
Ενότητα 4

Ανοιχτή Συνεργασία

Νίκος Καρακαπιλίδης

Industrial Management & Information Systems Lab, MEAD

University of Patras, Greece

nikos@mech.upatras.gr

Μαθησιακοί στόχοι

- Κατανόηση των βασικών χαρακτηριστικών των Κοινοτήτων Πρακτικής και των απαιτήσεών τους σε θέματα συνεργασίας
- Εισαγωγή στο Ελεύθερο Λογισμικό και στο Λογισμικό Ανοιχτού Κώδικα
- Κατανόηση των όρων Big Data, Open Data και Linked Data
- Ανάδειξη της ανάγκης ανάπτυξης και υιοθέτησης ανοιχτών λύσεων
- Γνωστοποίηση σχετικών διεθνών και ελληνικών πρωτοβουλιών και έργων

Περιεχόμενα ενότητας

- Κοινότητες Πρακτικής
- Ελεύθερο Λογισμικό / Λογισμικό Ανοιχτού Κώδικα
- Big Data
- Ανοιχτά Δεδομένα
- Συνδεδεμένα Δεδομένα
- Πρωτοβουλίες και έργα
- Πλατφόρμες Ανοιχτής Συνεργασίας

Συζήτηση



OPEN

Openness in what?

The Open Definition

- The Open Definition sets out principles that define “openness” in relation to **data** and **content**

“Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness)”

Πηγή: <http://opendefinition.org/>

Κοινότητες Πρακτικής

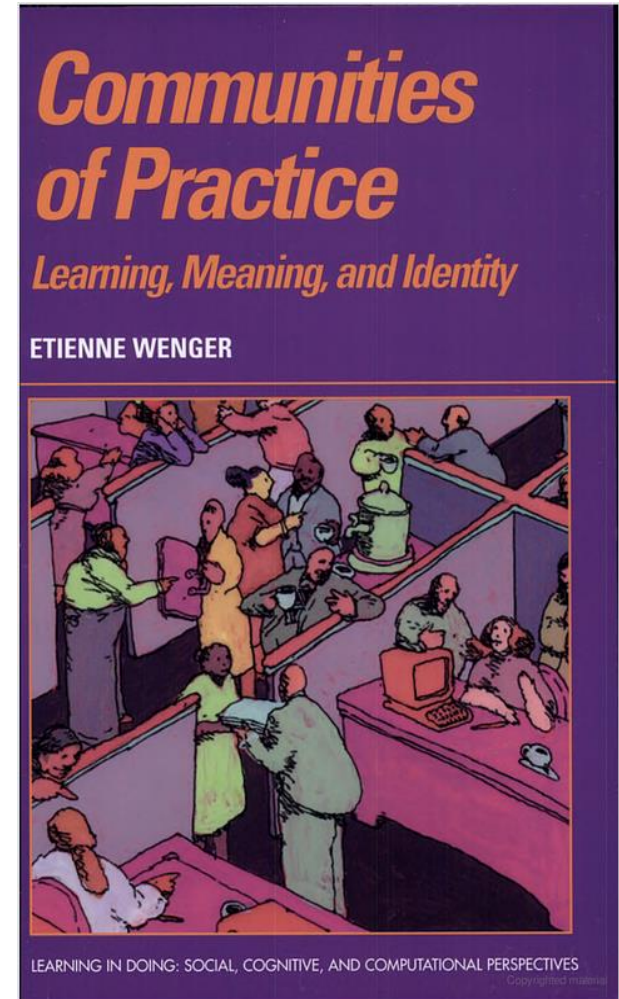
- Ορισμοί

- Communities of practice are groups of people who *share a concern or a passion* for something they do and *learn how to do it better* as they *interact regularly*

Πηγή: <http://wenger-trayner.com/>

- Ομάδες ανθρώπων που συμμερίζονται μια ανησυχία ή ένα πάθος για κάτι που κάνουν και μαθαίνουν πώς να το κάνουν καλύτερα μέσω της ανταλλαγής απόψεων

Πηγή: Μαρμαράς Ν. & Ναθαναήλ Δ., Κοινότητες Πρακτικής και Πληροφοριακά Συστήματα. Στο: Ν. Αβούρης, Χ. Καραγιαννίδης & Β. Κόμης (επιμ.), Συνεργατική Τεχνολογία: Συστήματα και Μοντέλα Συνεργασίας για Εργασία, Μάθηση, Κοινότητες Πρακτικής και Δημιουργία Γνώσης, Κλειδάριθμος, Αθήνα, 2008, σελ. 59-80.

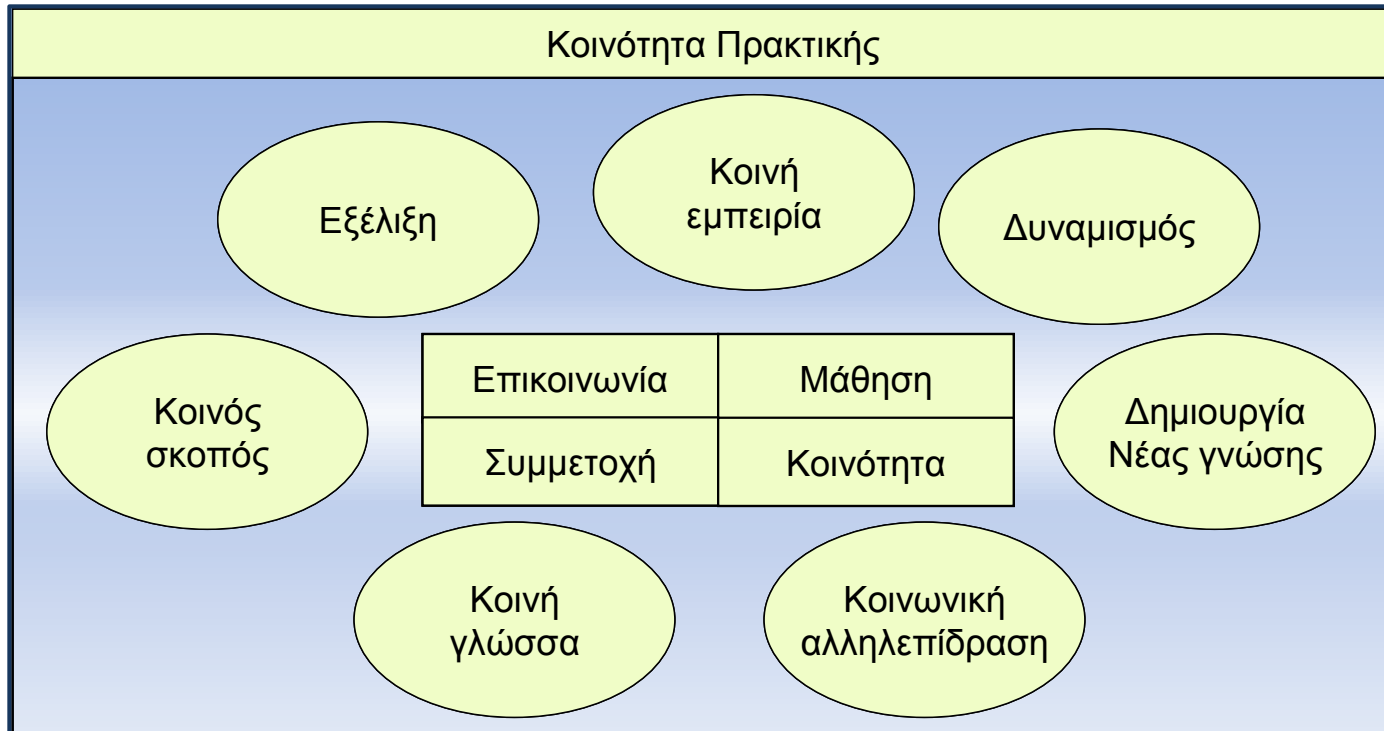


Cambridge University Press, 1999

Κοινότητες Πρακτικής (συν.)

