

KnowledgeCoin (KC)

*a reward and recognition mechanism to
incentivize proper data integration and sharing*

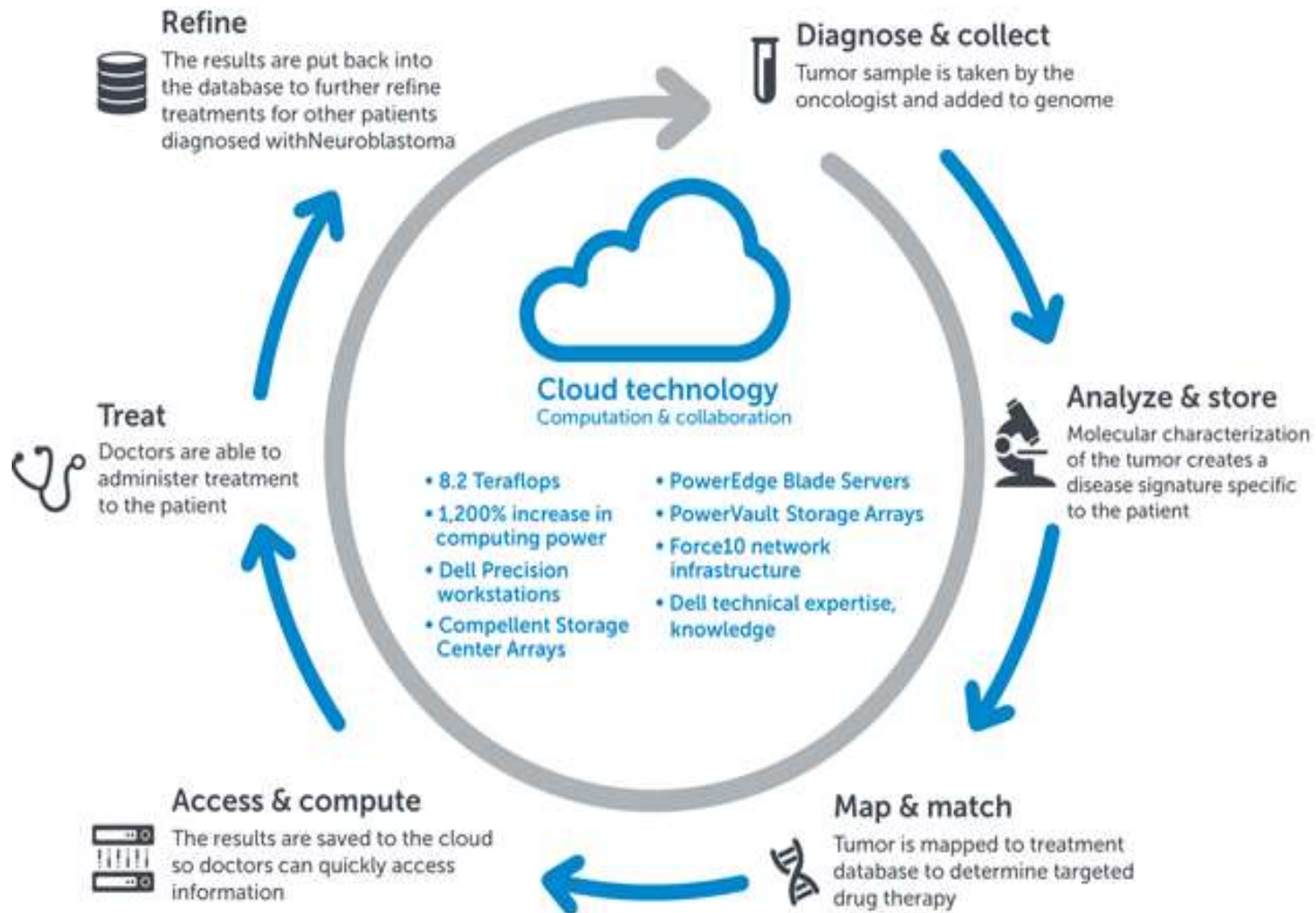
Francisco Couto

SAPO Labs Talks
14 de Outubro, 18-20h
Portugal Telecom, Picoas
Espaço Snack

DATA INTEGRATION AND SHARING

Can Cancer Treatment be Found in the Cloud?

Source: http://www.porterresearch.com/Resource_Center/Blog_News/Industry_News/2011/December/Can_Cancer_Treatment_be_Found_in_the_Cloud

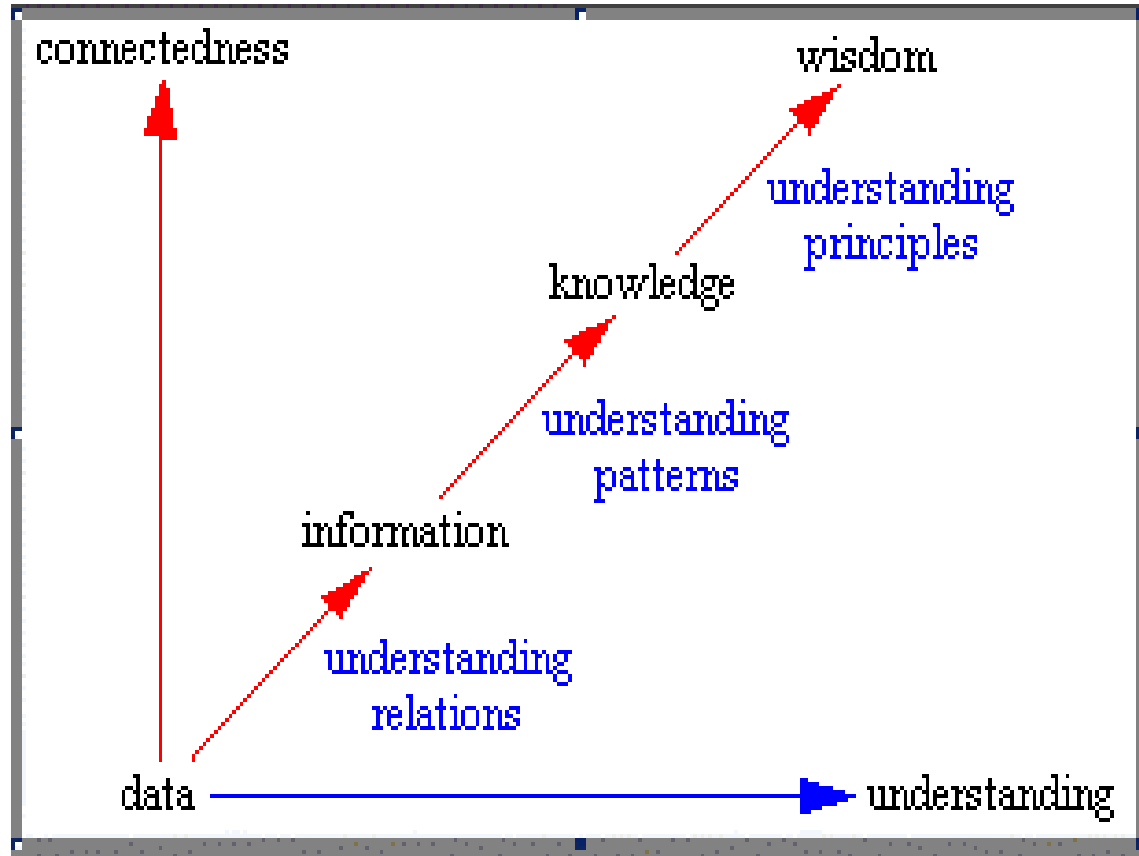


Value of Data

- the real value of data
 - can only be leveraged through its analysis
 - resulting in the acquisition of knowledge
- data integration and sharing
 - key requirements for an efficient data analysis
 - besides the big technological advances it
 - still remains an open issue

