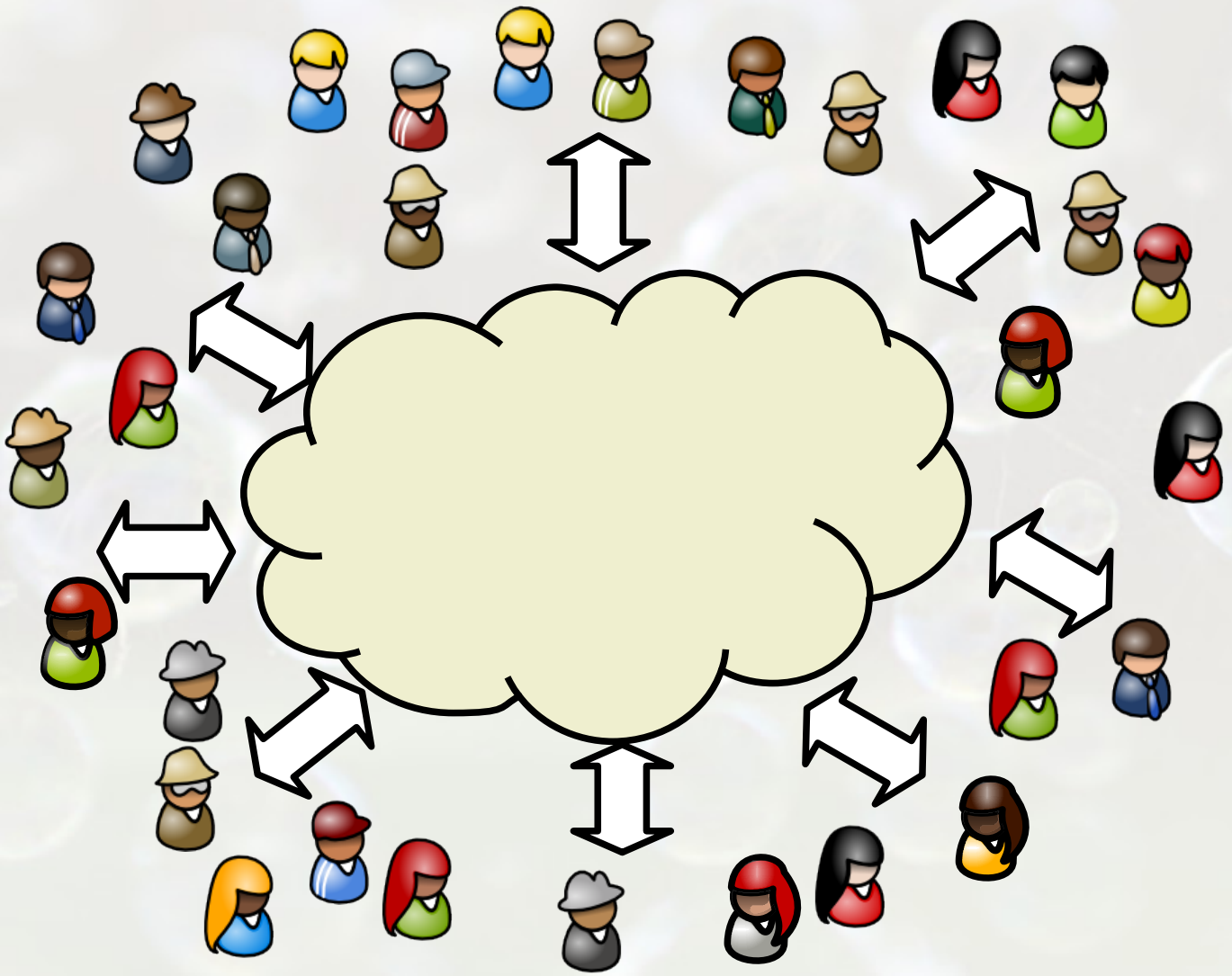
Folksonomies x Ontologies

- Folksonomies
  - tagging é feito em um ambiente social
  - as pessoas não estão bem categorizando

(Vander Wal, 2004)

- Folksonomies x Ontologias
  - Folksonomy – classificação emergente?