- Ομάδες ανθρώπων που συμμετέχουν σε μια διαδικασία **συλλογικής μάθησης**
- Συνήθως, τα μέλη των ΚΠ έχουν κοινούς στόχους ή/και ενδιαφέροντα, και είναι θετικά διακείμενα στο να **επικοινωνήσουν τη γνώση τους** στα υπόλοιπα μέλη της κοινότητας, επιδιώκοντας το να μάθουν ο ένας από τον άλλο
- Τρεις βασικές διαστάσεις
 - **Κοινή δράση**, η οποία συνεχώς επαναδιαπραγματεύεται από τα μέλη της κοινότητας
 - **Αμοιβαία δέσμευση**, η οποία συνδέει τα μέλη σε μια κοινωνική οντότητα
 - **Κοινοί πόροι**, που αναπτύσσονται από τα μέλη της ΚΠ στο πέρασμα του χρόνου

Τα βασικά χαρακτηριστικά των ΚΠ



Πηγή: Hildreth, P., Kimble, C. and Wright, P. (1998), Computer-mediated communications and communities of practice, in *Proceedings of Ethicomp'98*, March 1998, Erasmus University, The Netherlands, pp. 275-286.

Βασικές απαιτήσεις των ΚΠ

- Διαχείριση πρακτικών και δραστηριοτήτων
- Διαχείριση χρηστών, ρόλων και δικαιωμάτων
- Καθοδήγηση και αναζήτηση βοήθειας
- Ενίσχυση της ροής πληροφοριών και ενημέρωση
- Διαχείριση μεγάλου όγκου πληροφοριών
- Διαχείριση κοινωνικών δομών
- Έκφραση της άρρητης γνώσης
- Αξιοποίηση υπαρχόντων πόρων
- Επεξεργασία δεδομένων για την υποστήριξη διαδικασιών λήψης αποφάσεων
- Διαχείριση υποκειμενικότητας και ασάφειας

Ελεύθερο Λογισμικό /
Λογισμικό Ανοιχτού Κώδικα



CLOSE WINDOWS,
OPEN DOORS



Πηγή: <http://www.gnu.org/software/for-windows.html>

Ελεύθερο Λογισμικό / Λογισμικό Ανοιχτού Κώδικα

- Λογισμικό που μπορεί ελεύθερα να χρησιμοποιηθεί, διανεμηθεί, μελετηθεί, τροποποιηθεί και αναδιανεμηθεί
- Συντηρείται από κοινότητες με ισχυρή «κουλτούρα» κοινοκτημοσύνης και υποστήριξης των χρηστών

- Παραδείγματα

- Linux
- Firefox
- mySQL
- R



Linux



mozilla
Firefox[®]



Ελεύθερο Λογισμικό

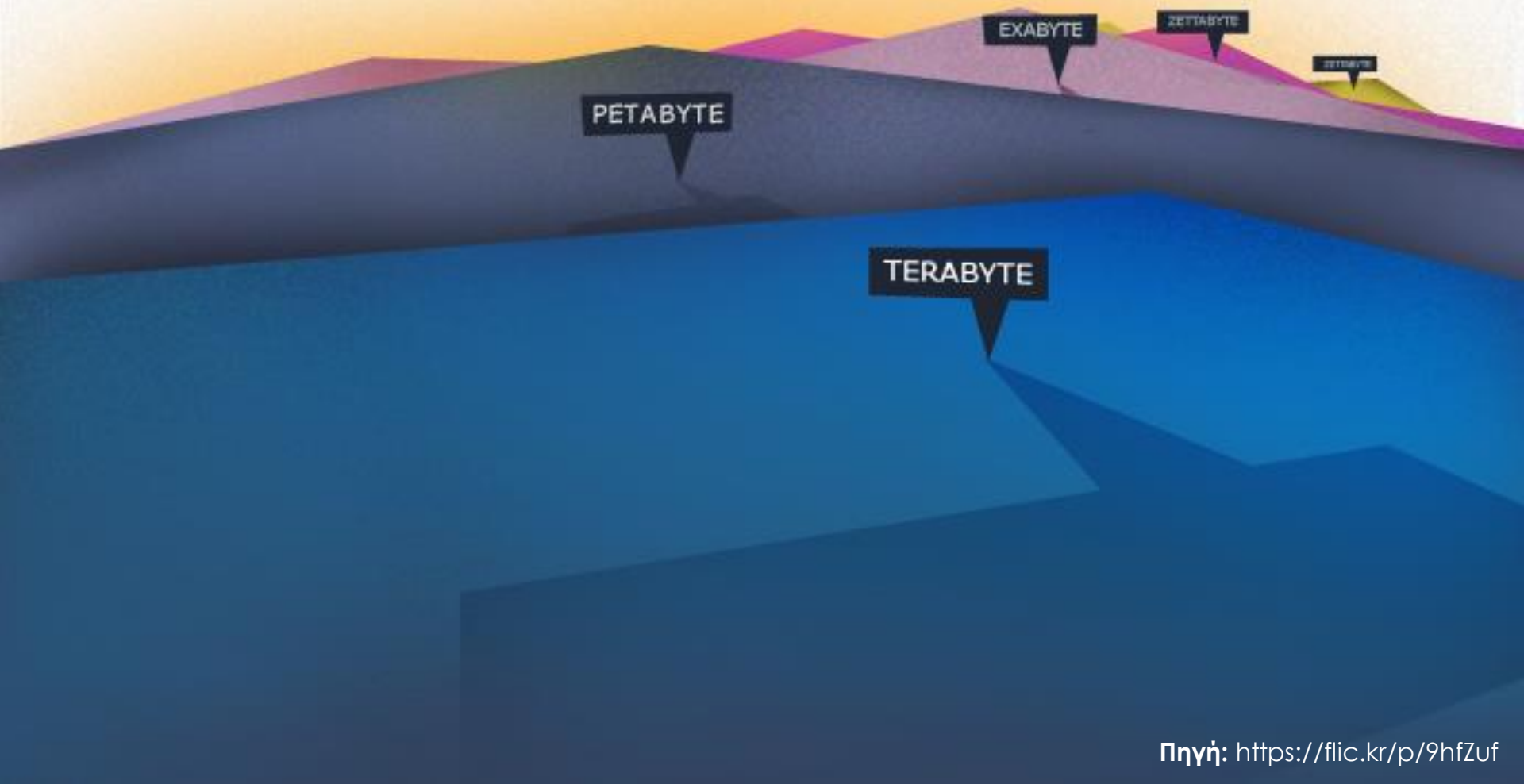
- Τέσσερις βασικές ελευθερίες
 - Η ελευθερία να εκτελούν το πρόγραμμα για οποιονδήποτε σκοπό (ελευθερία 0)
 - Η ελευθερία να μελετούν τον τρόπο λειτουργίας του προγράμματος και να το προσαρμόζουν στις ανάγκες τους (ελευθερία 1)*
 - Η ελευθερία να αναδιανέμουν αντίγραφα του προγράμματος ώστε να βοηθούν το συνάνθρωπο τους (ελευθερία 2)
 - Η ελευθερία να βελτιώνουν το πρόγραμμα και να δημοσιεύουν τις βελτιώσεις που έχουν κάνει στο ευρύ κοινό, ώστε να επωφεληθεί ολόκληρη η κοινότητα (ελευθερία 3)*

* Η πρόσβαση στον πηγαίο κώδικα είναι προϋπόθεση

ΕΛ/ΛΑΚ

- Ο όρος «Ελεύθερο Λογισμικό» παραπέμπει σε κοινωνικό κίνημα
- Ο όρος «Λογισμικό Ανοιχτού Κώδικα» παραπέμπει σε (ανοιχτή) μεθοδολογία ανάπτυξης
- Στην πράξη, εάν και υπάρχουν εξαιρέσεις, κάθε λογισμικό που χαρακτηρίζεται ελεύθερο είναι και ανοικτού κώδικα και αντίστροφα
- Μεγάλη γκάμα αδειών (licenses) → <http://opensource.org/licenses/alphabetical>