DIKW Pyramid

Data, Information, Knowledge, Wisdom

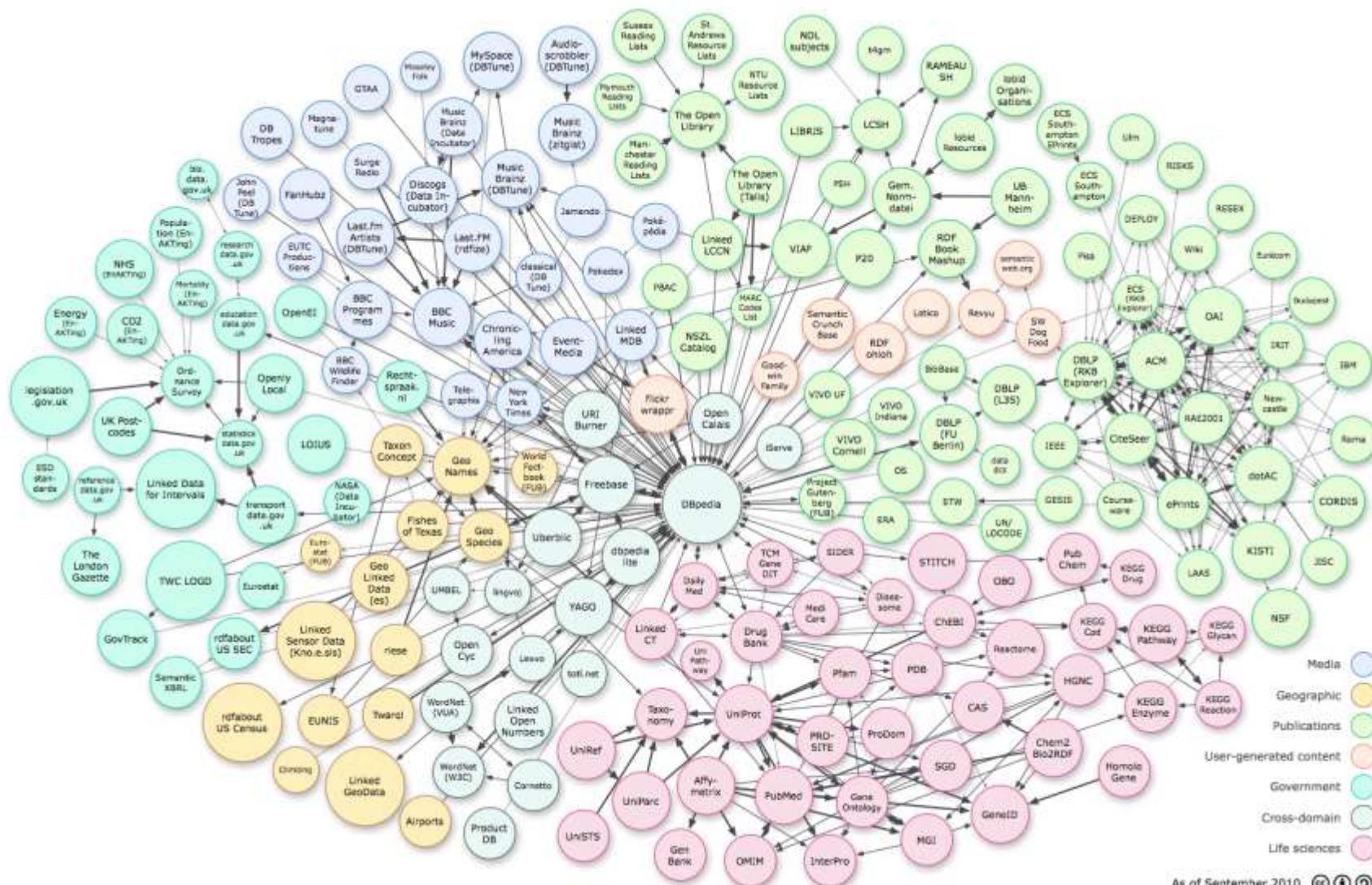


Source: <http://www.systems-thinking.org/dikw/dikw.htm>

Linked Data initiative

- well-defined set of recommendations for
 - exposing, sharing and integrating
 - data, information and knowledge
 - using **semantic web** technologies
- data integration and sharing
 - is achieved in the form of links
 - connecting the data elements themselves
 - and adding semantics to them

Linked Data Project



As of September 2010

Value of Semantics

Galileo integrated

- the direct results of his observations of Jupiter
- with careful and clear descriptions of how they were performed

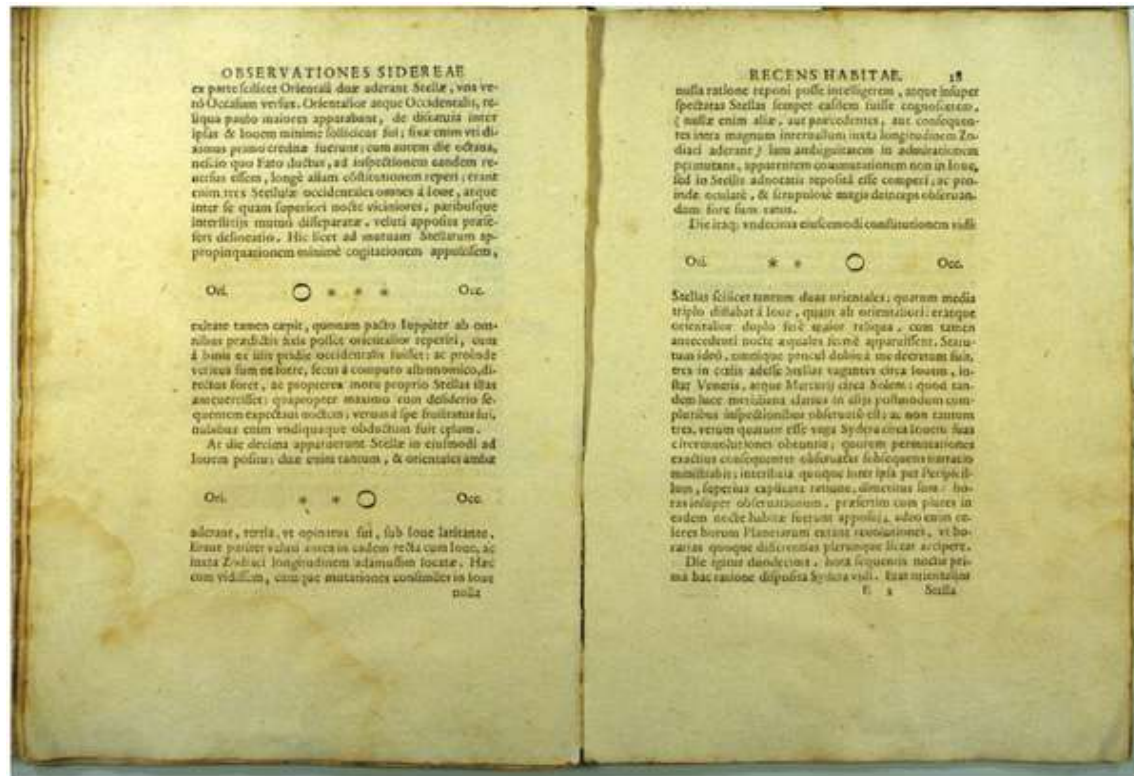


Figure 1. Two pages (scan) from Galilei's Sidereus Nuncius (“The Starry Messenger” or “The Herald of the Stars”), Venice, 1610.

SEMANTIC WEB

World Wide Web > Inventors

Robert Cailliau



Tim Berners-Lee



Cookies help us deliver our services. By using our services, you agree to our use of cookies.
[Learn more](#) [Got it](#)

History of the Web – World Wide Web Foundation

webfoundation.org/about/vision/history-of-the-web/

Tim Berners-Lee invented the World Wide Web in 1989, about 20 years after the first connection was established over what is today known as the Internet. At the time, Tim was a software engineer at CERN, the large particle physics laboratory near Geneva, Switzerland.

World Wide Web - Wikipedia, the free encyclopedia

en.wikipedia.org/wiki/World_Wide_Web

The **World Wide Web** Consortium (W3C) was founded by Tim Berners-Lee after he ...
Connected by the existing Internet, other websites were created around the ...

[History of the World Wide Web](#) - [Hypertext](#) - [Robert Cailliau](#) - [WWW \(disambiguation\)](#)

Tim Berners-Lee - Wikipedia, the free encyclopedia

en.wikipedia.org/wiki/Tim_Berners-Lee

He was honoured as the "Inventor of the **World Wide Web**" during the 2012 Summer Olympics "Tim Berners-Lee, Robert Cailliau, and the **World Wide Web**".

[First International Conference](#) - [Conway Berners-Lee](#) - [DFBCS](#) - [Mary Lee Woods](#)

[History of the World Wide Web](#) - Wikipedia, the free ...

World Wide Web

Computer



The World Wide Web is a system of interlinked hypertext documents that are accessed via the Internet. With a web browser, one can view web pages that may contain text, images, videos, and other multimedia and navigate between them via hyperlinks. [Wikipedia](#)

[Feedback](#)



World Wide Web > Inventors > Tim Berners-Lee

Robert Cailliau



Tim Berners-Lee



Cookies help us deliver our services. By using our services, you agree to our use of cookies.
[Learn more](#) [Got it](#)

Tim Berners-Lee - Wikipedia, the free encyclopedia

en.wikipedia.org/wiki/Tim_Berners-Lee

Sir Timothy John "Tim" Berners-Lee, OM, KBE, FRS, FREng, FRSA, DFBCS (born 8 June 1955) "Tim Berners-Lee, Robert Cailliau, and the World Wide Web".
First International Conference - Conway Berners-Lee - DFBCS - Mary Lee Woods

News for tim berners-lee



Sir Tim Berners-Lee speaks out on data ownership

[The Guardian](#) - 4 days ago

The data we create about ourselves should be owned by each of us, not by the large companies that harvest it, the **Tim Berners-Lee**, the ...

Tim Berners-Lee, Web Creator, Defends Net Neutrality
[New York Times \(blog\)](#) - 3 days ago

Berners-Lee: 'Computers are getting smarter. We're not'
[Telegraph.co.uk](#) - 4 days ago

[More news for tim berners-lee](#)

Tim Berners-Lee - World Wide Web Consortium

www.w3.org/People/Berners-Lee/

A graduate of Oxford University, **Tim Berners-Lee** invented the World Wide Web, an internet-based hypermedia initiative for global information sharing while at ...



