

Lavorare con i Linked Data

Nuovi modelli di organizzazione della conoscenza in ambiente digitale

Giovedì 16 giugno 2016, Aula D5

Campus Luigi Einaudi, Lungo Dora Siena, 100 A, Torino

I contesti dei Linked Data

Maurizio Vivarelli

Dipartimento di Studi storici, Università di Torino

maurizio.vivarelli@unito.it

Un seminario
promosso da



Sezione Piemonte



Sezione Piemonte



Sezione Piemonte



Sezione Piemonte



Punti di vista

Obiettivo di questo intervento è **discutere il «contesto» di produzione e comunicazione dell'informazione digitale** in cui i dati (inclusi quelli «linked») si situano.

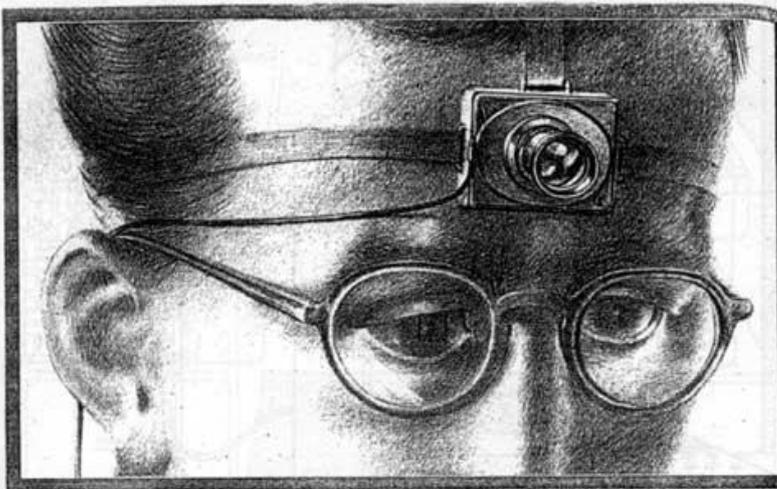
Seguendo la suggestione visiva di *Interstellar* possiamo pensare l'accesso all'informazione secondo **due prospettive**, fondate la prima su un approccio «razionale», e la seconda sulla «intelligenza» distribuita nel Web.

Per comprendere, od intuire, alcune di queste dinamiche è indispensabile **allargare il punto di vista dell'analisi**, rivolgendo le domande di Vannevar Bush su «come possiamo pensare» e «come potremmo pensare» sia a noi che alle macchine.

Due immagini dal film
Interstellar, regia di
Cristopher Nolan, 2014



As We May Think, As We Will Think



A SCIENTIST OF THE FUTURE RECORDS EXPERIMENTS WITH A TINY CAMERA FITTED WITH UNIVERSAL-FOCUS LENS. THE SMALL SQUARES IN THE EYEGLASS AT THE LEFT SIGHTS THE USE

AS WE MAY THINK

A TOP U. S. SCIENTIST FORESEES A POSSIBLE FUTURE WORLD IN WHICH MAN-MADE MACHINES WILL START TO THINK

by VANNEVAR BUSH

DIRECTOR OF THE OFFICE OF SCIENTIFIC RESEARCH AND DEVELOPMENT
Condensed from the Atlantic Monthly, July 1945

This has not been a scientists' war; it has been a war in which all have had a part. The scientists, burying their old professional competition in the demand of a common cause, have shared greatly and learned much. It has been exhilarating to work in effective partnership. What are the scientists to do next?

For the biologists, and particularly for the medical scientists, there can be little indication, for their war work has hardly required them to leave the old paths. Many indeed have been able to carry on their war research in their familiar peacetime laboratories. Their objectives remain much the same.

It is the physicists who have been thrown most violently off stride, who have left academic pursuits for the making of strange destructive gadgets, who have had to devise new methods for their unanticipated assignments. They have done their part on the devices that made it possible to turn back the enemy. They have worked in combined effort with the physicists of our allies. They have felt within themselves the stir of achievement. They have been part of a great team. Now one asks where they will find objectives worthy of their best.

There is a growing mountain of research. But there is increased evidence that we are being bogged down today as specialization extends. The investigator is staggered by the findings and conclusions of thousands of other workers—conclusions which he cannot find time to grasp, much less to remember, as they appear. Yet specialization becomes increasingly necessary for prog-

ress, and the effort to bridge between disciplines is correspondingly superficial.

Professionally our methods of transmitting and reviewing the results of research are generations old and by now are totally inadequate for their purpose. If the aggregate time spent in writing scholarly works and in reading them could be evaluated, the ratio between these amounts of time will be startling. Those who conscientiously attempt to keep abreast of current thought, even in restricted fields, by close and continuous reading will shy away from an examination calculated to show how much of the previous month's efforts could be produced on call.

Mendel's concept of the laws of genetics was lost to the world for a generation because his publication did not reach the few who were capable of grasping and extending it. This sort of catastrophe is undoubtedly repeated all about us as truly significant attainments become lost in the sea of the inconsequential.

Publication has been extended far beyond our present ability to make use of the record. The summation of human experience is being expanded at a prodigious rate, and the means we use for absorbing it—the most quest since the momentarily important item is the same as was used the days of square-rigged ships.

But there are signs of a change as new and powerful instrumentalities come into use. Photocells capable of seeing things in a physical sense, x-ray photography which can record what is seen or even what is felt, thermionic tubes capable of controlling potent forces under the guidance