Big Data

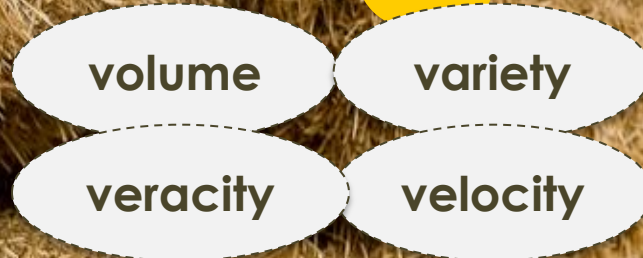


data
intensiveness

cognitive
complexity

Big

sense
making



decision
making

Data

human
reasoning

machine
reasoning

semantics

Data inflation

2

Data inflation

Unit	Size	What it means
Bit (b)	1 or 0	Short for “binary digit”, after the binary code (1 or 0) computers use to store and process data
Byte (B)	8 bits	Enough information to create an English letter or number in computer code. It is the basic unit of computing
Kilobyte (KB)	1,000, or 2^{10} , bytes	From “thousand” in Greek. One page of typed text is 2KB
Megabyte (MB)	1,000KB; 2^{20} bytes	From “large” in Greek. The complete works of Shakespeare total 5MB. A typical pop song is about 4MB
Gigabyte (GB)	1,000MB; 2^{30} bytes	From “giant” in Greek. A two-hour film can be compressed into 1-2GB
Terabyte (TB)	1,000GB; 2^{40} bytes	From “monster” in Greek. All the catalogued books in America’s Library of Congress total 15TB
Petabyte (PB)	1,000TB; 2^{50} bytes	All letters delivered by America’s postal service this year will amount to around 5PB. Google processes around 1PB every hour
Exabyte (EB)	1,000PB; 2^{60} bytes	Equivalent to 10 billion copies of <i>The Economist</i>
Zettabyte (ZB)	1,000EB; 2^{70} bytes	The total amount of information in existence this year is forecast to be around 1.2ZB
Yottabyte (YB)	1,000ZB; 2^{80} bytes	Currently too big to imagine

The prefixes are set by an intergovernmental group, the International Bureau of Weights and Measures.

Source: *The Economist*

Yotta and Zetta were added in 1991; terms for larger amounts have yet to be established.

Πηγή: The Economist, “All too much”, Feb 2010 → <http://www.economist.com/node/15557421>

“Data, data everywhere”

- Wal-Mart, a retail giant, handles more than **1m customer transactions every hour**, feeding databases estimated at more than 2.5 petabytes, the equivalent of 167 times the books in America's Library of Congress
- Facebook, a social-networking website, is home to **40 billion photos**
 - More than **30 billion pieces of content** (web links, news stories, blog posts, notes, photo albums, etc.) shared each month
- Decoding the human genome involves analyzing **3 billion base pairs**, which took **ten years the first time it was done, in 2003, but can now be achieved in one week**

Facebook statistics

Facebook Statistics	Data
Total number of monthly active Facebook users	1,440,000,000
Total number of mobile Facebook users	874,000,000
Increase in Facebook users from 2014 to 2015	12 %
Total number of minutes spent on Facebook each month	640,000,000
Percent of all Facebook users who log on in any given day	48 %
Average time spent on Facebook per visit	18 minutes
Total number of Facebook pages	74,200,000
Facebook Demographics	Data
Percent of 18-34 year olds who check Facebook when they wake up	48 %
Percent of 18-34 year olds who check Facebook before they get out of bed	28 %
Average number of friends per facebook user	130
Average number of pages, groups, and events a user is connected to	80
Average number of photos uploaded per day	205
Number of fake Facebook profiles	81,000,000
Global Facebook Reach Statistics	
Number of languages available on the Facebook site	70
Percent of Facebook users who are outside the United States	75 %

Twitter statistics

Twitter Company Statistics	Data
Total number of registered Twitter users	645,750,000
Total number of active Twitter users	289,000,000
Number of new Twitter users signing up everyday	135,000
Number of unique Twitter site visitors every month	190 million
Average number of tweets per day	58 million
Number of Twitter search engine queries every day	2.1 billion
Percent of Twitter users who use their phone to tweet	43 %
Percent of tweets that come from third party applicants	60%
Number of people that are employed by Twitter	2,500
Number of active Twitter users every month	115 million
Percent of Twitters who don't tweet but watch other people tweet	40%
Number of days it takes for 1 billion tweets	5 days
Number of tweets that happen every second	9,100

Πηγή: <http://www.statisticbrain.com/twitter-statistics>, Nov 2015

The Big Data fallacy

- More data doesn't mean you will get “proportionately” more information
 - In fact, the more data you have, the less information you gain as a proportion of the data
- The **value** of big data is often **overrated**
 - Its value is in the information that it can provide
 - Information is only the **non-redundant portions** of the data, which is a tiny and diminishing fraction of the overall data volume

**The data-information inequality:
information << data**

Big Data concerns

- Data storage and archiving
 - Where and how to store the data?
 - Traditional data storage technologies (e.g. Relational DBMSs) can't handle Big Data
- Data preparation (preprocessing)
 - Manipulation of data into a form suitable for further analysis and processing (preparation is about constructing a dataset from one or more data sources to be used for further exploration and processing)
 - Raw data cannot be processed and must be checked for accuracy

Big Data concerns (cont.)

- Data visualization
 - Clearly communicating information
 - Gaining of insights
- Data mining
 - Discovering consistent patterns and relationships between variables in large data sets
 - Extracting information from a data set and transform it into an understandable structure for further use

Big Data white paper

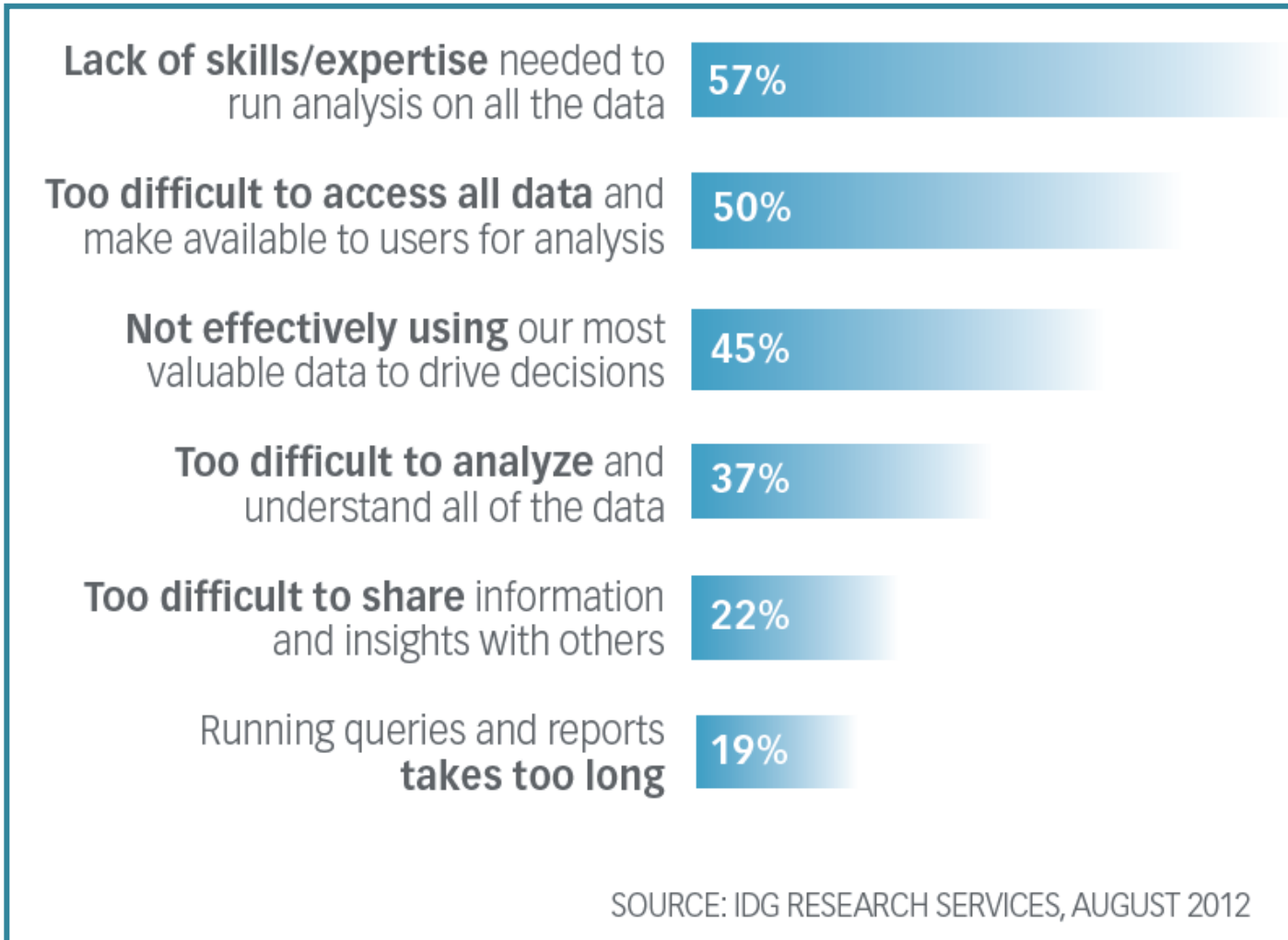
- In spite of the tremendous advances made in computational analysis, there remain many patterns that humans can easily detect but computer algorithms have a hard time finding
- With Big Data, the use of separate systems becomes prohibitively expensive ... Big Data has made it necessary to run heterogeneous workloads on a single infrastructure that is sufficiently flexible to handle all these workloads