[More images](#)

Tim Berners-Lee

Computer Scientist

Sir Timothy John "Tim" Berners-Lee, OM, KBE, FRS, FREng, FRSA, DFBCS, also known as "TimBL", is an English computer scientist, best known as the inventor of the World Wide Web. [Wikipedia](#)

Born: June 8, 1955 (age 59), London, United Kingdom

Nationality: British

Books: [Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by its Inventor](#)

Parents: [Mary Lee Woods](#), [Conway Berners-Lee](#)

Education: [The Queen's College, Oxford \(1973–1976\)](#), [Emanuel School \(1969–1973\)](#)



Article

Talk

Read

View source

View history

Search

Tim Berners-Lee

From Wikipedia, the free encyclopedia

Sir Timothy John "Tim" Berners-Lee, OM, KBE, FRS, FEng, FRSA, DFBCS (born 8 June 1955), also known as "**TimBL**", is an English [computer scientist](#), best known as the inventor of the [World Wide Web](#). He made a proposal for an information management system in March 1989,^[4] and he implemented the first successful communication between a [Hypertext Transfer Protocol](#) (HTTP) client and server via the Internet sometime around mid November of that same year.^{[5][6][7][8][9]}

Berners-Lee is the director of the [World Wide Web Consortium](#) (W3C), which oversees the Web's continued development. He is also the founder of the [World Wide Web Foundation](#), and is a senior researcher and holder of the [Founders Chair](#) at the MIT Computer Science and Artificial Intelligence Laboratory (CSAIL).^[10] He is a director of the [Web Science Research Initiative](#) (WSRI),^[11] and a member of the advisory board of the MIT Center for Collective Intelligence.^{[12][13]}

In 2004, Berners-Lee was [knighted](#) by Queen [Elizabeth II](#) for his pioneering work.^{[14][15]} In April 2009, he was elected a foreign associate of the [United States National Academy of Sciences](#).^{[16][17]} He was honoured as the "Inventor of the World Wide Web" during the 2012 Summer Olympics opening ceremony, in which he appeared in person, working with a vintage NeXT Computer at the [London Olympic Stadium](#).^[18] He tweeted "This is for everyone",^[19] which instantly was spelled out in LCD lights attached to the chairs of the 80,000 people in the audience.^[18]

Contents [hide]

1 Early life

2 Career

3 Current work

4 Awards and honours

Sir Tim Berners-Lee



Berners-Lee in 2012

Born	<div>Timothy John Berners-Lee</div> <div>8 June 1955 (age 59)^[1]</div> <div>London, England</div> <div>United Kingdom</div>
Residence	United Kingdom and United States ^[2]
Nationality	British
Alma mater	The Queen's College, Oxford
Occupation	Computer scientist
Employer	World Wide Web Consortium
	University of Southampton
	Plessey

- Main page
- Contents
- Featured content
- Current events
- Random article
- Donate to Wikipedia
- Wikimedia Shop

- Interaction
- Help
- About Wikipedia
- Community portal
- Recent changes
- Contact page

- Tools
- What links here
- Related changes
- Upload file
- Special pages
- Permanent link
- Page information
- Wikidata item
- Cite this page

- Print/export
- Create a book
- Download as PDF
- Printable version

- Languages

D About: Tim Berners-Lee	
dbpedia.org/page/Tim_Berners-Lee	
dbpedia-owl:alias	<ul style="list-style-type: none"> TimBL
dbpedia-owl:almaMater	<ul style="list-style-type: none"> dbpedia:The_Queen's_College,_Oxford
dbpedia-owl:award	<ul style="list-style-type: none"> dbpedia:Royal_Designers_for_Industry dbpedia:Order_of_the_British_Empire dbpedia:Royal_Academy_of_Engineering dbpedia:Order_of_Merit dbpedia:Royal_Society
dbpedia-owl:birthDate	<ul style="list-style-type: none"> 1955-06-08 (xsd:date)
dbpedia-owl:birthName	<ul style="list-style-type: none"> Timothy John Berners-Lee
dbpedia-owl:birthYear	<ul style="list-style-type: none"> 1955-01-01 (xsd:date)
dbpedia-owl:employer	<ul style="list-style-type: none"> dbpedia:World_Wide_Web_Consortium dbpedia:Massachusetts_Institute_of_Technology dbpedia:University_of_Southampton dbpedia:Plessey
dbpedia-owl:knownFor	<ul style="list-style-type: none"> dbpedia:World_Wide_Web dbpedia:Massachusetts_Institute_of_Technology
dbpedia-owl:occupation	<ul style="list-style-type: none"> dbpedia:Computer_scientist dbpedia:Tim_Berners-Lee__1
dbpedia-owl:parent	<ul style="list-style-type: none"> dbpedia:Mary_Lee_Woods dbpedia:Conway_Berners-Lee
dbpedia-owl:personFunction	<ul style="list-style-type: none"> dbpedia:Tim_Berners-Lee__2
dbpedia-owl:religion	<ul style="list-style-type: none"> dbpedia:Unitarianism
dbpedia-owl:thumbnail	<ul style="list-style-type: none"> http://commons.wikimedia.org/wiki/Special:FilePath/Tim_Berners-Lee_2012.jpg?width=300
dbpedia-owl:title	<ul style="list-style-type: none"> Professor
dbpedia-owl:viafId	<ul style="list-style-type: none"> 85312226
dbpedia-owl:wikiPageExternalLink	<ul style="list-style-type: none"> http://computemagazine.com/man-who-invented-world-wide-web-gives-new-definition/ http://identi.ca/timbl http://info.cern.ch/hypertext/WWW/TheProject.html, http://news.bbc.co.uk/2/hi/technology/4132752.stm http://videolectures.net/iswc06_lee_itbl/ http://www.technologyreview.com/Infotech/13784/ http://www.w3.org/History/19921103-hypertext/hypertext/WWW/TheProject.html http://www.w3.org/People/Berners-Lee/
dbpedia-owl:wikiPageId	<ul style="list-style-type: none"> 30034 (xsd:integer)
dbpedia-owl:wikiPageRevisionID	<ul style="list-style-type: none"> 605407364 (xsd:integer)
dbpprop:almaMater	<ul style="list-style-type: none"> dbpedia:The_Queen's_College,_Oxford
dbpprop:alternativeNames	<ul style="list-style-type: none"> TimBL
dbpprop:awards	<ul style="list-style-type: none"> * OM * KBE * OBE * RDI * FRS * FREng
dbpprop:birthDate	<ul style="list-style-type: none"> 1955-06-08 (xsd:date)
dbpprop:birthName	<ul style="list-style-type: none"> Timothy John Berners-Lee
dbpprop:birthPlace	<ul style="list-style-type: none"> United Kingdom London, England

Definition

- ***“The Semantic Web is not a separate Web***
 - but an extension of the current one,***
- ***In which information is given***
 - well-defined meaning,***
 - better enabling computers***
 - and people to work in cooperation.”***

Tim Berners-Lee

Comparison of WWW and SW

FEATURE	WWW	SEMANTIC WEB
Fundamental component	Unstructured content	Formal statements
Primary audience	Humans	Applications
Links	Indicate location	Indicate location and meaning
Primary vocabulary	Formatting instructions	Semantics and logic
Logic	Informal/nonstandard	Description logic

Hebeler, John, et al. Semantic web programming. John Wiley & Sons, 2011.



Semantics as Metadata

ReBIL: Relating Biological Information
through Literature

Francisco José Moreira Couto

DOUTORAMENTO EM INFORMÁTICA
ESPECIALIDADE BIOINFORMÁTICA

2006

Tese orientada pelo Prof. Doutor Mário Jorge Costa Gaspar da Silva
e pelo Prof. Doutor Pedro Maldonado Coutinho

Title: ReBIL: Relating Biological
Information through Literature

Authors: [Couto, Francisco M.](#)

Advisors : [Mário J. Silva](#)

Keywords: Bioinformatics
Text mining
Data mining
Gene and protein automatic
annotation

Issue Date: May-2006

Publisher: Department of Informatics,
University of Lisbon

Series/Report no.: di-fcul-tr-06-6

Abstract: Bioinformatics ...

DOI: <http://hdl.handle.net/10455/3127>

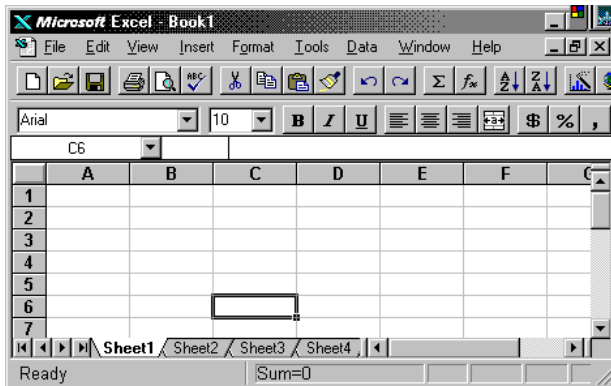
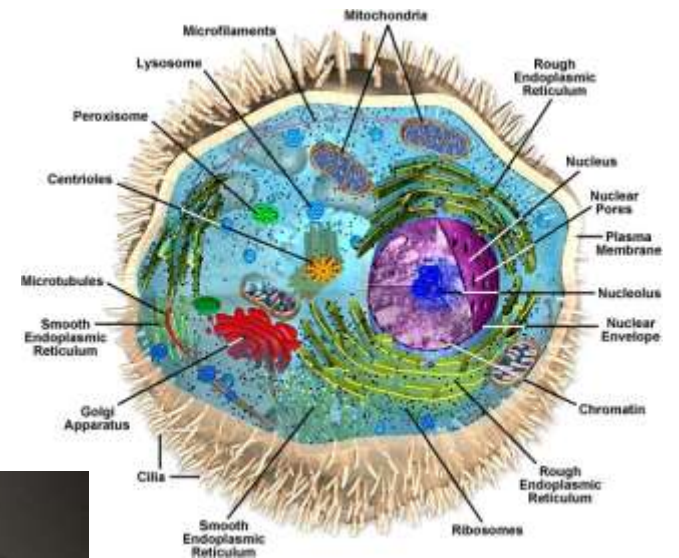
Appears in Collections: [Doctoral Theses](#)

Ambiguity

```
<meta name="keywords"  
  content="building, ontologies" />
```

- Does building mean constructing an ontology?
- Or ontologies that focus on constructing buildings?