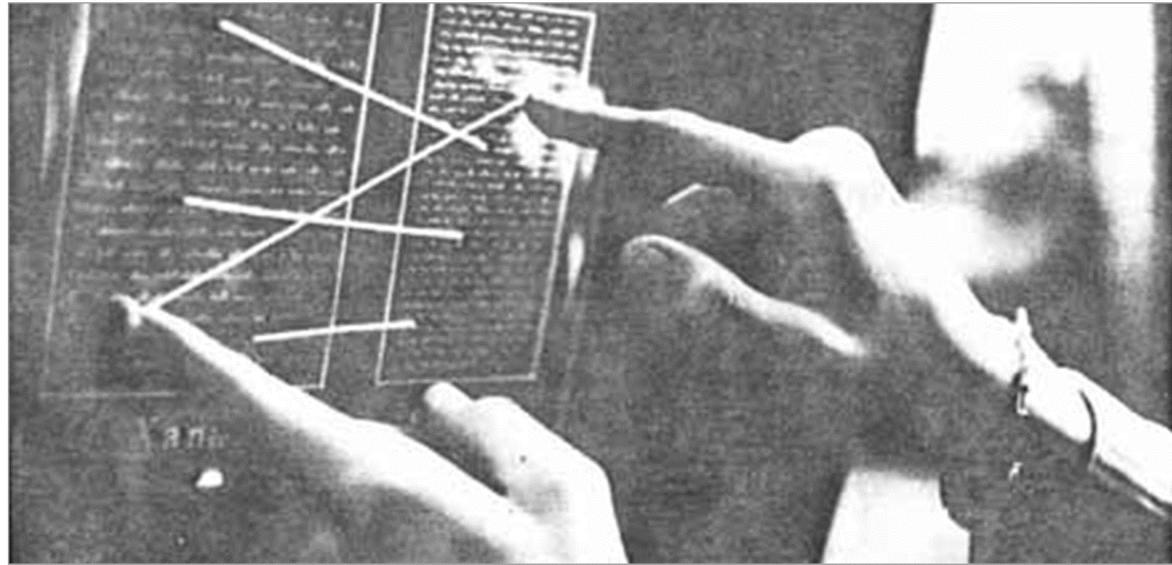
112

「LIFE」1945年9月10日号に特集された「AS WE MAY THINK」のページ

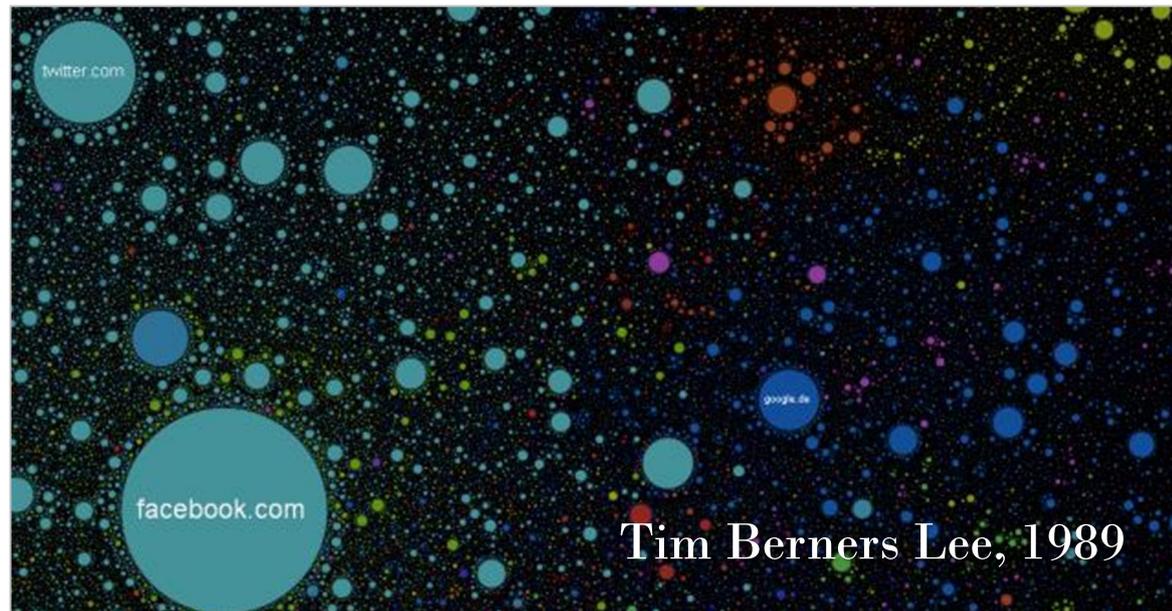
「LIFE」1945年9月10日号より引用

「AS WE MAY THINK」From the Atlantic Monthly, July 1945

Vannevar Bush, 1945



Ted Nelson, 1965

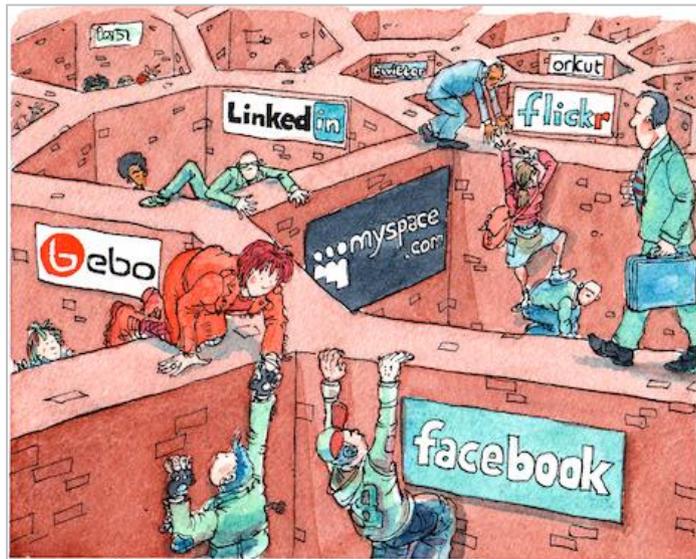
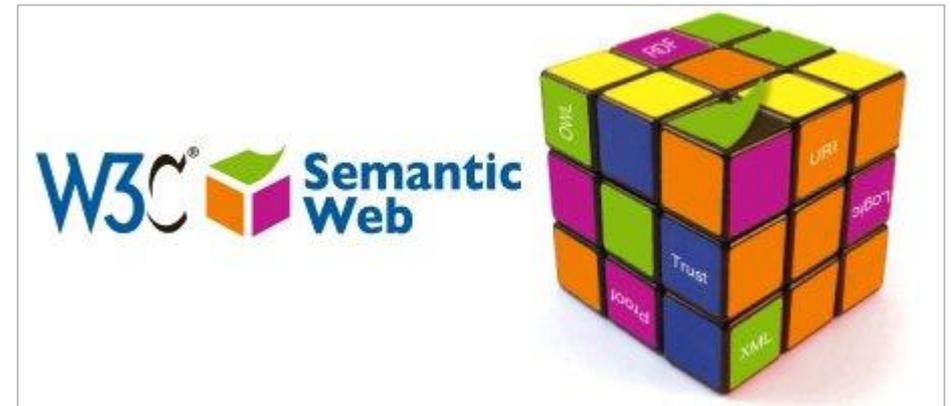


Tim Berners Lee, 1989

Web semantico e Web sociale

Web semantico

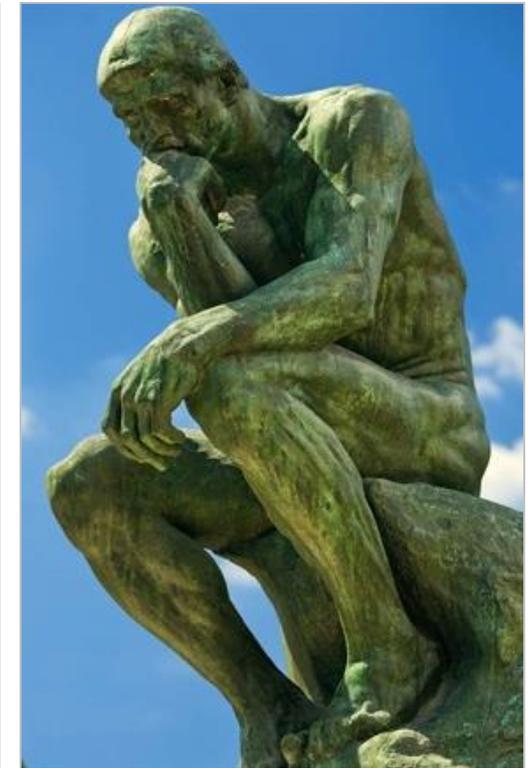
Tim Berners-Lee, James Hendler,
Ora Lassila, *The Semantic Web*,
«Scientific American», 2001