Divyakant Agrawal, UC Santa Barbara
Philip Bernstein, Microsoft
Elisa Bertino, Purdue Univ.
Susan Davidson, Univ. of Pennsylvania
Umeshwar Dayal, HP
Michael Franklin, UC Berkeley
Johannes Gehrke, Cornell Univ.
Laura Haas, IBM
Alon Halevy, Google
Jiawei Han, UIUC
H. V. Jagadish, Univ. of Michigan (Coordinator)
Alexandros Labrinidis, Univ. of Pittsburgh
Sam Madden, MIT
Yannis Papakonstantinou, UC San Diego
Jignesh M. Patel, Univ. of Wisconsin
Raghu Ramakrishnan, Yahoo!
Kenneth Ross, Columbia Univ.
Cyrus Shahabi, Univ. of Southern California
Dan Suciu, Univ. of Washington
Shiv Vaithyanathan, IBM
Jennifer Widom, Stanford Univ.

Big Data white paper (cont.)

- Ideally, analytics for Big Data will not be all computational – rather it will be designed explicitly to have a **human in the loop**
- It is rarely enough to provide just the results ... rather, one must provide supplementary information that **explains how each result was derived**, and based upon precisely what inputs
- Systems with a **rich palette of visualizations** become important in conveying to the users the results of the queries in a way that is best understood in the particular domain

Big Data challenges



Open Data

- Definition
 - Open data is data that can be **freely used, re-used and redistributed by anyone** - subject only, at most, to the requirement to attribute and share alike
- Interoperability issue
 - Lots of datasets but **little or no ability to combine them together** into the larger systems where the real value lies

The logo consists of a horizontal rectangle with a thin black border. The word "OPEN" is written in white, uppercase letters on a light blue background. The word "DATA" is written in white, uppercase letters on a darker blue background.

OPEN DATA

Open Data Issues

- Availability and Access
 - The data must be **available as a whole** and at no more than a **reasonable reproduction cost**, preferably by downloading over the internet
 - The data must also be available in a **convenient and modifiable form**
- Re-use and Redistribution
 - The data must be provided under terms that **permit re-use and redistribution** including the intermixing with other datasets
- Universal Participation
 - Everyone must be able to use, re-use and redistribute - there should be **no discrimination against fields of endeavor or against persons or groups**
 - For example, **'non-commercial' restrictions** that would prevent **'commercial' use**, or restrictions of use for certain purposes (e.g. only in education), **are not allowed**

Open Content

- The term "open content" describes any **copyrightable work** (traditionally excluding software, which is described by other terms like "open source") that is licensed in a manner that **provides users with free and perpetual permission** to engage in the 5R activities:
 - **Retain:** the right to make, own, and control copies of the content (e.g., download, duplicate, store, and manage)
 - **Reuse:** the right to use the content in a wide range of ways (e.g., in a class, in a study group, on a website, in a video)
 - **Revise:** the right to adapt, adjust, modify, or alter the content itself (e.g., translate the content into another language)
 - **Remix:** the right to combine the original or revised content with other open content to create something new (e.g., incorporate the content into a mashup)
 - **Redistribute:** the right to share copies of the original content, your revisions, or your remixes with others (e.g., give a copy of the content to a friend)

Linked Data

- The term “Linked Data” refers to a set of best practices for publishing and connecting structured data on the Web
- Linked Data is about using the Web to connect related data that wasn't previously linked, or using the Web to lower the barriers to linking data currently linked using other methods
- Wikipedia: “A term used to describe a recommended best practice for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web using URIs and RDF”

Linked Data - Key technologies

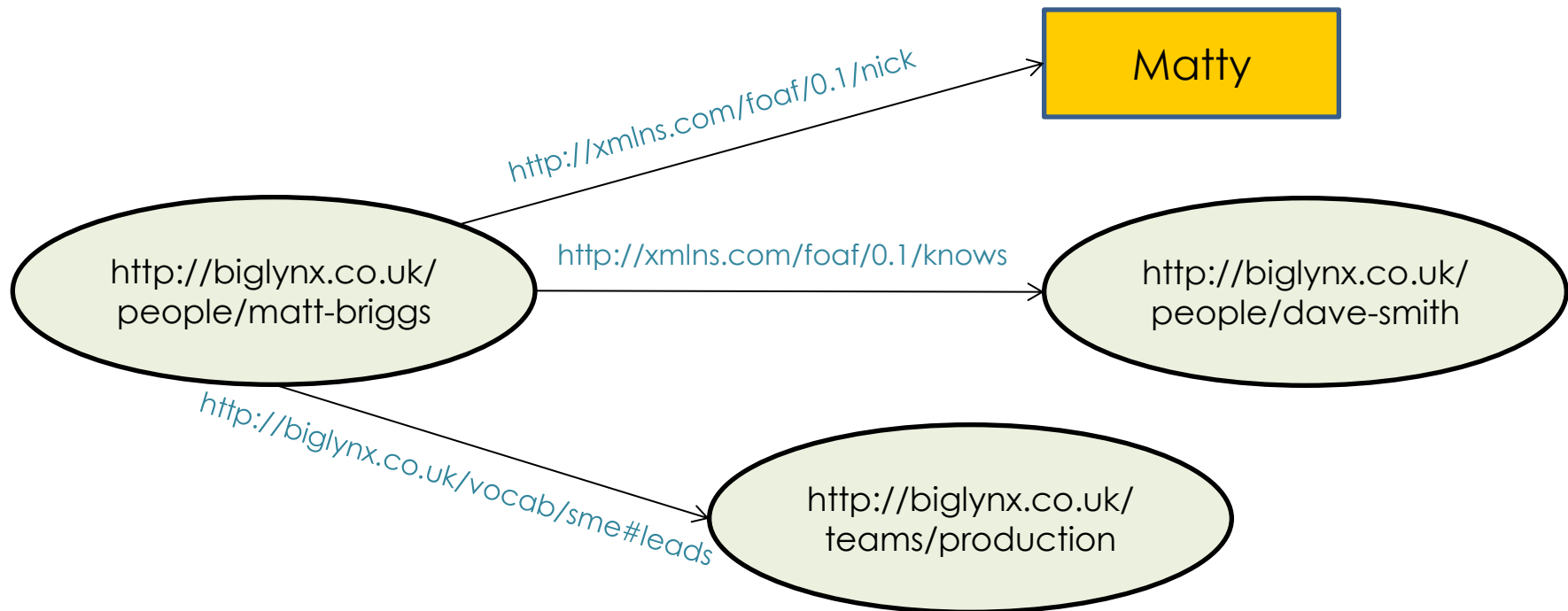
- URIs (Uniform Resource Identifiers)
 - A generic means to **identify entities or concepts** (e.g., a digital resource, a concept within a vocabulary) in the world
 - A URI can be either a **Uniform Resource Locator (URL)**, a **Uniform Resource Name (URN)**, or both
- HTTP (Hypertext Transfer Protocol)
 - A simple yet universal mechanism for **retrieving resources**, or descriptions of resources
 - The **foundation of data communication** for the World Wide Web
- RDF (Resource Description Framework)
 - A generic **graph-based data model** with which to structure and link data that describes “things” in the world
 - A description of a resource is represented as **a number of triples (subject, predicate, object)**