Cell



Different Perceptions of Reality



Metadata

http://en.wikipedia.org/wiki/Sintra_Collar

- **Title:** Sintra Collar
- **Local:** Portugal
- **Date:** 1250BC-800BC
- **Type:** gold



```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">

  <rdf:Description rdf:about="http://en.wikipedia.org/wiki/Sintra_Collar">

    <dc:description>
      Gold collar. It was made from three circular sectioned and tapering gold
      bars that are fused at the ends forming a penannular neck-ring.
    </dc:description>

    <dc:date>1250BC-800BC (circa)</dc:date>

    <dc:location>
      Sintra, Portugal
      http://yboss.yahooapis.com/geo/placementfinder?woeid=748874
    </dc:location>

    <dc:type>
      Gold
      http://purl.obolibrary.org/obo/CHEBI_30050
    </dc:type>

  </rdf:Description>

</rdf:RDF>
```

Ontologies

- Field of Knowledge Representation
 - to denote computational systems for describing specific domains of reality

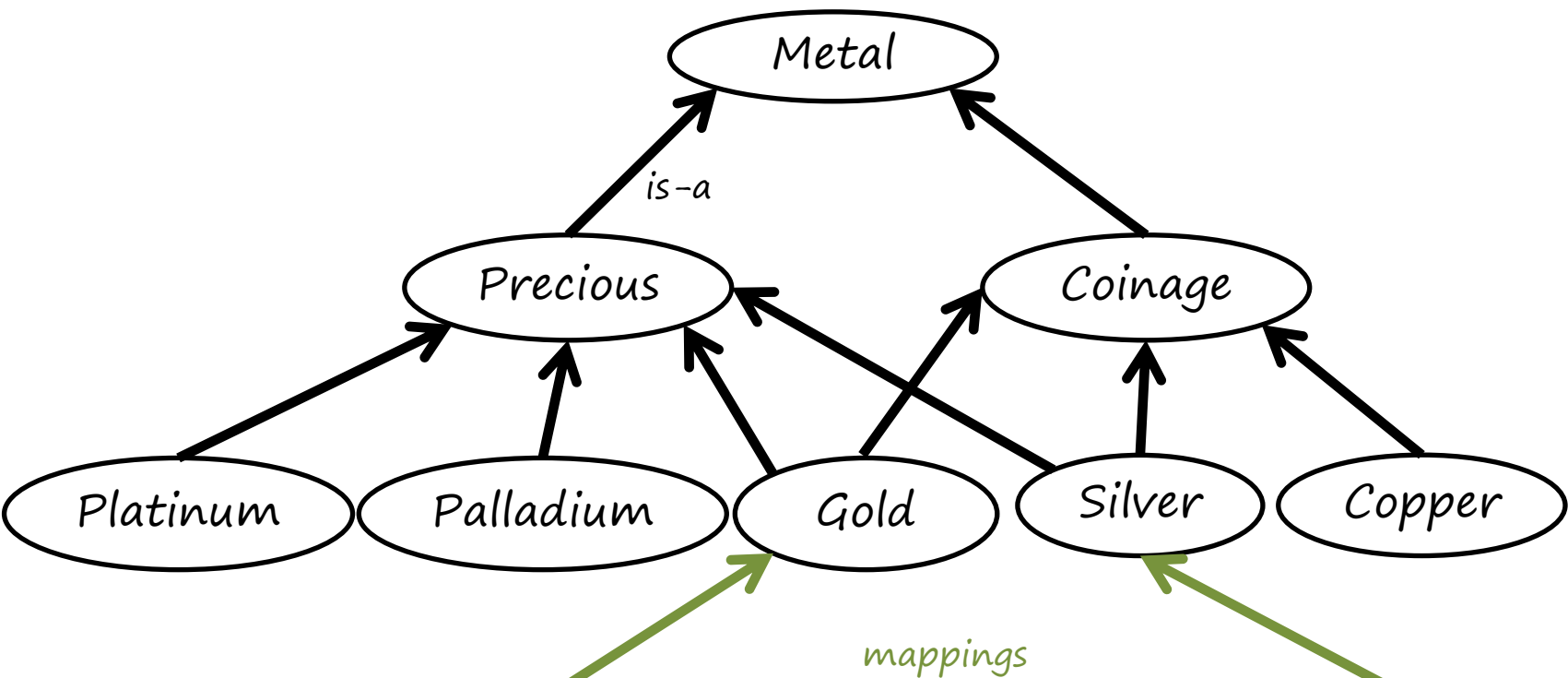
*“An ontology is a specification of a conceptualization.”
by Tom Gruber, 1993*
- Semantic Web applications
 - use many ontologies
 - each chosen for a required information area

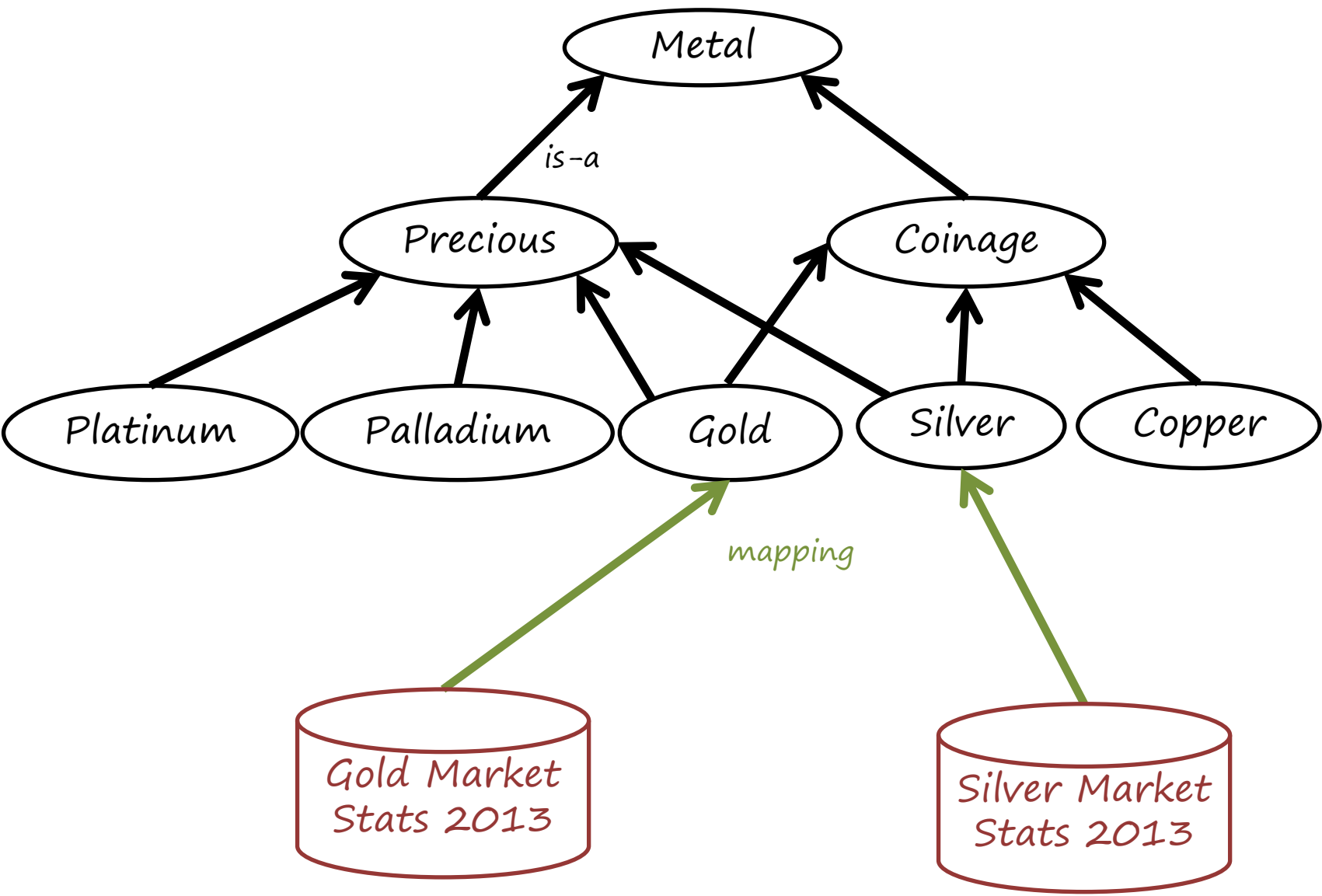
Metadata

http://en.wikipedia.org/wiki/American_Silver_Eagle

- **Title:** American Silver Eagle
- **Local:** USA
- **Date:** 1916
- **Type:** silver







METADATA CREATION

Human Factor

- authors cannot recover the data
 - associated with their own published works

<http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.1003542>
- *“you have no idea the number of excuses people come up with to hang onto their data and not give it to you, even though you've paid for it as a taxpayer”*

Tim Berners-Lee, TED talk, 2009

http://www.ted.com/talks/tim_berners_lee_on_the_next_web

Data-sharing policies

- adherence is inconsistent and scarce
- needs motivation to do it correctly
 - enforcing does not work
- *“to encourage data sharing, systematic reward and recognition mechanisms are necessary”.*

Roadmap of the European Strategy Forum on Research Infrastructures (ESFRI)

<https://zenodo.org/record/8304>

Reward and recognition mechanisms

- motivate the community
 - To become strongly committed in sharing data
- we must formally define:
 1. what needs to be rewarded and recognized;
 2. and measure its value in quantitative and objective way.