Web sociale

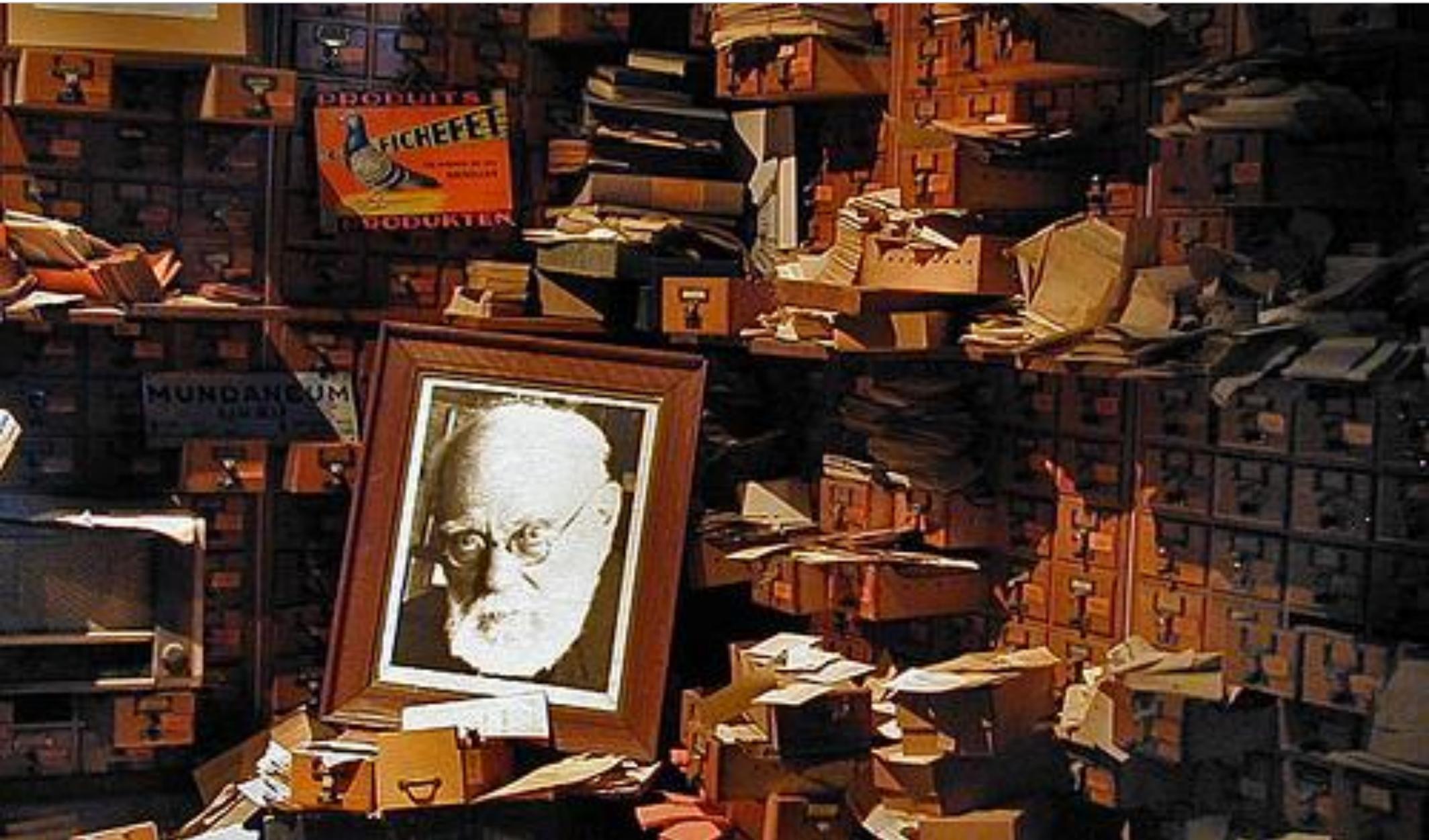
Howard Rheingold, nel sito web *Electric Minds*, citato da «Time»
nel 1996

L'informazione nei *Big Data*



- Il 90% dell'informazione digitale esistente è stata creata negli ultimi due anni.
- La quantità di tutti i dati esistenti è riprodotta ogni 48 ore.
- Ogni giorno vengono prodotti 2.5 quintilioni (10^{30}) di dati (un nuovo Google ogni quattro giorni).
- Il traffico dei dispositivi mobili nel 2013 è stato di 18 volte maggiore dell'intero traffico di Internet nel 2000.

Ordine e disordine



Elaborazione di una immagine di Paul Otlet

Ordine e disordine

«l'ordinamento dei documenti, in quanto supporti fisici indivisibili di registrazioni di simboli, può avvenire soltanto secondo uno dei caratteri di riconoscimento; così che **un ordine realizzato di documenti esclude necessariamente tutti gli altri**. [...] Per ovviare a questa limitazione sono stati introdotti i cataloghi (sotto forma di schede mobili, riunite o registrate in volume), cioè degli insiemi ordinati di notizie che si riferiscono, ciascuna, a uno dei connotati distintivi dei documenti, vuoi in quanto formula prestabilita, vuoi in quanto posizione o definizione di uno spazio semantico».

Alfredo Serrai, *Biblioteconomia come scienza. Introduzione ai problemi e alla metodologia*, Firenze, Olschki, 1973.

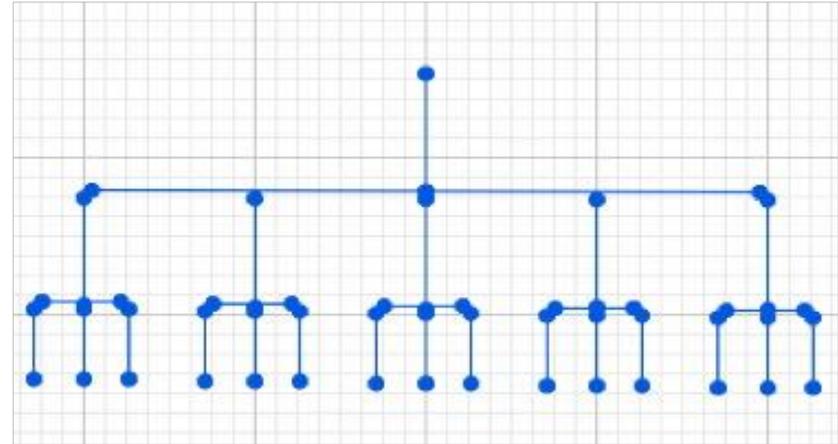
«Codeste **ambiguità, ridondanze e deficienze** ricordano quelle che il dottor Franz Kuhn attribuisce a un'enciclopedia cinese che s'intitola *Emporio celeste di conoscimenti benevoli*. Nelle sue remote pagine è scritto che gli animali si dividono in (a) appartenenti all'Imperatore, (b) imbalsamati, (c) ammaestrati, (d) lattonzoli, (e) sirene, (f) favolosi, (g) cani randagi, (h) inclusi in questa classificazione, (i) che s'agitano come pazzi, (j) innumerevoli, (k) disegnati con un pennello finissimo di pelo di cammello, (l) eccetera, (m) che hanno rotto il vaso, (n) che da lontano sembrano mosche».