The RDF Data Model

[<http://biglynx.co.uk/people/matt-briggs> <http://xmlns.com/foaf/0.1/nick> "Matty"]

[<http://biglynx.co.uk/people/matt-briggs> <http://xmlns.com/foaf/0.1/knows>
<http://biglynx.co.uk/people/dave-smith>]

[<http://biglynx.co.uk/people/matt-briggs> <http://biglynx.co.uk/vocab/sme#leads>
<http://biglynx.co.uk/teams/production>]



The Linked Data Paradigm

- Use URIs as names for “things”
- Use HTTP URIs so that people can look up those names
- When someone looks up a URI, provide useful information (using standards such as RDF* and SPARQL)
- Include links to other URIs, so that they can discover more “things”



Linked Open Data (LOD)

- The basic LOD idea
 - Publish **data** (not documents) of public interest on the Web
 - **Interlink** these data with those of other data sources
- **Humans and applications** can **easily access data** using the appropriate Web technologies and **follow the links** provided in order to obtain further contextual information

Linked Data Search Engines

The screenshot shows a Google search for "birth date of Madonna". The search bar is circled in red. The results page features a structured data card for Madonna's birth date: "August 16, 1958 (age 57 years)". Below this, there are smaller cards for other celebrities: Jennifer Lopez (July 24, 1969), Michael Jackson (August 29, 1958), and Britney Spears (December 2, 1981). To the right, a detailed profile for Madonna is shown, including her birth date, full name, children, spouse, songs, and events. Three callout boxes with arrows point to specific elements: the first points to the search bar, the second to the structured data card, and the third to the detailed profile.

use of structured data from the Web to enrich search results

use of extracted data to directly answer simple factual questions

major search engines have begun to evolve into answering engines which rely on structured data from the Web

birth date of Madonna

Web Images Videos News Maps More Search tools

About 53,800,000 results (0.46 seconds)

Madonna / Date of birth

August 16, 1958 (age 57 years)

Jennifer Lopez July 24, 1969

Michael Jackson August 29, 1958

Britney Spears December 2, 1981

Feedback

Madonna - Film Actress, Singer - Biography.com
www.biography.com/people/madonna-9394994

Early Life. Singer, performer and actress Madonna Louise Veronica Ciccone was born in Bay City, Michigan, on **August 16, 1958**, to parents Silvio "Tony" Ciccone and Madonna Fortin.

Madonna (entertainer) - Wikipedia, the free encyclopedia
[https://en.wikipedia.org/wiki/Madonna_\(entertainer\)](https://en.wikipedia.org/wiki/Madonna_(entertainer))

Madonna was born to Catholic parents Silvio Anthony "Tony" Ciccone and Madonna Louise Fortin (c. 1933 – December 1, 1963) in Bay City, Michigan, on **August 16, 1958**.
[Rebel Heart - Madonna singles discography](#) - [Madonna albums discography](#) - [Evita](#)

Madonna - Biography - IMDb
www.imdb.com/name/nm0000187/bio

Jump to [Mini Bio](#) - The remarkable, hyper-ambitious Material Girl who never stops reinventing herself, **Madonna** is a seven-time Grammy Award-winner ...

Actress: Madonna, herescope for birth date 16 August

Madonna
Singer-songwriter

Madonna Louise Ciccone is an American singer, songwriter, actress, and businesswoman. She achieved popularity by pushing the boundaries of lyrical content in mainstream popular music and imagery in her music videos, which became a fixture on MTV. [Wikipedia](#)

Born: August 16, 1958 (age 57), Bay City, Michigan, United States
Full name: Madonna Louise Ciccone
Children: Lourdes Maria Ciccone Leon, David Banda Mwale Ciccone Ritchie, Rocco Ritchie, Mercy James
Spouse: Guy Ritchie (m. 2000–2008), Sean Penn (m. 1985–1989)

Songs

Bitch I'm Madonna	2015	Rebel Heart
La Isla Bonita	1986	True Blue
Like A Virgin	1984	Like a Virgin
Like a Prayer	1989	Like a Prayer
Don't Cry for Me Argent...	1996	Evita

Events

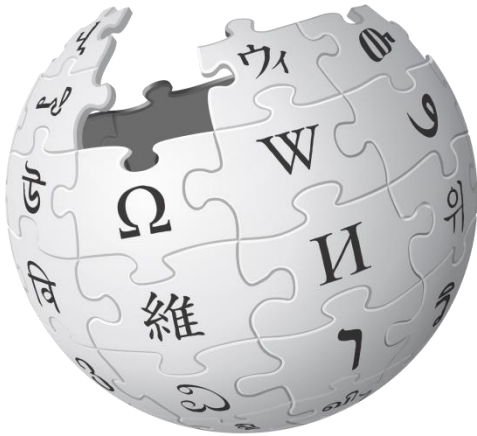
Barcelona, Spain Palau Sant Jordi	Tue Nov 24
Barcelona, Spain Palau Sant Jordi	Wed Nov 25
Antwerp, Belgium Antwerps Sportpaleis	Sat Nov 28

Profiles

DBpedia

- A crowd-sourced community effort to extract structured information from Wikipedia ...
 - ... and make it available on the Web
- The whole DBpedia data set (as of Sep 2014)
 - describes 4.58 million entities, out of which 4.22 million are classified in a consistent ontology, including 1,445,000 persons, 735,000 places, 123,000 music albums, 87,000 films, 19,000 video games, 241,000 organizations, 251,000 species and 6,000 diseases
 - features labels and abstracts for these entities in up to 125 different languages, 25.2 million links to images and 29.8 million links to external web pages
 - in addition, it contains around 50 million links into other RDF datasets