Crowdsourcing



Content, Behavioral and Graph analysis

- Link and tag analysis
- Link prediction
- Sentiment analysis

Network Science

Complex Networks

- Developed steadily since 1999
- Discrete systems are represented in terms of entities and relationships

(Luciano da F. Costa, 2013)

Complex Networks

- Network with non-trivial topological features

Complex Networks

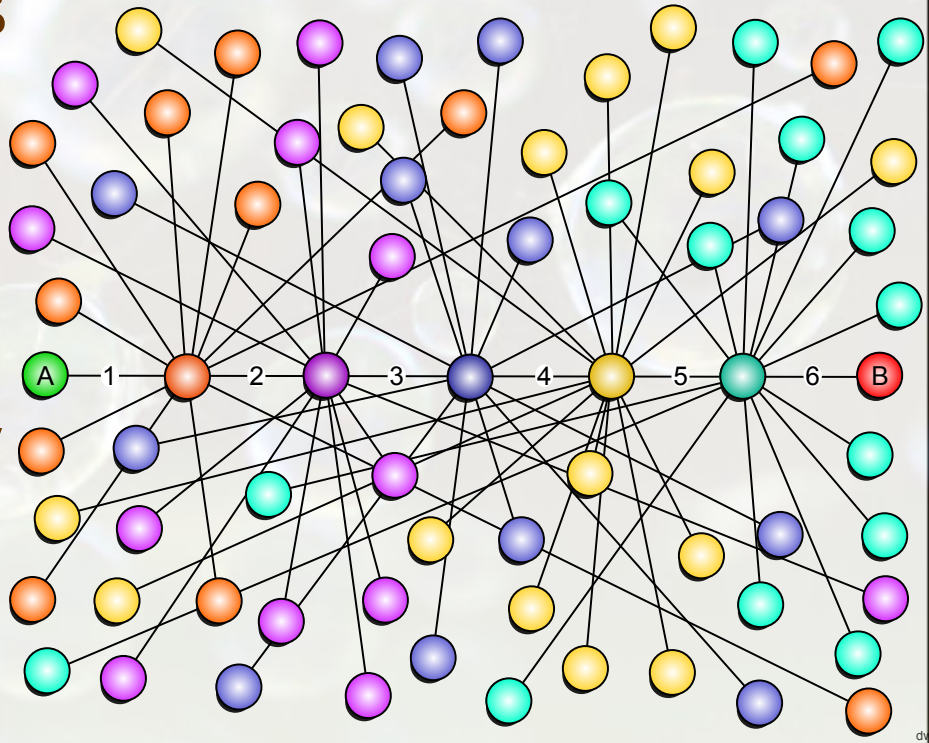
- Network with **non-trivial topological features**
  - “Real world”-like networks
    - non purely regular
    - Non purely random
  - Scale-free Property
  - Small-world Network