Jorge Luis Borges, *L'idioma analitico di John Wilkins*, in *Altre inquisizioni*, Milano, Feltrinelli, 1973.

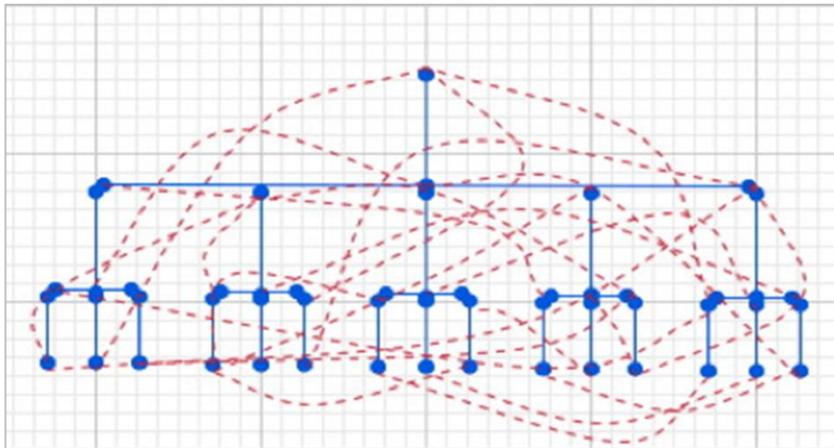
Ordine e disordine

Dall'albero alla rete

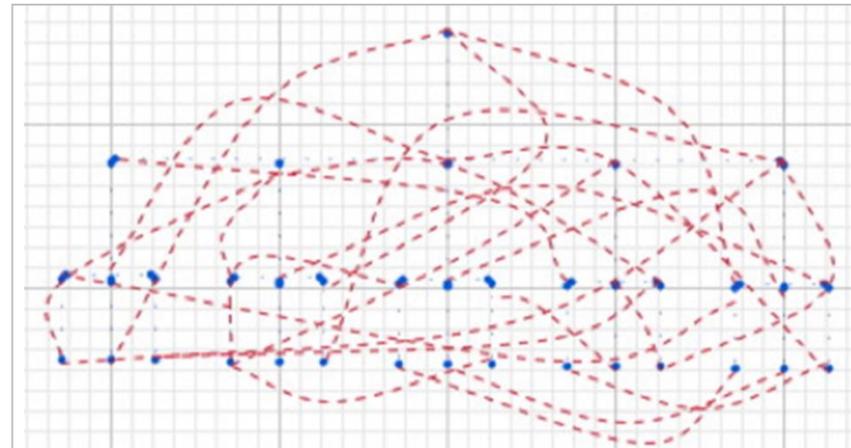
Clay Shirky, *Ontology is Overrated: Categories, Links, and Tags*,
http://www.shirky.com/writings/ontology_overrated.html?goback=.gde_1838701member_179729766