Συζήτηση



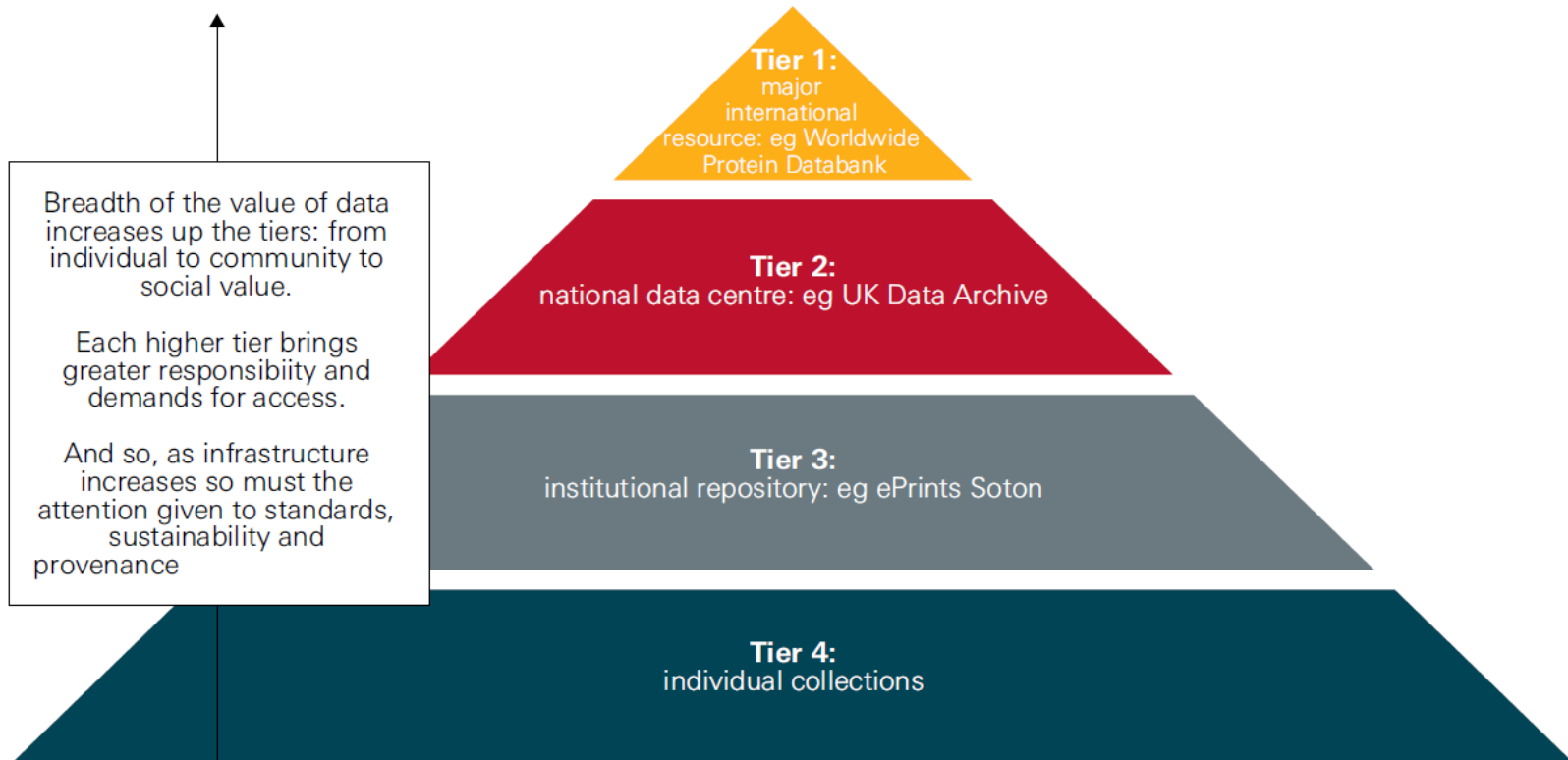
WIKIPEDIA
The Free Encyclopedia

Web of
documents



Web of
data

The Data Pyramid – rising value and permanence



Πηγή: "Science as an open enterprise". The Royal Society Science Policy Centre report. June 2012.
Available from: http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf

Πρωτοβουλίες & έργα

Talk - Action = Shit



Brendan Dawes is a designer exploring the interaction of objects, people, technology and art using form and code with an eclectic mix of digital and analog materials.

<http://www.brendandawes.com/>

ERCIM NEWS

www.ercim.eu



ERCIM News

Special theme: Scientific Data sharing and re-use

Number 100, January 2015

ISSN 0926-4981

<http://ercim-news.ercim.eu/images/stories/EN100/EN100-web.pdf>

Also in this issue:

Keynote:
by Carlos Morais, European Commission

Joint ERCIM Actions:
100 issues of ERCIM News
ERCIM 25 Years Celebration

Research and Innovation:
Simulations Show How
Lightning Creates Antimatter

Science 2.0 Repositories

The screenshot shows a web browser window titled "John Smith Home". The interface includes a navigation bar with "Home", "Workspace", "Notifications", and "Messages". A search bar for "Search RAs" is present. The main content is divided into three sections:

- My Research Activities** (7 items): A list of activities including Research Activity 1, 2, and 3, with sub-items like Experiment 2, Dataset 2, Experiment 3, and Processes 3.1 and 3.2.
- Trending Research Activities**: A list of five activities, such as "Harmonize the data according to the EU recommended vocabularies" and "Lorenzino's master thesis".
- Research Activities Feed**: A feed of activity updates. The first entry, "Process 1", shows a table preview and a comment from Leonardo. The second entry, "Process 2", shows a map preview and a comment from Elisabeth. A third entry by Carol is partially visible. An "Actions" menu is overlaid on the second entry, listing "View experiment", "Repeat experiment", and "Repurpose experiment".

Each feed entry includes a timestamp, a "Comment" button, social media icons (thumbs up/down, share), and "Comments" and "Likes" counts.

Science 2.0 Repositories (cont.)

The screenshot displays a web interface for a research activity. At the top, there are navigation tabs for Home, Workspace, Notifications, and Messages. The main title is "Research Activity Example". Below this, there are statistics for the activity: Rating 6.8, Citations 132, and Likes 345.

The central content area features an **Abstract** section with the text: "Examining the TCGA Methylation (Glioblastoma multiforme) dataset, to assess whether we can use it as a means for identification of potential biomarkers for clinical development. To bridge from the Illumina Human Methylation 27/450 CG# identifier codes back to standard NCBI Gene ...". Below the abstract is an **Experiments** section with a table:

Processes	Input	Output	Result	Actions
P1	Dataset 1	Dataset 2	Positive	⚙️ ▼
P2	Dataset 2	Dataset 3	Negative	⚙️ ▼
...
...
...

An **Explore Dataset 3** modal window is open, showing a **Table** and a **Chart** visualization.

At the bottom, there is a **Comments and Discuss** section with two comments:

- Paul** said: "We have dataset containing large numbers of proteins belonging to PCB and unknown sources. We have no idea about their nucleotide sequences, but we are interested in understanding codon preference in these proteins." (Rating: 4 stars)
- Francis** said: "I have a dataset from a single gene being sequenced in 10000 subjects, how do I compare the variants found in my dataset with [Dataset 2](#)?" (Rating: 5 stars)

On the left sidebar, there are sections for **Research Products** (12 items) and **Persons involved** (7 persons). The Research Products section includes Manuscripts, Workflows (WF 1, WF 2), and Datasets (Dataset 1, Dataset 2 with File set 2.1 and File set 2.2). The Persons involved section lists P. Schrute, F. Halpert, and P. Opti with their contact information.



Research Data Sharing without barriers



Share your news with RDA, write to news@rd-alliance.org

Search input field with magnifying glass icon

- Home
- Organisation
- Working and Interest Groups
- Plenary Meetings
- News & Events
- Early Career Programmes
- About

6th PLENARY || PROGRAMME || Climate Change Data Challenge || Experimentation Day || PhotoGallery || Recordings ||

Request for Comments

- DSA-WDS Partnership for Certification WG Recommend...**
By Herman Stehouwer
- Data Citation Recommendation**
By Herman Stehouwer
- Chemistry Research Data Interest Group**
By Kathleen Fontaine On 30/10/2015
- Array Database Working Group**
By Peter Baumann On 19/10/2015

RDA Outputs

RDA group outputs are focused on tangibly accelerate progress for global data sharing and increase data-driven innovation.

- [VIEW ALL](#)
- [DOWNLOAD the RDA Outputs booklet](#)

Focus

Call for Nominations and Expressions of Interest for RDA Council

The Governance Documents of RDA and the RDA Foundation indicate that Council members are to serve three-year (renewable) terms, with three members selected each year. Members of RDA are encouraged to submit names for the three open Council positions no later than 31 December 2015.

[READ MORE](#)

News

Only One Week Left Until "The Future of the Commons: Data, Software and Beyond"

On Wednesday, November 18, 2015, RDA/US, CENDI and the National Federation of

Recent Group Activity

RDA Members



EUROPEAN DATA PORTAL (BETA)

Search Datasets

[Advanced Search \(SPARQL\)](#)
BETA VERSION
You can access the data tools from here

Latest News

- 13/11/2015
Follow the European Data Portal on social media
- 12/11/2015
French governments' Open Data budget will more than double in 2016
- 10/11/2015
The Open Data Charter: A Roadmap for Using a Global Resource

[News Archive](#)

Browse Categories

- Agriculture, Fisheries, Forestry & Foods
- Energy
- Regions & Cities
- Transport
- Economy & Finance
- International Issues
- Government & Public Sector
- Justice, Legal System & Public Safety
- Environment



EUDAT: the collaborative Pan-European infrastructure providing research data services, training and consultancy for



Researchers



Research Communities



Research Infrastructures & Data Centres



B2DROP

Sync and Exchange Research Data

Read more!
use



B2SHARE

Store and Share Research Data

Read more!
use



B2SAFE

Replicate Research Data Safely

Read more!



B2STAGE

Get Data to Computation

Read more!