Complex Networks
Scale-free Property

- Degree distribution → power law

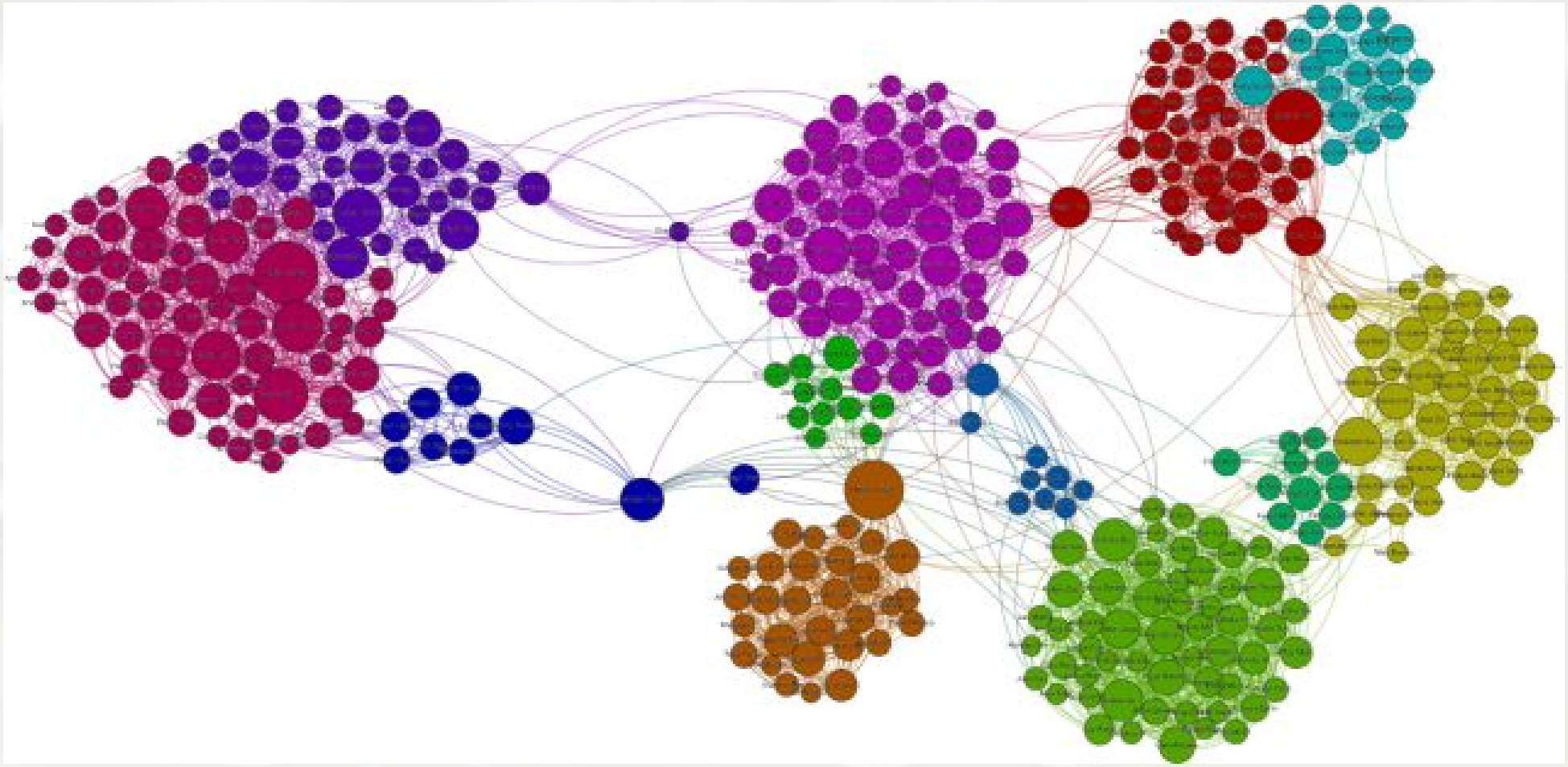
Small-world Network

- Six degrees of separation
  - The small world hypothesis
  - Everybody (everything) is at most six steps way
  - Described by the writer Frigyes Karinthy (1929)
  - Tested experimentally by Stanley Milgram (1967)



Daniel-walker [http://commons.wikimedia.org/w/index.php?title=User:Dannie-walker]
Adapted Laurensvan Lieshout [http://commons.wikimedia.org/wiki/User:LaurensvanLieshout]