Modello gerarchico



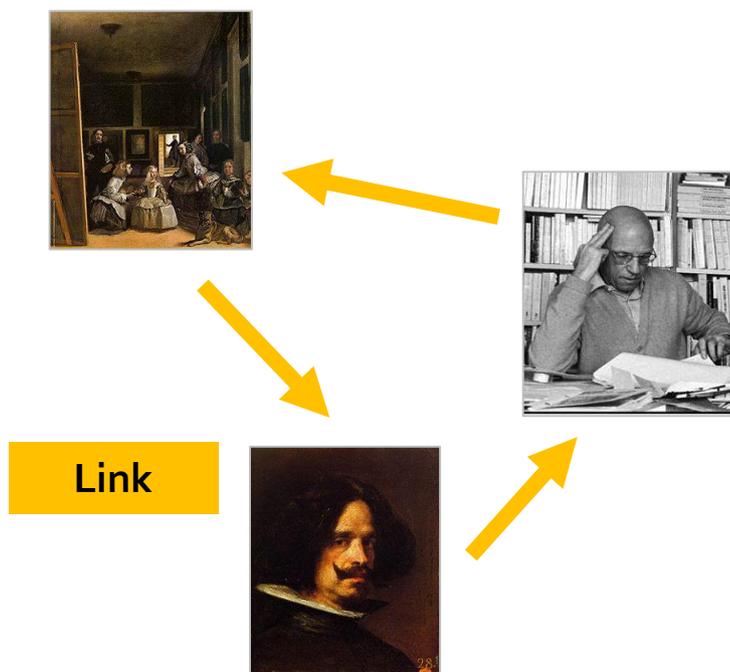
Modello ibrido



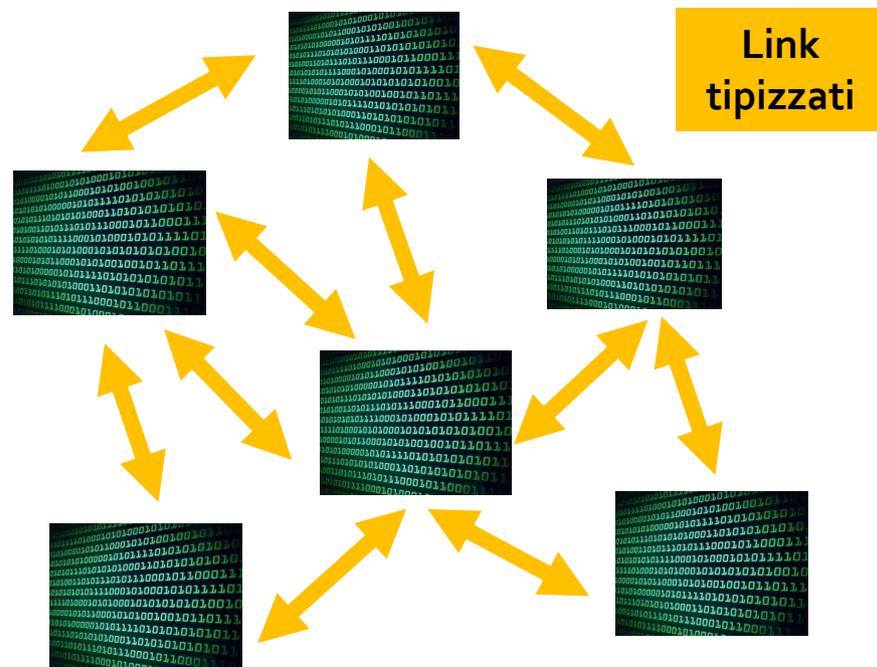
Modello reticolare

Verso il Web dei dati

Web dei documenti

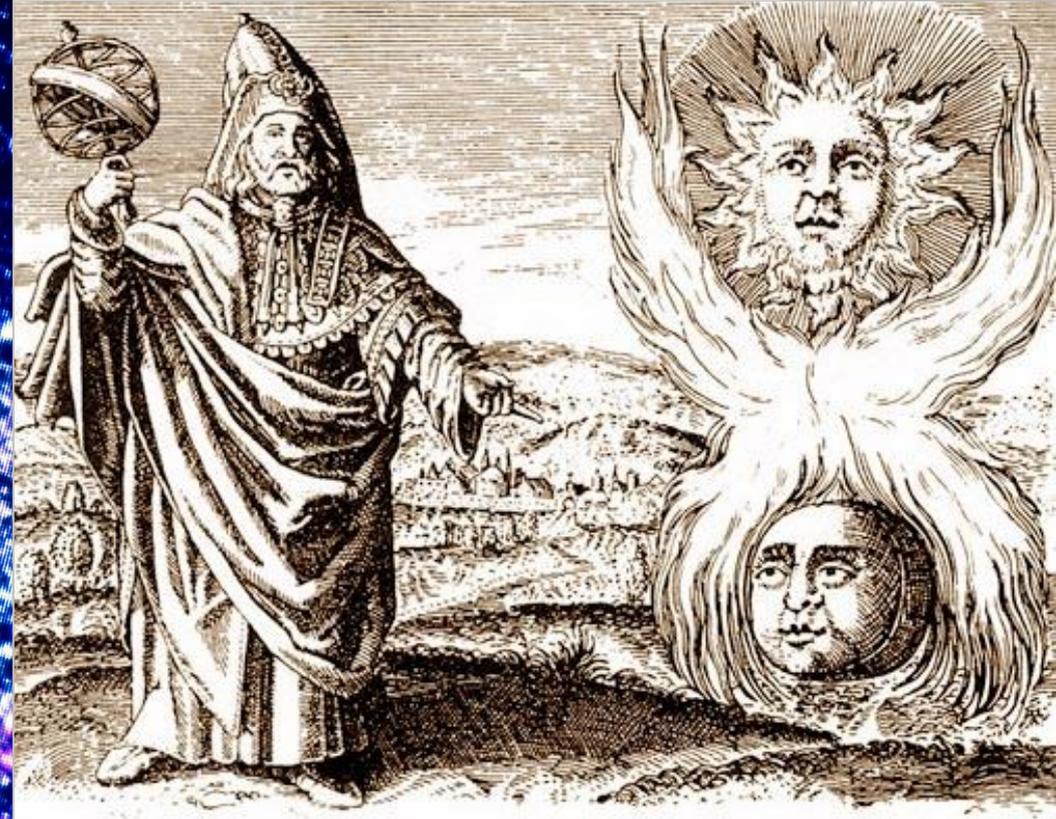


Web dei dati



Web dei documenti	Web dei dati
Risorse informative	Entità
Link tra documenti	Link tra entità
Dati non strutturati	Dati strutturati
Semantica implicita	Semantica esplicita
Destinato agli umani	Destinato alle macchine ed agli umani

Le reti tra «arte» e «scienza»



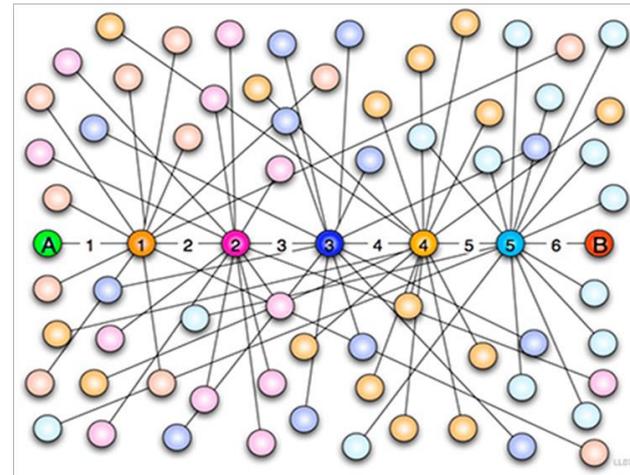
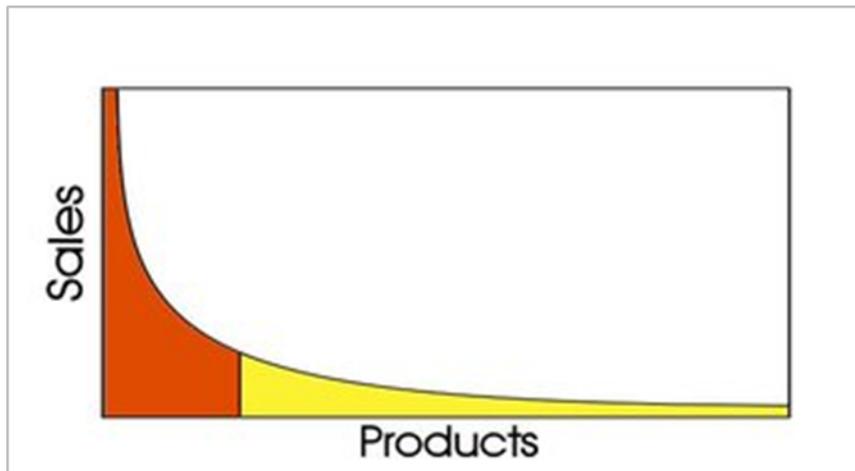
Hermes Trismegisto, in Daniel Stolz von Stolzenberg, *Viridarium Chymicum*, 1624