B2FIND

Find Research Data

Read more!
use



TECHNICAL AND HUMAN INFRASTRUCTURE FOR OPEN RESEARCH

The THOR mission

THOR is a 30 month project funded by the European Commission under the Horizon 2020 programme. It will establish seamless integration between articles, data, and researchers across the research lifecycle. This will create a wealth of open resources and foster a sustainable international e-infrastructure. The result will be reduced duplication, economies of scale, richer research services, and opportunities for innovation.

THOR launches later this year, and follows on from the successful [ODIN](#) (ORCID and DataCite Interoperability Network) project. The project has four concrete aims:

1. **Establishing interoperability**
2. **Integrating services**
3. **Building capacity**
4. **Achieving sustainability**

The project will meet these aims by defining relations between contributors, research artefacts (including data), and organizations. We will incorporate these relationships into the ORCID and DataCite systems. We will also expand existing linkages between different types of identifiers and versions of artefacts to improve interoperability across platforms and integrate ORCID iDs into production systems for article and data submission services in pilot communities and beyond.

RECENT POSTS

- [Contributor Information in DataCite Metadata](#)
October 12, 2015
- [Differences between ORCID and DataCite Metadata](#)
September 18, 2015
- [The THOR Ambassador Programme is GO!](#)
September 16, 2015
- [Interactive API docs for ORCID](#)
September 11, 2015
- [Digging into DataCite Metadata using R](#)
September 8, 2015

MORE FROM ORCID

- [Coding with the GitHub Octocat](#)
- [All About #ORCID15](#)
- [ORCID Around the World](#)



PARTICIPATE

SEARCH

MONITOR

SUPPORT

OPEN ACCESS

Supporting Open Science



Deposit Publications & Data

Link Research Results

Validate / Register Repository

Content policy

A **BRIDGE** BETWEEN **ALL** RESEARCH STAKEHOLDERS
AND THE WORLD OF **SCHOLARLY PUBLICATION**

OpenAIRE is a network of Open Access repositories, archives and journals that support Open Access policies. It goes beyond the traditional publications aggregator by interconnecting entities related to scholarly communication (publications, research data, funding, people, organizations, data sources) allowing users to navigate alongside a rich information space graph and provides a wide range of services, from deposition to statistics.

OpenAIRE has started out as a policy support mechanism for the EC (FP7 pilot and H2020 OA policies), with the aim to be *the* European scholarly communication hub providing its services to many European funders.

The OpenAIRE team is open to collaborations. Please contact us to find more on our activities and see how we can be of further assistance.

OpenAIRE is an EC funded project (FP7 246686 and 283595)



OpenAIRE has access to 13,336,838 publications and 11,018 datasets from 6,053 data sources. These involve 98,410 projects and 17,961 organizations.

See
what it's
all about

WATCH THE VIDEO





The European Open Science Cloud for Research



Leading European initiatives, [EUDAT](#), [LIBER](#), [OpenAIRE](#), [EGI](#) and [GEANT](#) share their joint vision for the European Open Science Cloud for Research which includes eight elements of success for a concrete contribution to the Digital Single Market.



The Open Science Cloud, part of the European Commission's Digital Single Market Strategy, will Europe for the benefit of innovation and growth. Today's joint statement sets out the partners' strategic governance as a contribution towards the practical realisation of the EC's vision.

The European Open Science Cloud initiative is strategically important as Open Science is a key driver of innovation. To harness its full value and reap the fruits of public and private investment, Europe needs to focus on analysis, sharing, reuse and preservation of research data on which innovative services can be developed.

As Europe can and must build on decades of public investment in scientific infrastructures—existing and emerging, cloud services, scientific software and institutional and community data repositories—by co-

- Basic elements for its success:**
1. Open in design, participation and use
 2. Publicly funded & governed
 3. Research-centric
 4. Comprehensive
 5. Diverse & distributed
 6. Interoperable
 7. Service-oriented & protocol-centric
 8. Social connecting diverse communities

CC Search

search.creativecommons.org

creative commons

CC Search

Find content you can share, use and remix

collaboration

I want something that I can... use for *commercial purposes*;
 modify, adapt, or build upon.

Search using:

- Europeana
Media
- Flickr
Image
- Google
Web
- Google Images
Image
- Jamendo
Music
- Open Clip Art Library
Image
- SpinXpress
Media
- Wikimedia Commons
Media
- YouTube
Video
- Pixabay
Image
- ccMixer
Music
- SoundCloud
Music

Τα **Creative Commons** είναι ένας μη κερδοσκοπικός οργανισμός που έχει ως βασικό στόχο την υποστήριξη του ανοικτού και προσβάσιμου διαδικτύου, καθώς και τη χρήση, διανομή και αξιοποίηση της γνώσης και της δημιουργικότητας μέσω ελεύθερων αδειών

Δράσεις ΕΛ/ΛΑΚ



<https://ellak.gr/>



Ανοιχτή διακυβέρνηση
opengov.ellak.gr



Ανοιχτά Δεδομένα
opendata.ellak.gr



Ανοιχτά Πρότυπα
openstandards.ellak.gr



Ανοιχτό Λογισμικό
mathe.ellak.gr



Ανοιχτό Περιεχόμενο
mycontent.ellak.gr



Ανοιχτό Hardware
openhardware.ellak.gr



Άδειες CC
creativecommons.gr



Ελεύθερο Λογισμικό
Εκπαίδευσης
edu.ellak.gr



Search - Greek Linked Open Data

[Filter by location »](#)

Coastline of Greece

[api/sparql](#) [rdf, tar.bz2](#)

This dataset describes the geometry of the coastline of Greece. For each area contained in the shapefile a unique URI is created and a spatial literal is attributed to it. The spatial...

Corine Land Cover of Greece

[api/sparql](#) [rdf, tar.bz2](#)

The Corine Land Cover project22 is an activity of the European Environment Agency that collects data regarding the land cover of European countries. The project uses a hierarchical...

Greek Administrative Geography

[api/sparql](#) [rdf, tar.bz2](#)

Kallikratis is the most recent administrative organization of Greece as explained in the relevant Wikipedia page. The hierarchical administrative organization of Greece according to...

Rivers and Lakes of Greece

[rdf, tar.bz2](#)

This dataset contains various information about greek rivers and lakes. For each object contained in the shapefile a unique URI is created and a spatial literal is attributed to it. The...

Public Buildings

[rdf, tar.bz2](#)

The Public Buildings dataset provides information regarding buildings of public administration, being sorted in categories according to the services they provide. For each building we...

Public Transport

[rdf/xml, bz2](#)

The Public Transport dataset provides information regarding stops and stations of greek public transport, such as name, code as well as a spatial literal indicating its position on the...

Resource formats

- [rdf](#) (3)
- [api/sparql](#) (3)
- [rdf, tar.bz2](#) (2)
- [nt](#) (2)
- [tar.bz2, rdf](#) (1)
- [png](#) (1)
- [pdf](#) (1)
- [n3](#) (1)

Other access

You can also access this registry using the [API](#) (see [API Docs](#)).

Ανοικτά δεδομένα, προσβάσιμα σε όλους!

Μάθετε περισσότερα

Γαλάζιες σημαίες 2010

Αναζήτηση Δεδομένων



Τελευταίες Ενημερώσεις

18, November, 2015

Σημεία Πώλησης εισιτηρίων

Οργανισμός Αστικών Συγκοινωνιών Αθηνών

18, November, 2015

Δρομολόγια Αστικών Συγκοινωνιών Αθήνας

Οργανισμός Αστικών Συγκοινωνιών Αθηνών



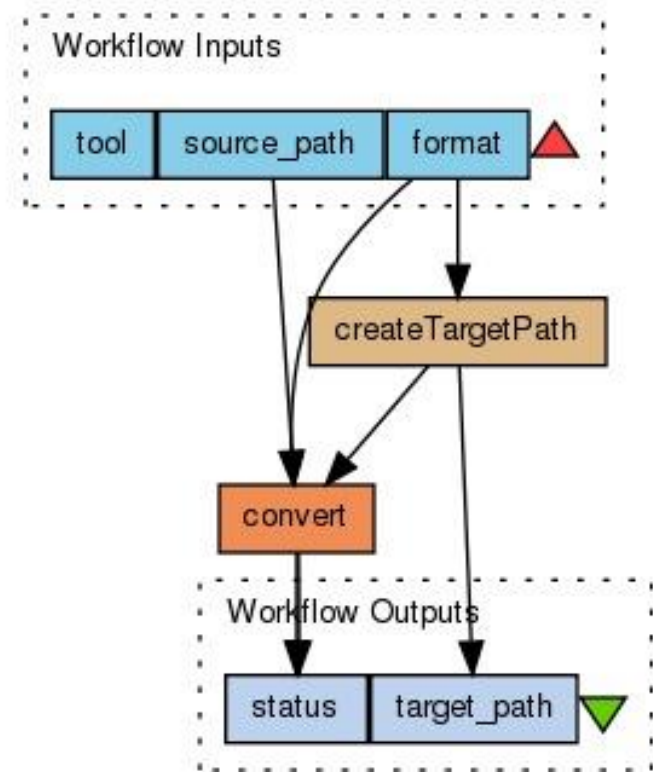
Αφήστε τα σχόλιά σας

Open Platforms



Workflow Management Systems

- Also known as Workflow Engines
- Provide the infrastructure for the set-up, execution and monitoring of a sequence of tasks
- Describe what one wants to do
 - building on certain data
 - running certain services
 - exploiting other resources ...