Graph Topology



Metrics

- relevance
- connectivity
- reputation
- influence
- similarity

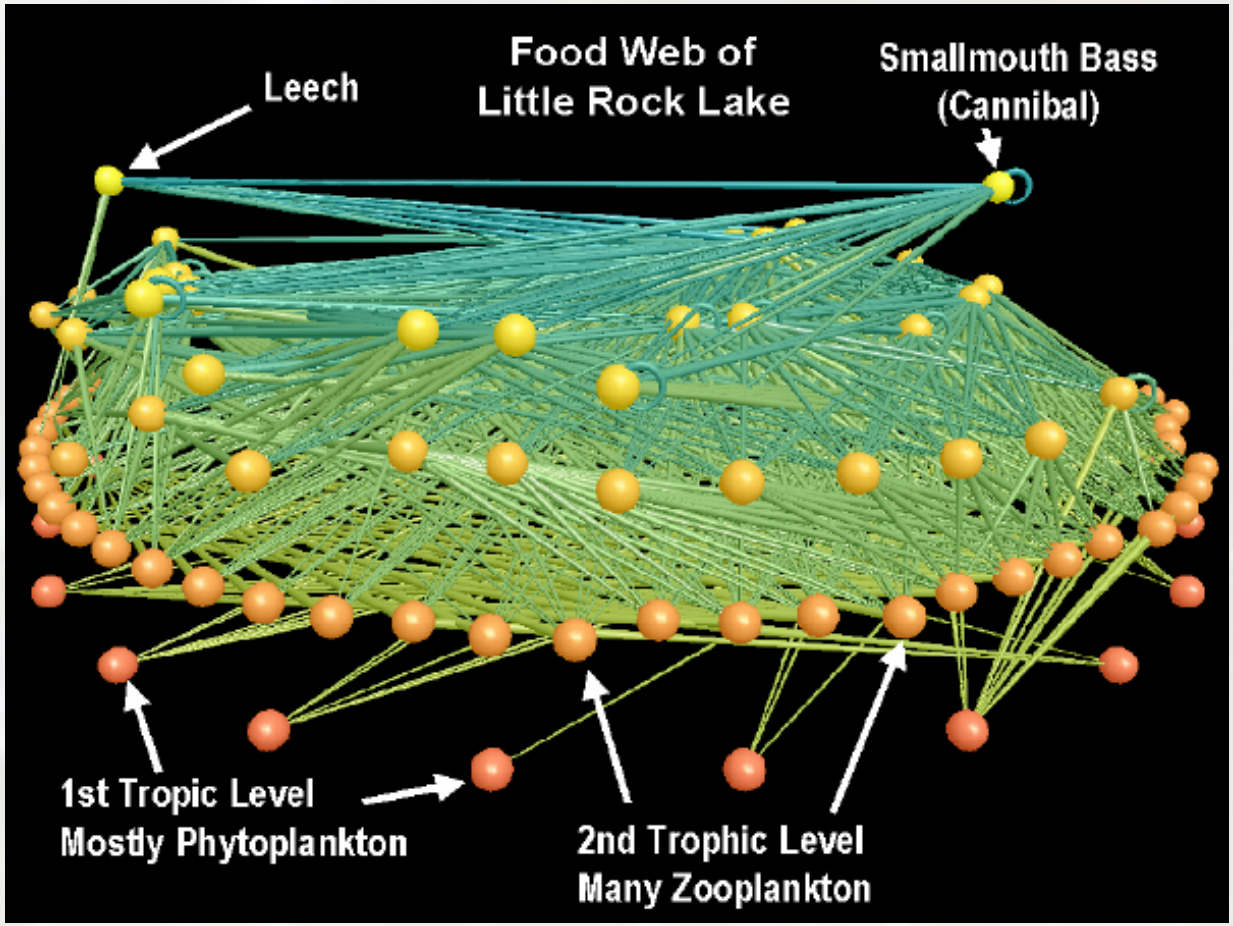
Complex Networks
Examples

http://www-personal.umich.edu/~mejn/networks/

Complex Networks