Proper data integration and sharing

- *“is time-consuming to do properly, the reward systems aren't there and neither is the stick”*

Steven Wiley in Nature, 2011

<http://www.nature.com/news/2011/110914/full/news.2011.536.html>

- Only careful and clear semantic characterizations
 - in the form of metadata
 - enable us to find the raw data
 - and retrieve it and explored it

Assumption

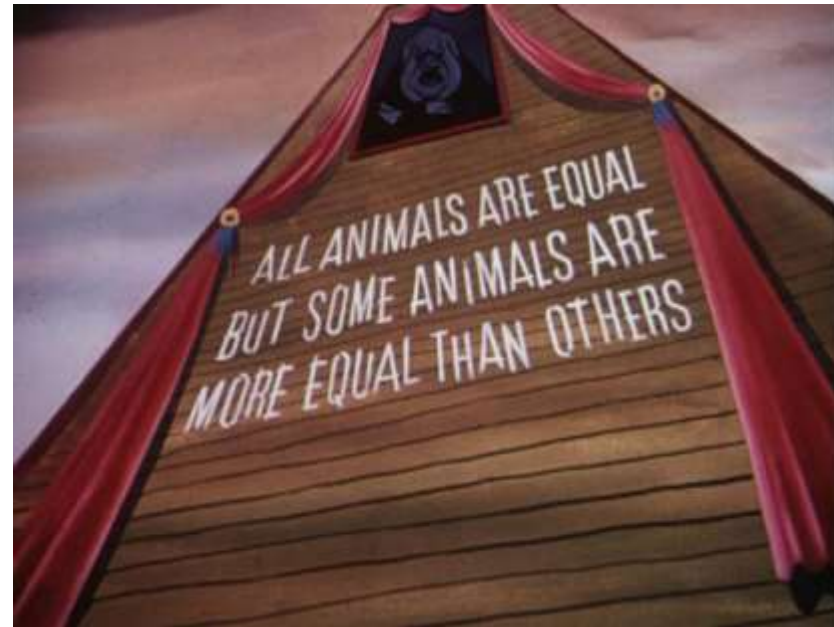
- rewarding and recognizing
 - metadata sharing and integration
 - on the semantic web
 - using standard and controlled vocabularies
- promotes and intensifies
 - scientific collaboration and progress

Knowledge rating

- Metadata as a set of links, where:

*all the links are equal,
but some links
are more equal
than others*

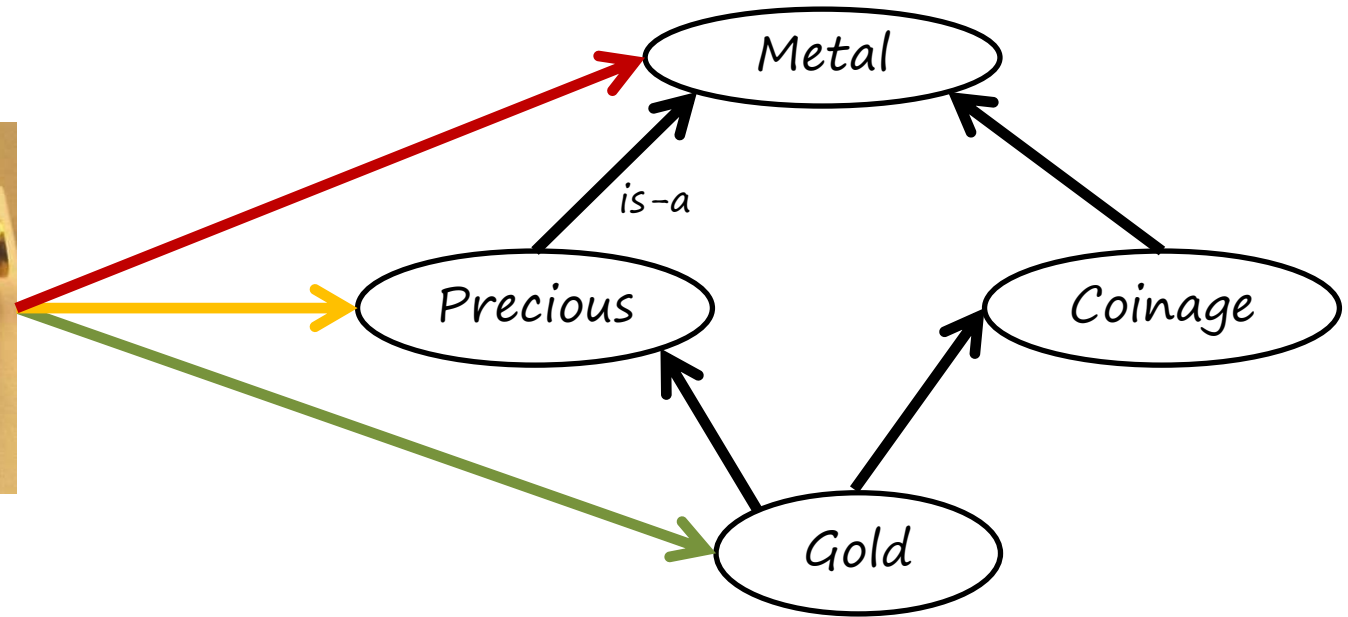
(adaption of George Orwell's quote of Animal Farm)



Hypothesis

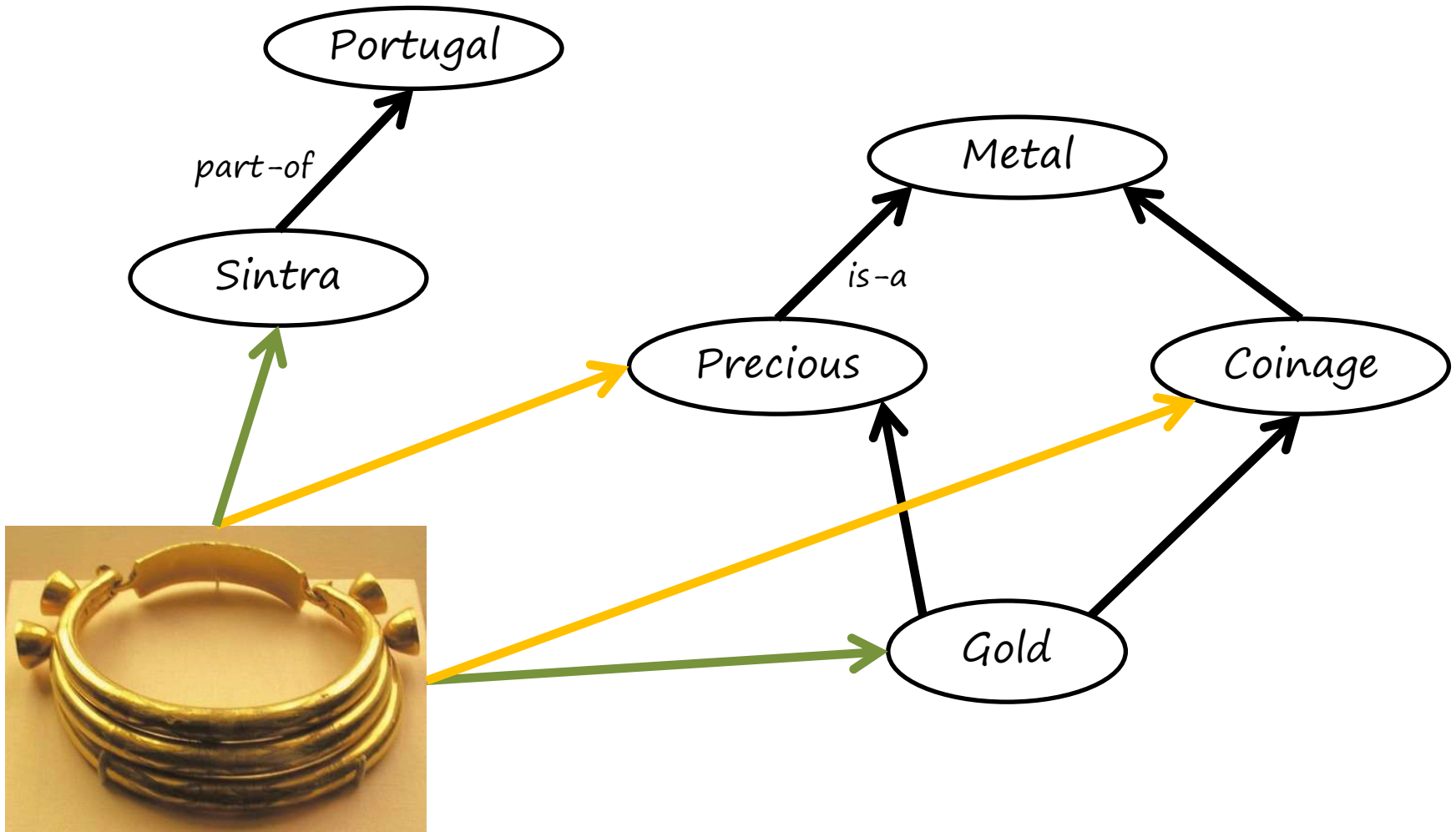
*The **metadata integration and sharing** value of a dataset, dubbed as **knowledge rating**, is proportional to the **specificity** and **distinctiveness** of its mappings to ontology concepts in relation to all the others datasets in the **Linked Data Cloud**.*