Albert-László Barabási

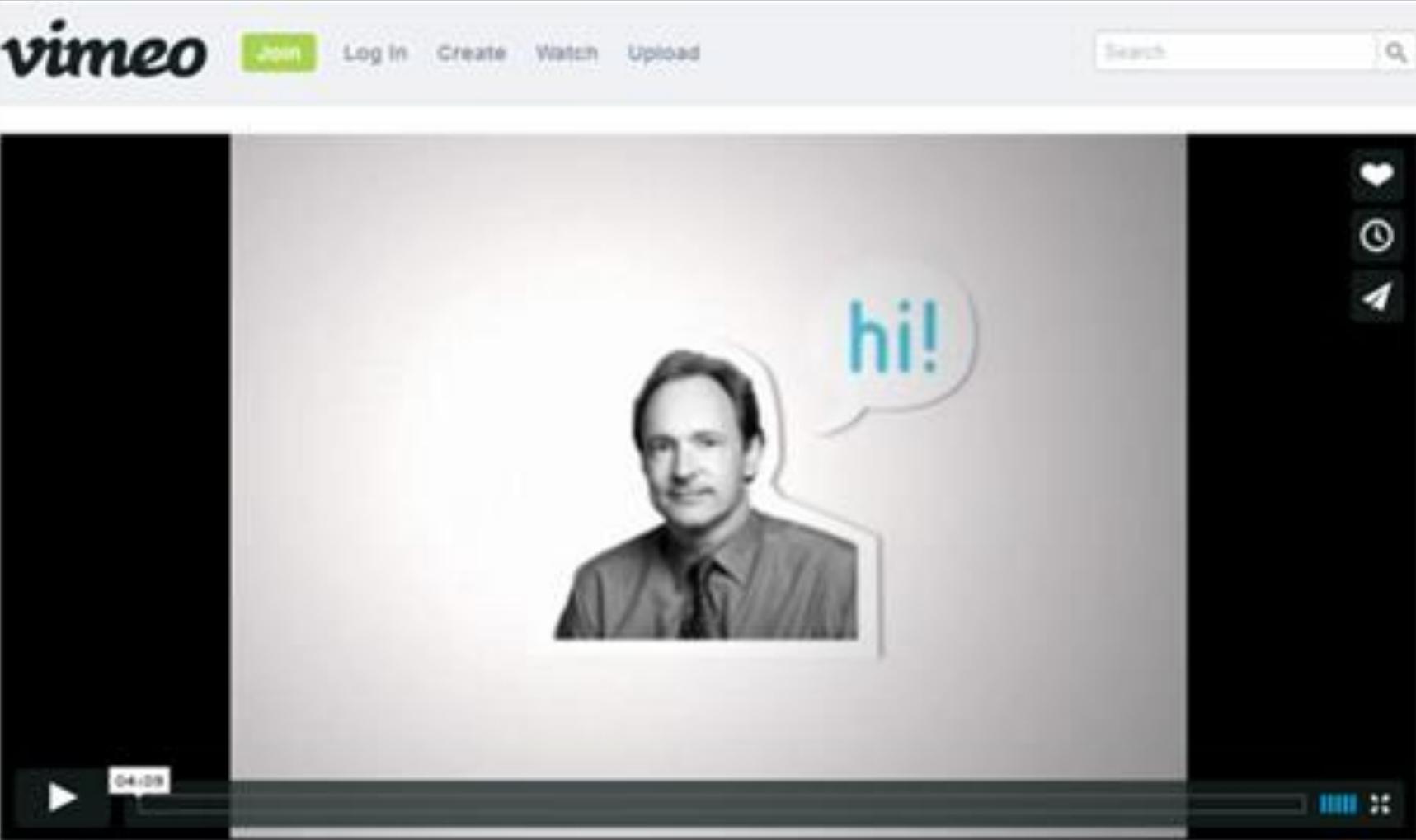
Leggi di potenza

L'applicazione di queste osservazioni al Web ha messo in evidenza che i link si distribuiscono non in modo casuale, ma seguendo **leggi di potenza**, che nelle reti complesse la maggioranza dei nodi ha solo pochi link, e che questi nodi coesistono con pochi altri nodi che hanno un altissimo numero di link, detti hub o connettori, essenziali per garantire coesione alla rete, ed evitarne la frammentazione.

Le più note leggi di potenza sono costituite dalla cosiddetta regola 80/20 (principio di Pareto, lunga coda) e dai sei gradi di separazione.



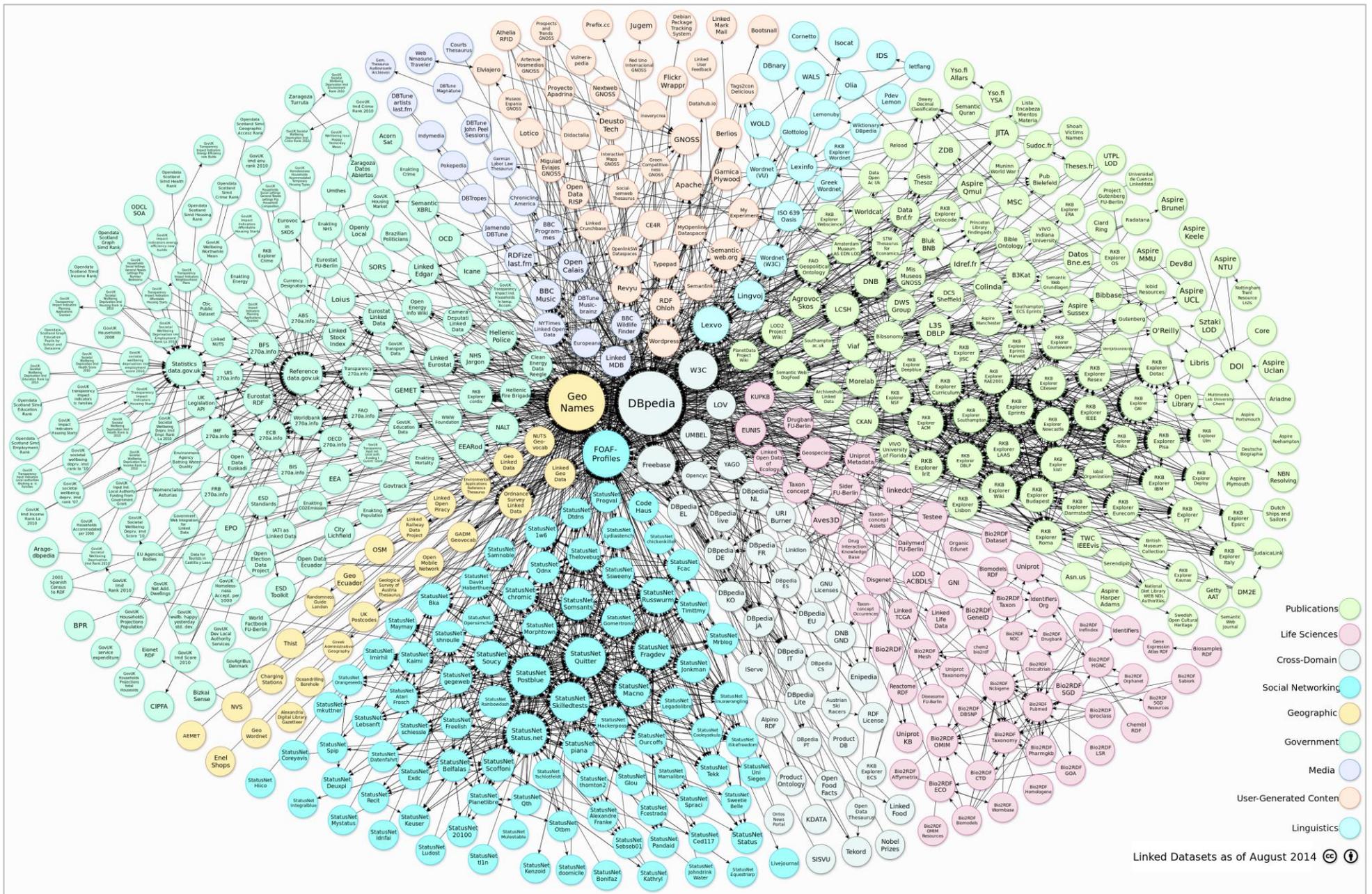
Linked Open Data



The image shows a screenshot of a Vimeo video player. At the top left is the Vimeo logo, followed by navigation links: "Join", "Log In", "Create", "Watch", and "Upload". A search bar is located at the top right. The video player itself shows a video with a man's head inside a speech bubble that says "hi!". Below the video player, the title "Linked Open Data: che cosa sono?" is displayed, along with the uploader "Europeana" and a timestamp of "2 years ago". A description below the title reads: "Europeana.eu - animazione semplice per spiegare cosa Linked Open Data è e perché è una buona cosa, sia per gli utenti e per i fornitori di dati."