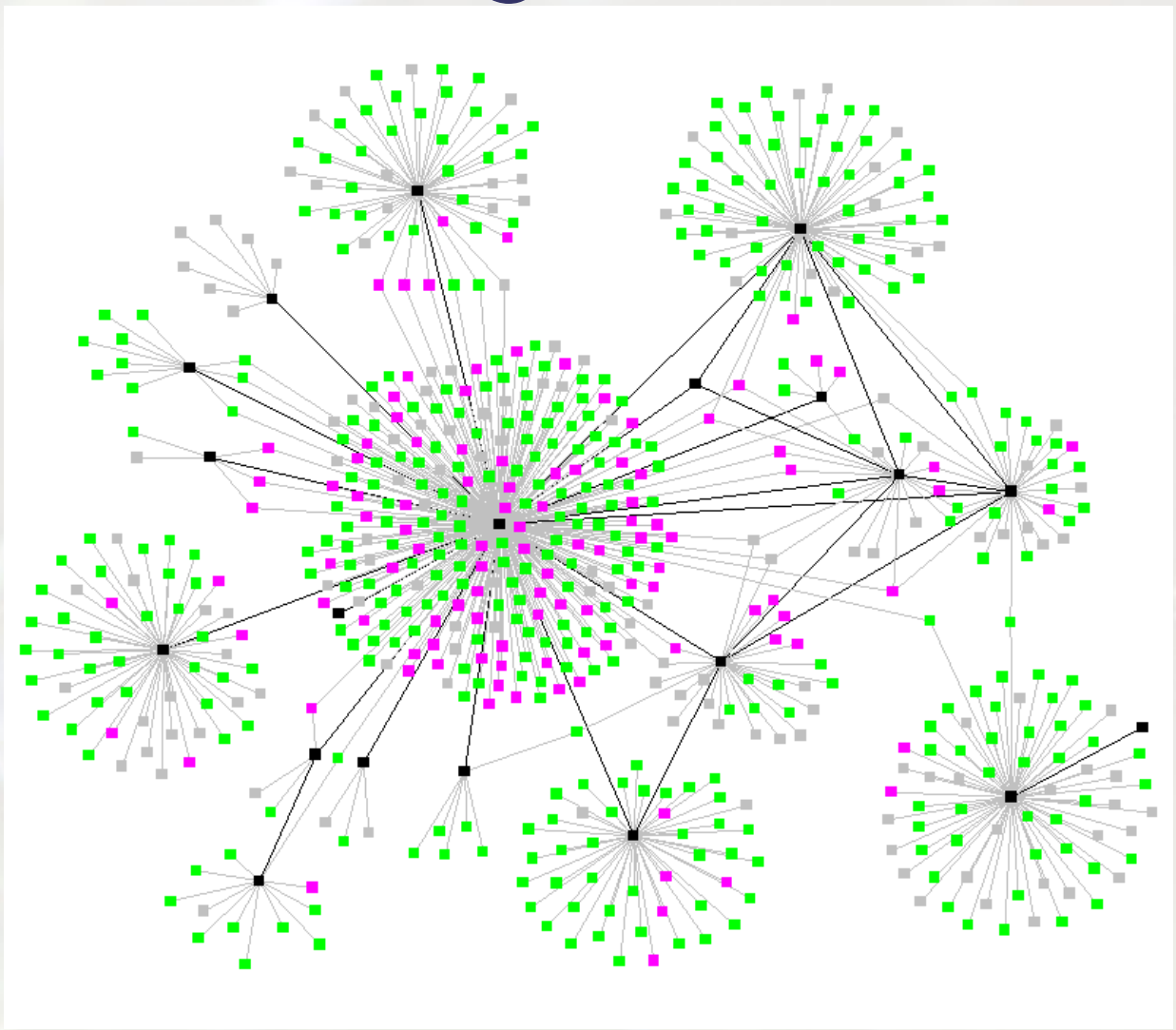
- Physical relationships
  - neurons – nodes; connections – edges
- Force relationships
  - grains – nodes; force vectors – edges
- Social relationships
- Conceptual relationships