Specificity Information Content

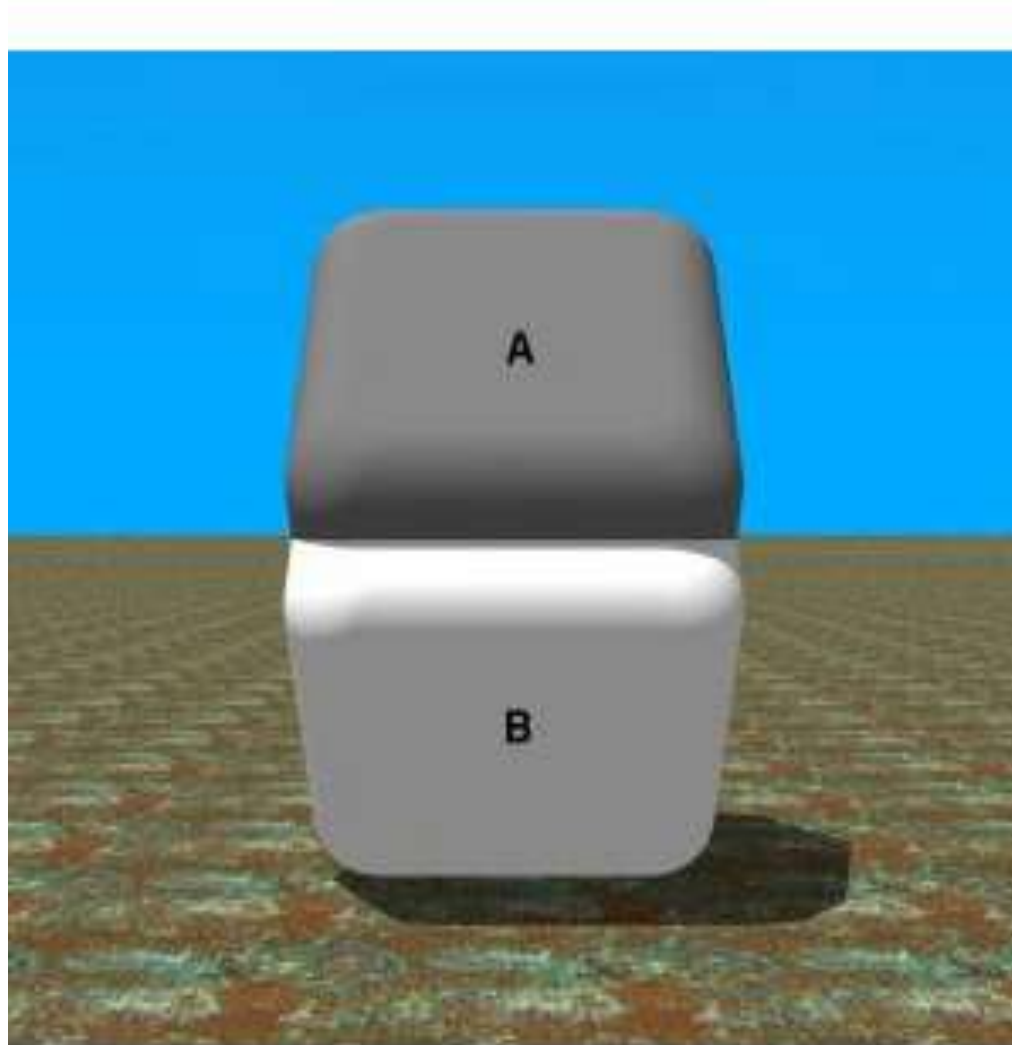


Distinctiveness

Semantic Similarity



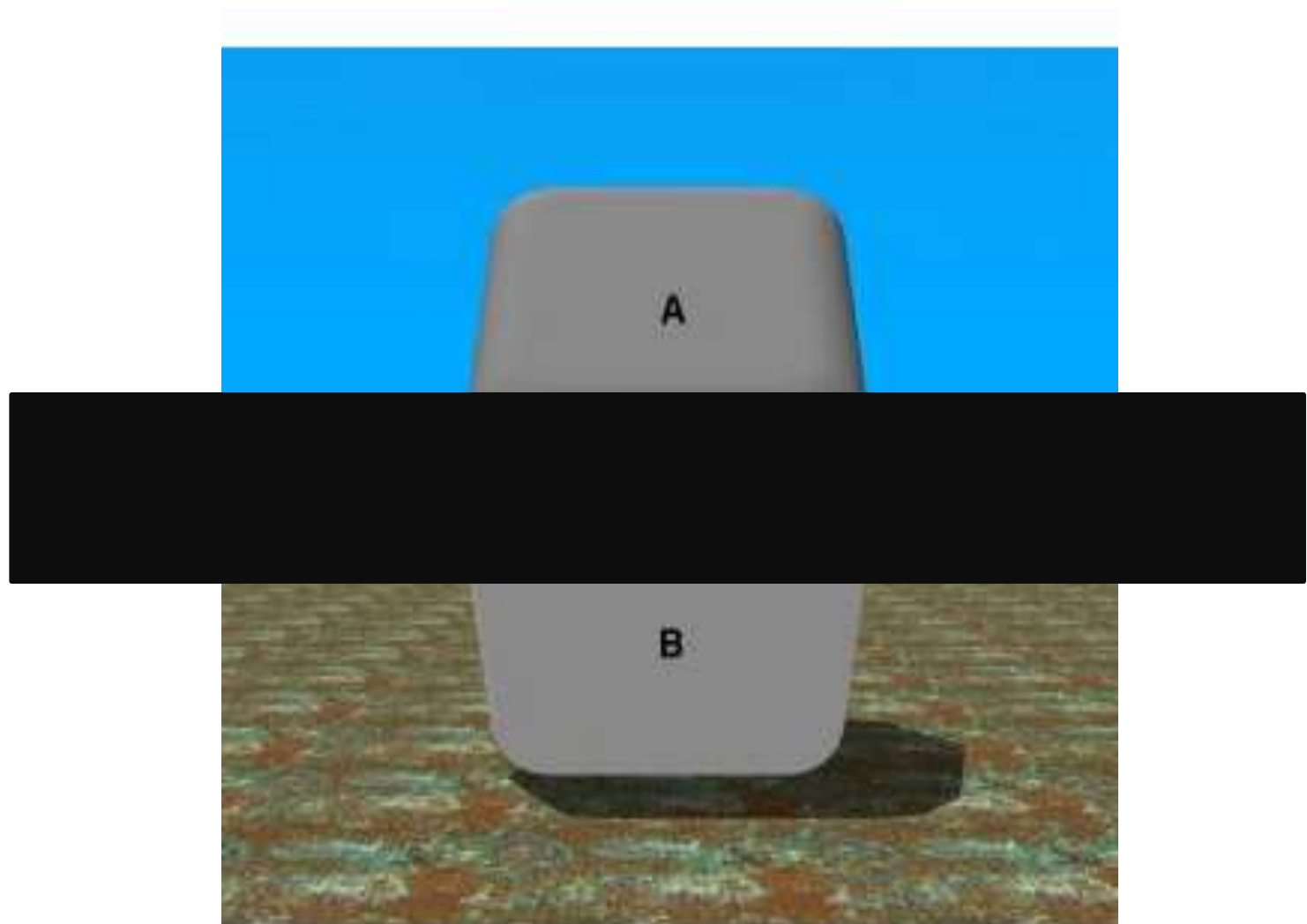
Are we good at measuring similarity?



Source:

<http://www.grandparents.com/food-and-leisure/did-you-know/optical-illusion-pictures>

Are we good at measuring similarity?



All similar but different brands



INFORMATION CONTENT AND SIMILARITY

Semantic Similarity

“A semantic similarity measure is a **function** that, given two ontology terms or two sets of terms annotating two entities, returns a numerical value reflecting the closeness in **meaning** between them.”

C. Pesquita, D. Faria, A. Falcão, P. Lord, and F. Couto, Semantic similarity in biomedical ontologies, PLoS Computational Biology, vol. 5, no. 7 (e1000443)

F. Couto and H. Pinto, The next generation of similarity measures that fully explore the semantics in biomedical ontologies, Journal of Bioinformatics and Computational Biology, vol. 11, no. 1371001, 2013

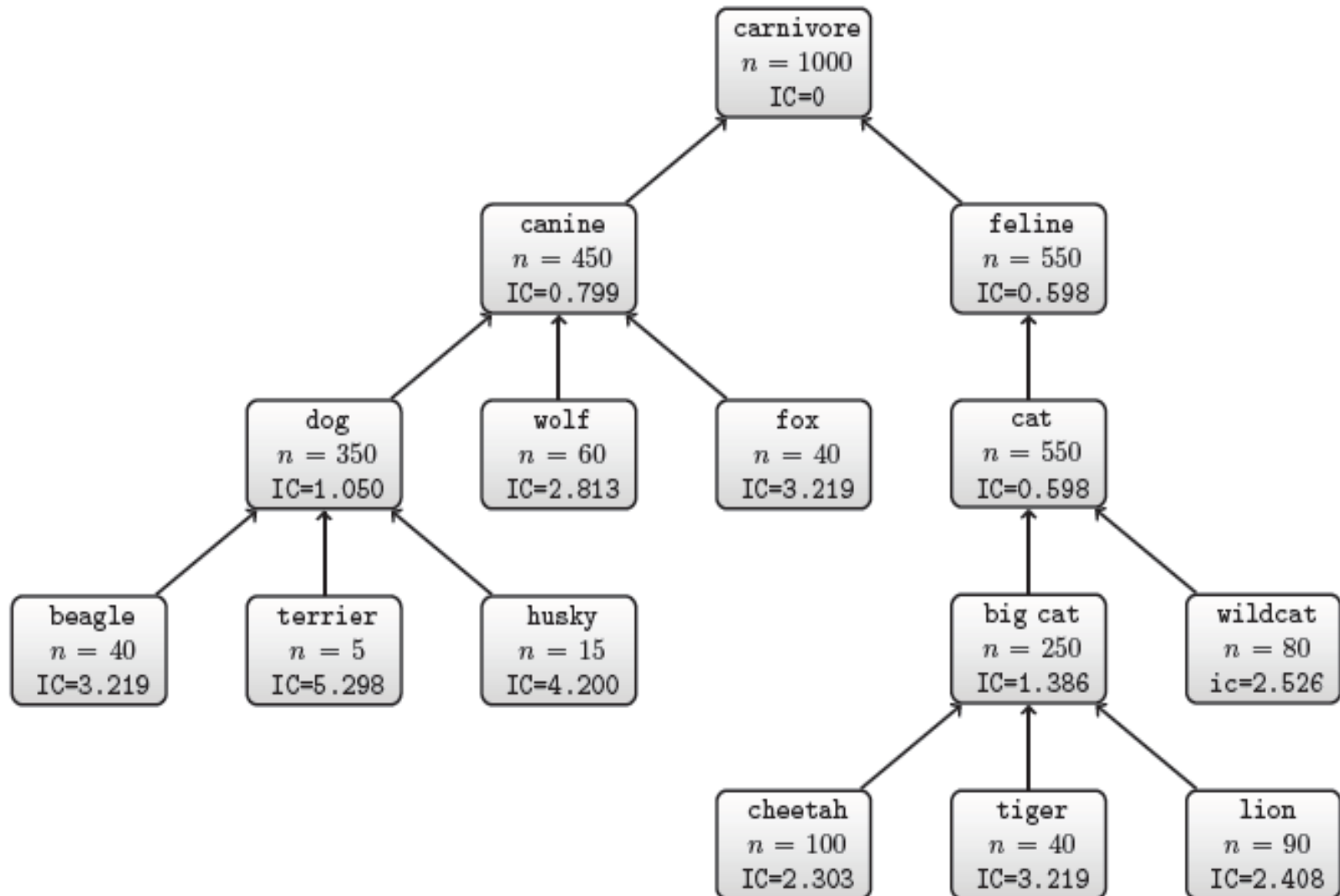
Semantic similarity

- The more information two terms share in common, the more similar they are.
- The information shared between two terms is indicated by the information content of their most informative common ancestor: MICA