<https://vimeo.com/49232562>

Linked Open Data



Linked Datasets as of August 2014 (CC BY)

The Linking Open Data cloud diagram, 30 agosto 2014

Le funzioni della tripla RDF

RDF Triple

subject
verb
object

Armstrong

visited

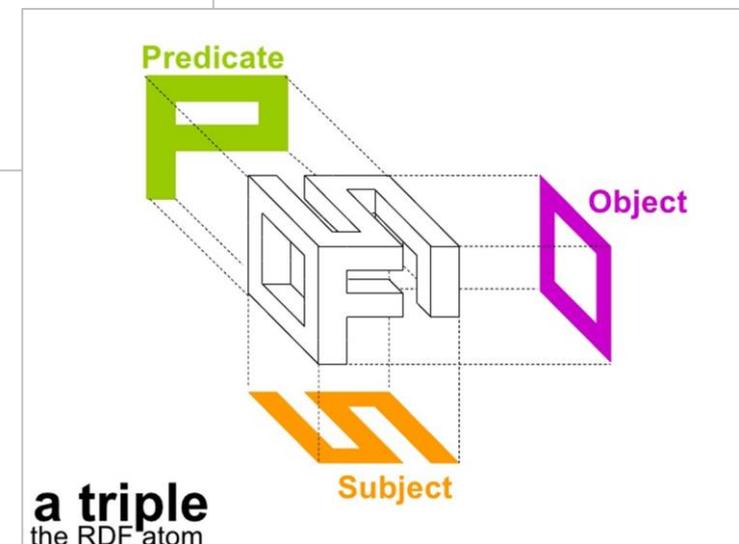
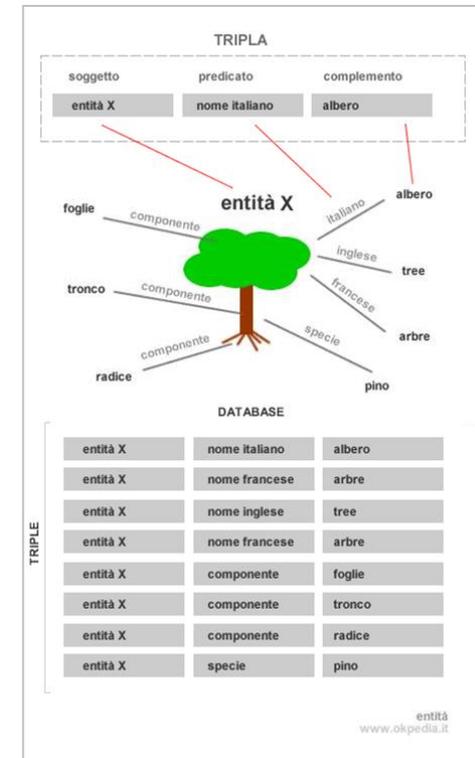
the moon, 1969