Freshwater food web



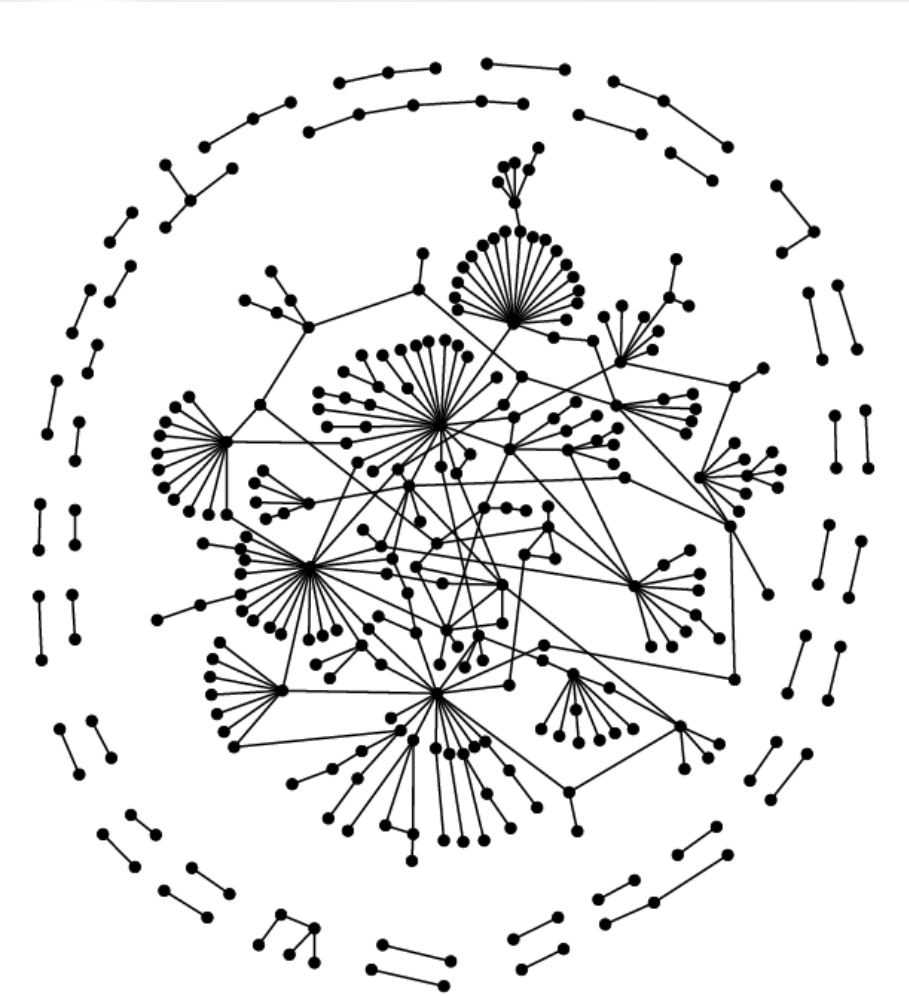
Freshwater food web: Neo Martinez and Richard Williams.

Contagion of TB



Contagion of TB, books on politics: Valdis Krebs, www.orgnet.com.

Yeast proteins



Yeast proteins: Sergei Maslov and Kim Sneppen,
Specificity and stability in topology of protein networks,
Science 296, 910-913 (2002).

André Santanchè
http://www.ic.unicamp.br/~santanche

Licença

- Estes slides são concedidos sob uma Licença Creative Commons. Sob as seguintes condições: Atribuição, Uso Não-Comercial e Compartilhamento pela mesma Licença.
- Mais detalhes sobre a referida licença Creative Commons veja no link: http://creativecommons.org/licenses/by-nc-sa/3.0/

- Fotografia de capa e fundos: web-drops por Jeremy Hiebert [http://www.flickr.com/photos/jeremyhiebert/] dispinível em http://www.flickr.com/photos/jeremyhiebert/6081389428/