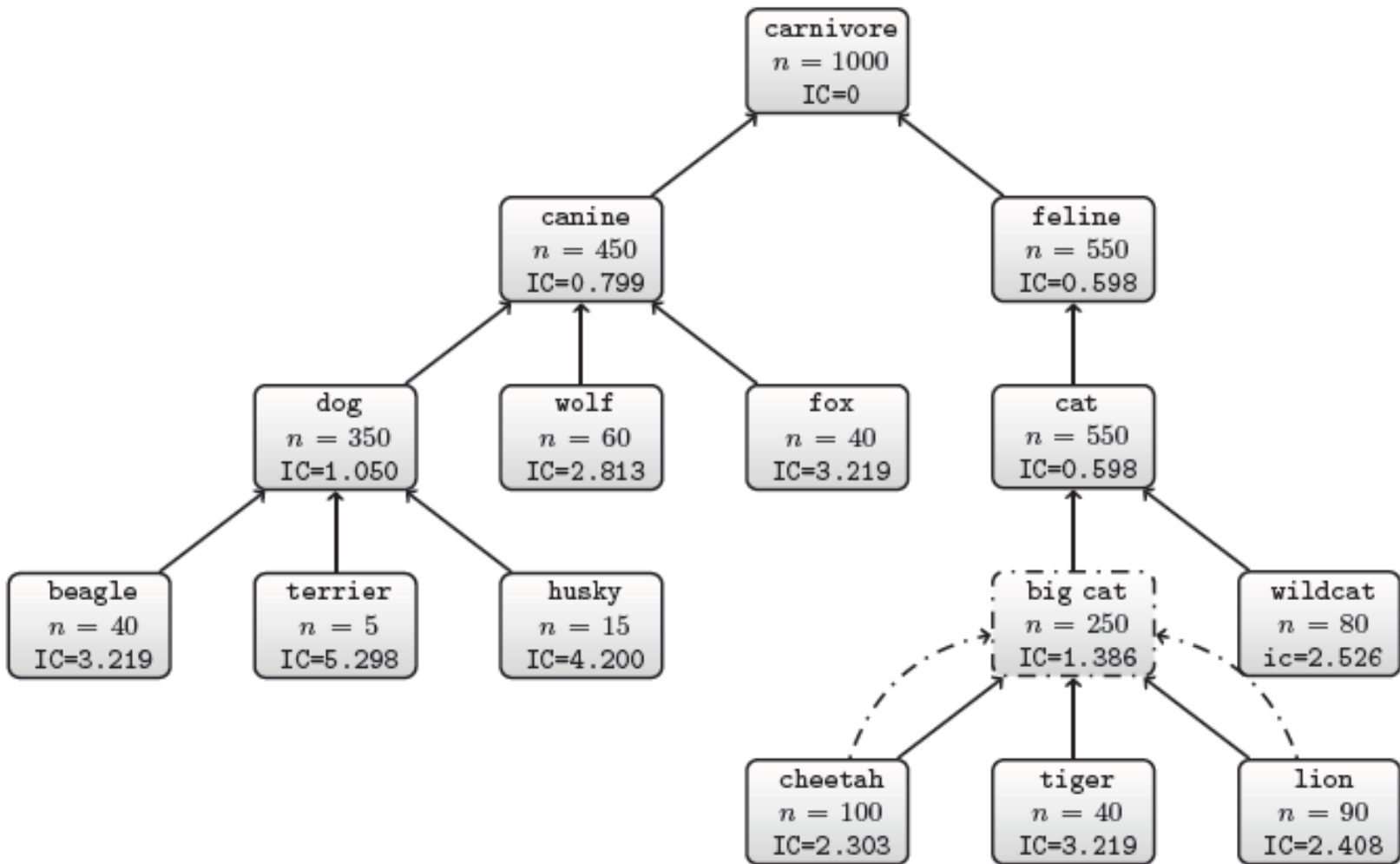
$$\text{sim}(t_1, t_2) = \max_{t \in \text{Anc}(t_1) \cap \text{Anc}(t_2)} IC(t) = -\log(p(t_{MICA}))$$

IC Example

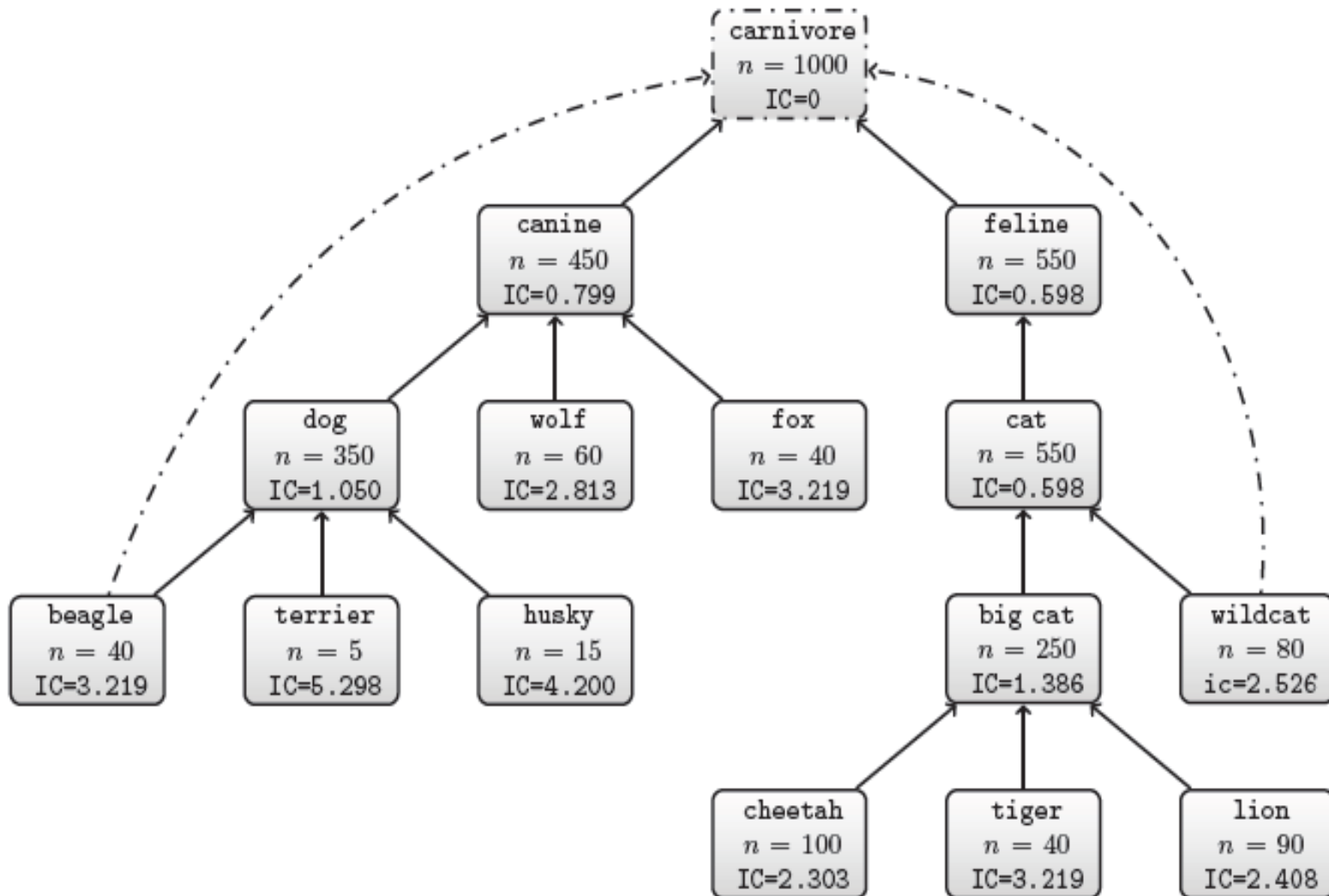
Introduction to Bio-Ontologies by P. Robinson and S. Bauer