WIKIPEDIA
The Free Encyclopedia

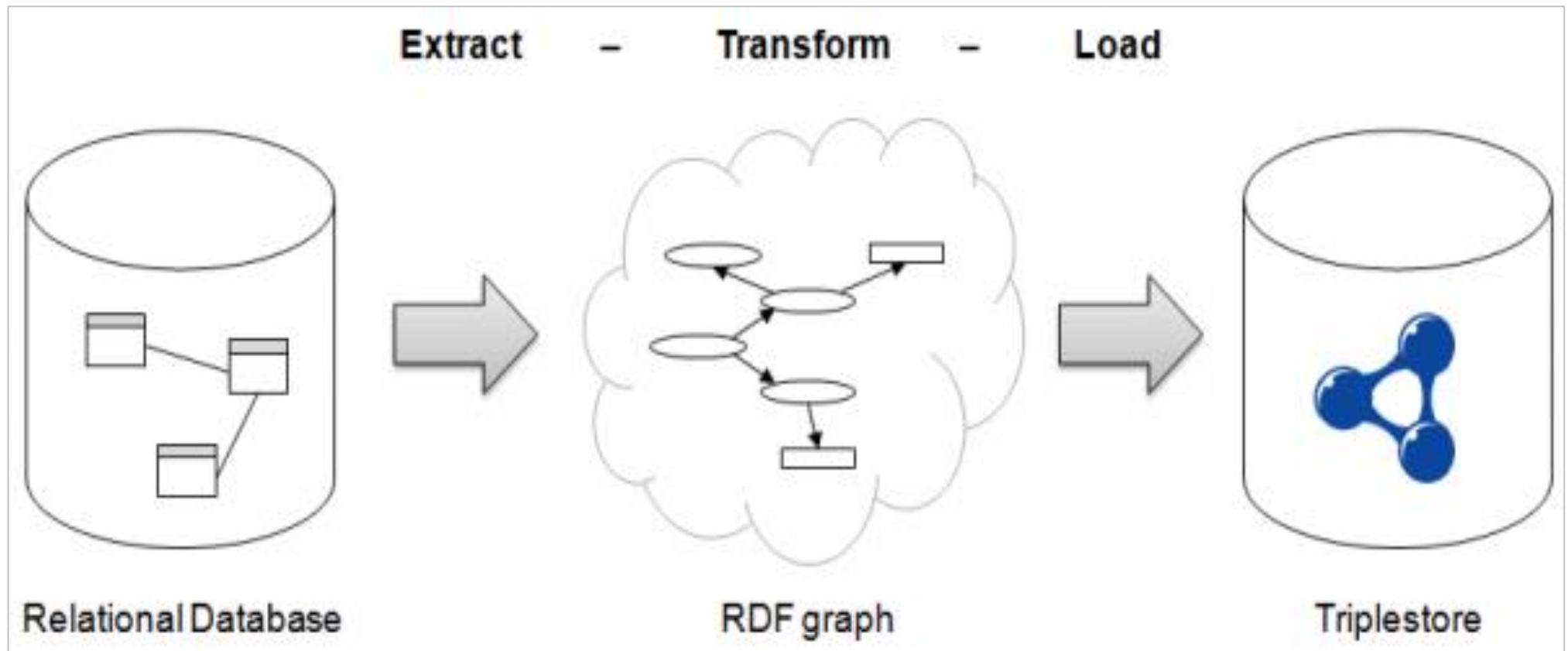






Modelli di rappresentazione delle funzioni della tripla RDF

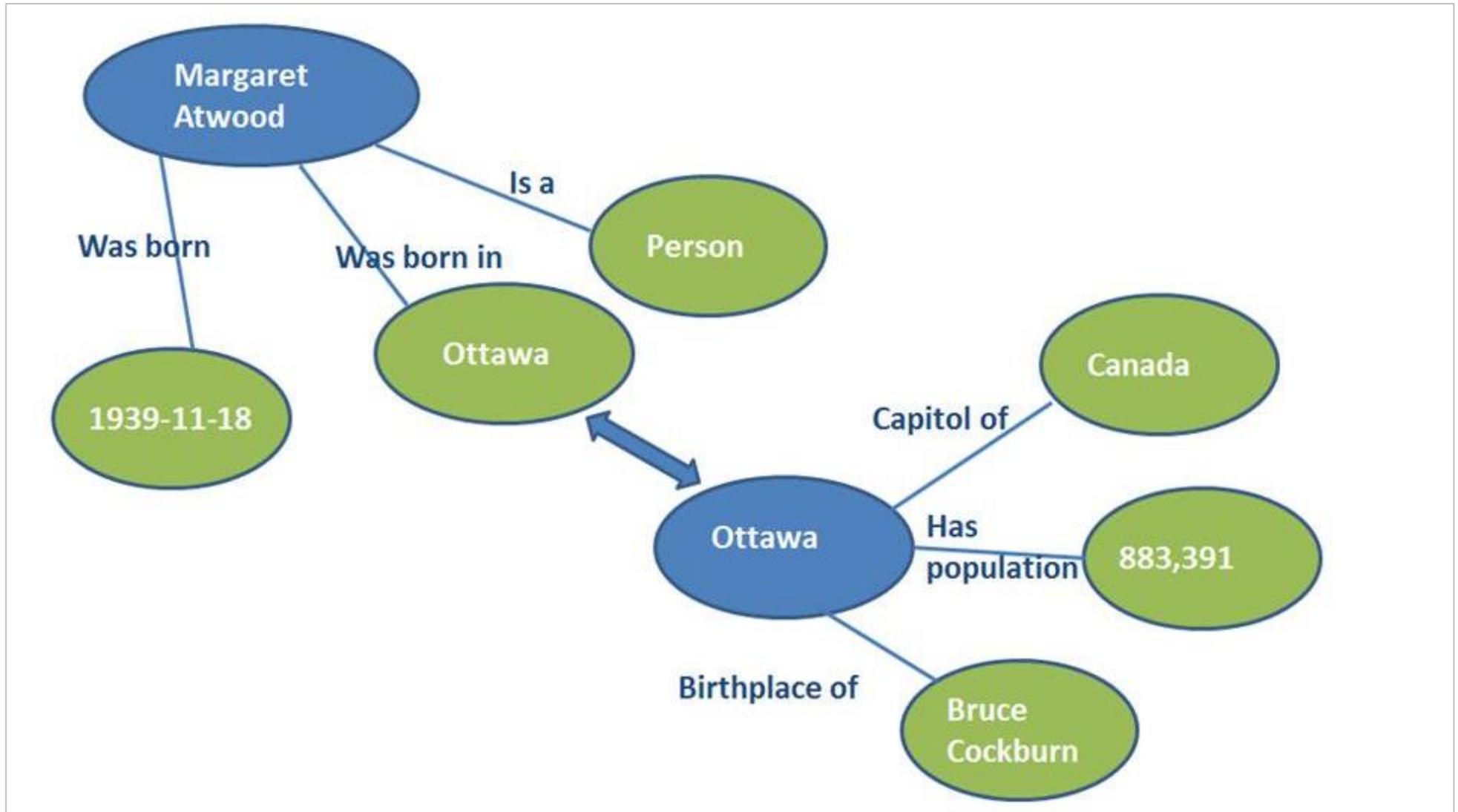
Triplificare



Processo di formazione di una tripla

http://wiki.cubist-project.eu/index.php?title=Data_Preparation

Costruire relazioni



Alison Hitchens, *Linked Open Data and Libraries*,
IST Friday morning seminar, uWaterloo, 2013

I Linked Data e il futuro del catalogo

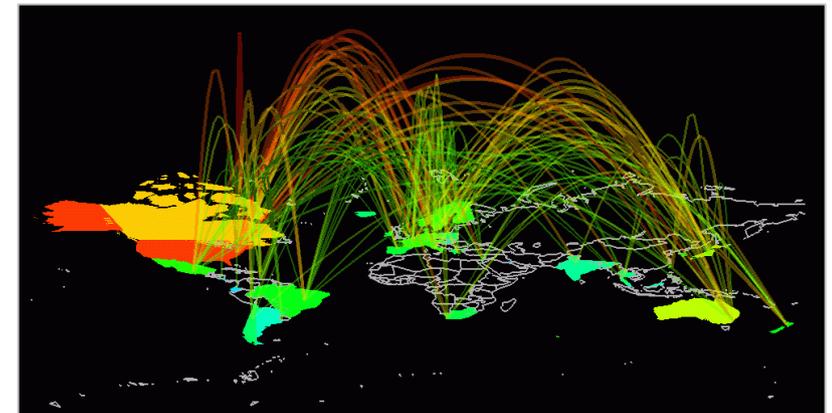
«I cataloghi non sono, infatti, integrati nel web, non sono interrogabili, pur essendo il web il luogo in cui la maggior parte degli utenti lavora, gioca, opera e crea altra informazione. La questione, dunque, è: **Come modificare i cataloghi e i dati perché siano del web e non solo sul web?** È proprio la filosofia che sottende la tecnologia dei linked data che può offrire un interessante punto di partenza per il raggiungimento di questo obiettivo strategico, **pena la morte dei cataloghi**, abbandonati dagli utenti a favore di altri strumenti di reperimento dell'informazione, come i motori di ricerca».

Mauro Guerrini - Tiziana Possemato, *Linked Data: a New Alphabet for the Semantic Web*, «JLIS.it», 4, 2013, 1. DOI: 10.4403/jlis.it-6305.

Futuro del catalogo e culture della complessità

«L'obiettivo [...] è **comprendere la complessità**. Per raggiungere questo scopo dobbiamo andare al di là della struttura e della topologia, e interrogarci sulle **dinamiche che hanno luogo nei link**. Le reti sono soltanto lo scheletro della complessità, i meccanismi su cui si articolano i processi che fanno pulsare il mondo [...] Anche se più avanti ci sono ancora molti "leones" già comincia a delinearsi, continente dopo continente, **la forma di un mondo nuovo**».

A.-L. Barabási, *Link*, 2004



Mappa di Internet, 2005

Ogni linea è tracciata tra due nodi, che rappresentano indirizzi IP.

Colori delle linee: **blu scuro**: net, ca, us; **verde**: com, org; **rosso**: mil, gov, edu; **giallo**: jp, cn, tw, au, de; **magenta**: uk, it, pl, fr; oro: br, kr, nl; **bianco**: sconosciuto.



Grazie per
l'attenzione