$$\text{Sim}(\text{cheetah}, \text{lion}) = \text{IC}(\text{big cat}) = 1.386$$



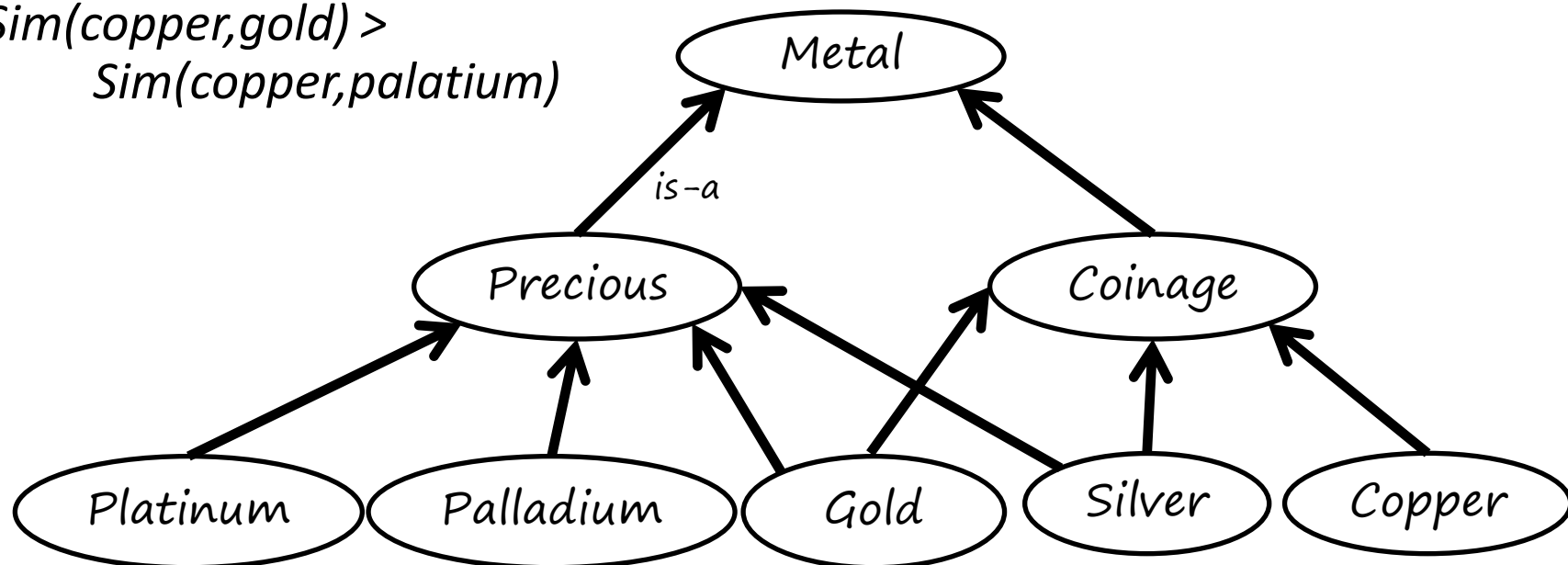
$$\text{sim}(\text{wildcat}, \text{beagle}) = \text{IC}(\text{carnivore}) = 0$$



Example

F. Couto and M. Silva, Disjunctive shared information between ontology concepts: application to Gene Ontology, Journal of Biomedical Semantics,, 2011

- $IC(copper) > IC(coinage) > IC(metal)$
- $Sim(copper, gold) \sim IC(coinage)$
- $Sim(copper, palatium) \sim IC(metal)$
- $Sim(copper, gold) > Sim(copper, palatium)$



Rating, recognizing and rewarding metadata integration and sharing on the semantic web, in 3rd International Workshop on Methods for Establishing Trust of (Open) Data (METHOD), 2014

KNOWLEDGE COIN



Cryptocurrency Components

- **Wallet**, define the amount of coins assigned to an individual, i.e. the balance;
- **Mining**, the process of supplying coins to some wallets, i.e. money creation;
- and **Transaction**, the process of transferring coins between two wallets, i.e. money exchange
- KC is not a Cryptocurrency

KC Mining



- When new knowledge is created
 - a new scientific article
 - references a supporting dataset
 - properly integrated in the Linked Data Cloud
 - level of integration measured by **knowledge ratings**
- Recognition statements
 - specify a different distribution of the KCs
 - Not uniformly distributed between author wallets
 - that may even include other wallets.

KC Inflation

- #KC in circulation will keep growing
 - with the growth of new scientific articles
 - may trigger some distrust on its value
- But, scientific articles are peer reviewed
 - So their number is always limited
 - and adhesion to real scientific advances guaranteed

KC Wallet



- Different from conventional cryptocurrency wallets
- the KCs will use a public identifier of its owner
 - for example using the ORCID
 - provides a persistent digital identifier unique to each researcher <http://orcid.org/>

New Knowledge



BMC Cancer | Full text | E: x
www.biomedcentral.com/1471-2407/14/518

Research article **Highly accessed** **Open Access**

Extracellular matrix signatures of human primary metastatic colon cancers and their metastases to liver

Alexandra Naba^{1,2*}, Karl R Clauser³, Charles A Whittaker⁴, Steven A Carr³, Kenneth K Tanabe⁵ and Richard O Hynes^{1,2*}

* Corresponding authors: Alexandra Naba anaba@mit.edu - Richard O Hynes rohynes@mit.edu **Author Affiliations**

- 1 David H. Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
- 2 Howard Hughes Medical Institute, Massachusetts Institute of Technology, 02139 Cambridge, MA, USA
- 3 Proteomics Platform, Broad Institute of MIT and Harvard, 02142 Cambridge, MA, USA
- 4 David H. Koch Institute for Integrative Cancer Research - Barbara K. Ostrom Bioinformatics and Computing facility at the Swanson Biotechnology Center, 02139 Cambridge, MA, USA
- 5 Division of Surgical Oncology, Massachusetts General Hospital Cancer Center, Boston 02114, MA, USA

For all author emails, please [log on](#).

BMC Cancer 2014, **14**:518 doi:10.1186/1471-2407-14-518

and survival.

Availability of supporting data

In addition to the supporting data included as additional files, the raw GSEA data may be downloaded from: http://rowley.mit.edu/Hynes/Naba_GSEA_ColonCancer/ [website](#). The raw mass spectrometry data accompanying this publication have been deposited in the public proteomics repository MassIVE (<http://massive.ucsd.edu> [website](#)) using the identifier: MSV000078555. The data should be accessible at <ftp://MSV000078555:a@massive.ucsd.edu> [website](#).

Abbreviations

ECM: Extracellular matrix; LC-MS/MS: Liquid chromatography and tandem mass spectrometry; OGE: Off-gel electrophoresis.w.

Competing interests

The authors declare that they have no competing interests.

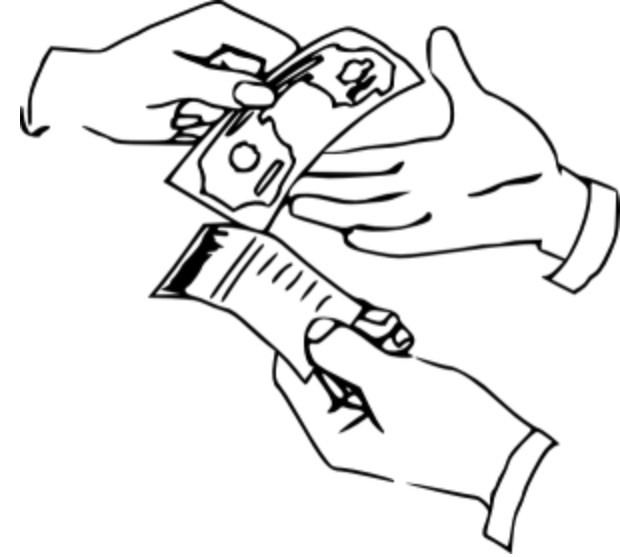
Authors' contributions

Conception and design of the experiments: AN, KRC, SAC, KKT, ROH. Development of methodology: AN, KRC. Acquisition of data: AN, KRC. Analysis and interpretation of data: AN, KRC, CAW, ROH. Writing of the manuscript: AN, KRC, CAW, SAC, KKT, ROH. Study approval: AN, KRC, CAW, ROH. All authors have read and approved the submitted manuscript.

Acknowledgements

The authors wish to thank Dr. Jia Wang from the Barbara K. Ostrom Bioinformatics

KC Transaction



- every KCs transaction
 - will be public, including:
 - the identification of researchers,
 - the amount of KCs exchanged,
 - and the references to the datasets involved.
- provide a public recognition of each individual contribution
 - to the intensification of data integration
 - and sharing using semantic web technologies

Scenario 1

- An exchange of a dataset
 - from the data provider to the data consumer
 - may include recognition statements
 - May incentivize the data provider
 - to properly characterize the dataset
 - more KCs when the data consumer publishes the work as a scientific article

Scenario 2

- The enrichment of a public dataset
 - done by a data curator
 - to satisfy a request from any researcher,
 - inclusively the data owner
- knowledge ratings can be used to
 - estimate the fair amount of KCs to be exchanged

Scenario 3

- a journal may offer publication fees reductions
 - in exchange of KCs
 - used to better characterize some datasets
 - supporting some previously published articles
 - improve the visibility of its articles
 - and in the end increase the journal impact factor

Scenario 4

- Data providers may include in their public funding proposals
 - the amount of KCs recently acquired
 - to show the importance and impact
 - of their work to reviewers

Issues to be studied

- knowledge ratings implementation
 - their validation, aggregation, performance, exceptions, and extension to any mappings besides the ontological ones;
- potential abuses
 - such as the creation of spam mappings and other security threats;
- central trusted authority for the KC ?!?
 - or the peer-to-peer mechanisms used by bitcoin;
- Study use case scenarios for the KC
 - exchange of datasets and
 - their characterization based on KC transactions

And to implement
KnowledgeCoins
we need RealCoins...



<http://metarate.fc.ul.pt>

PhD scholarship opportunity
@Phd Programme Biosys at ULisboa



Financed by the *Portuguese Fundação para a Ciência e Tecnologia* through the Multiannual Funding Programme and the project SOMER (PTDC/EIA-EIA/119119/2010); and by the European Commission through the BiobankCloud project under the Seventh Framework Programme (grant #317871).