Πηγή: <http://www.myexperiment.org/workflows/4674.html>

Apache Taverna



- **Open source** software tool for designing and executing workflows
- Enables users to integrate diverse software components, including SOAP or REST **Web services**
- Services to be used in Taverna workflows can be discovered through **BioCatalogue** (a public, centralized and curated registry of Life Science Web services - <https://www.biocatalogue.org/>)
- Taverna workflows can be shared with other people through the **myExperiment** social web site for scientists (<http://www.myexperiment.org/home>)



Service panel

Filter: Clear

Import new services

Available services

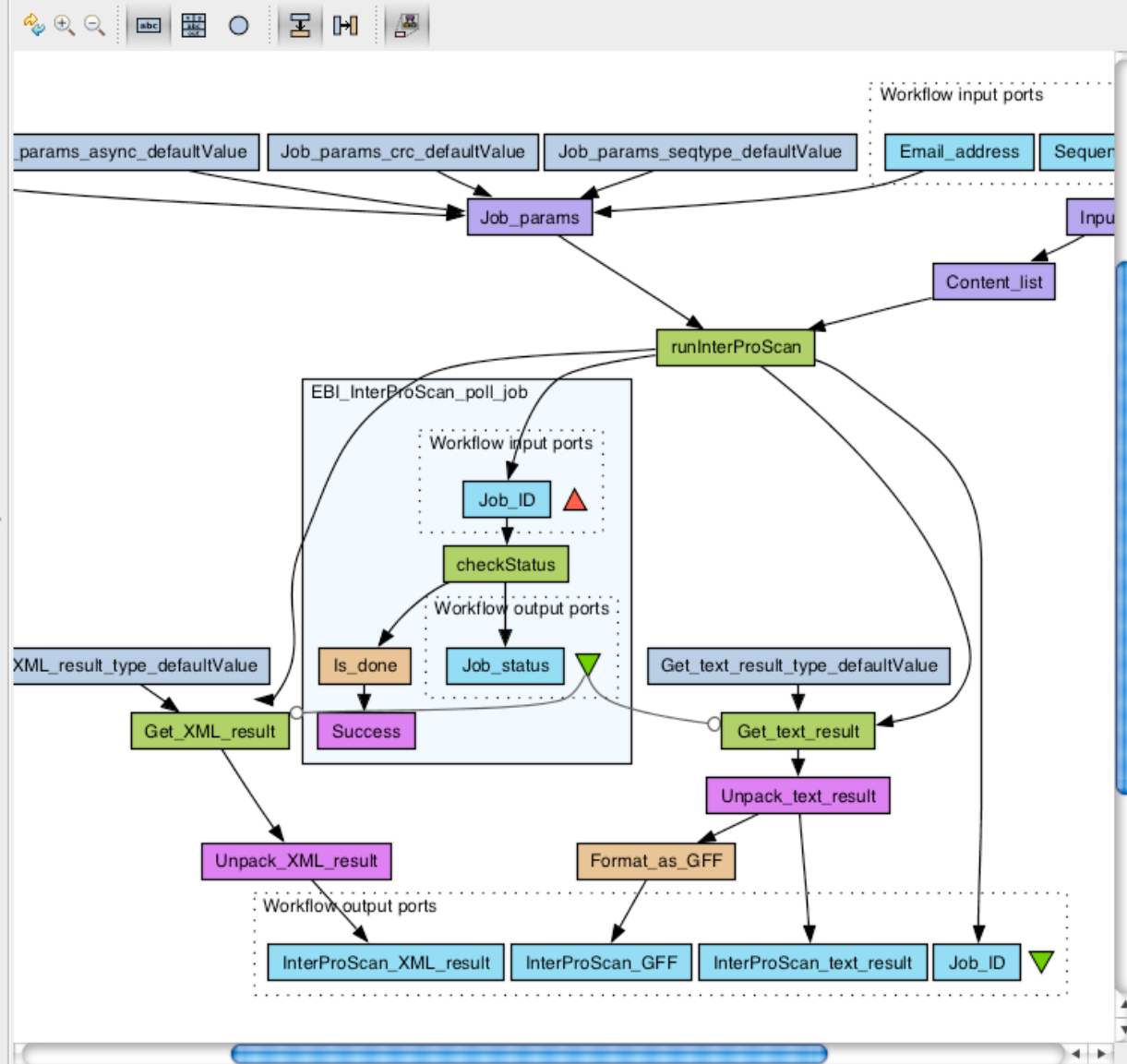
- Service templates
- Local services
- Biomart @ <http://www.biomart.org/biomart/martservice>
- Biomoby @ <http://moby.ucalgary.ca/moby/MOBY-Central.pl>
- Soaplab @ <http://www.ebi.ac.uk/soaplab/services/>
- WSDL @ <http://eutils.ncbi.nlm.nih.gov/entrez/eutils/soap/eutil>
- WSDL @ <http://soap.bind.ca/wsd1/bind.wsd1>
- WSDL @ <http://soap.genome.jp/KEGG.wsd1>
- WSDL @ <http://www.ebi.ac.uk/ws/services/urn:Dbfetch?wsdl>
- WSDL @ <http://www.ebi.ac.uk/xembl/XEMBL.wsd1>

Workflow explorer Details Validation report

EBI_InterProScan

- Workflow input ports
 - Email_address
 - Sequence_or_ID
- Workflow output ports
 - InterProScan_GFF
 - InterProScan_text_result
 - InterProScan_XML_result
 - Job_ID
- Services
 - Content_list
 - WSArrayOfData
 - output
 - EBI_InterProScan_poll_job
 - Job_ID
 - Job_status
 - Format_as_GFF
 - interproscan_text

Workflow diagram



The BioCatalogue: providing a curated catalogue of life science Web services

BioCatalogue currently has [1186 services](#), [250 service providers](#) and [828 members](#) [i](#)

[i](#) Helpful Links

[Getting started with the Catalogue](#)

[General info on Web services](#)

[Best practice guide for developing Web services](#)

[SSI's top tips on creating Web services](#)

[Executing Web services](#)

[Turn your command line application\(s\) into Web services](#)

[i](#) Citing BioCatalogue

Bhagat, J., Tanoh, F., Nzuobontane, E., Laurent, T., Orłowski, J., Roos, M., Wolstencroft, K., Aleksejevs, S., Stevens, R., Pettifer, S., Lopez, R., Goble, C.A.

BioCatalogue: a universal catalogue of web services for the life sciences

Nucl. Acids Res. (2010) 38: 689-694
doi:10.1093/nar/gkq394

"Web Services are hard to find"

DISCOVER

- Find the right Web Service
- Powerful search and filtering
- Information from providers and community

[More info](#)

"My Web Services are not visible"

REGISTER

- Easily register Web Services
- Instantly available to everyone
- Providers can advertise, describe and monitor their Services

[More info](#)

"Web Services are poorly described"

ANNOTATE

- Anyone can describe and annotate
- Ongoing expert curation
- Social curation by the community

[More info](#)

"Web Services are volatile"

MONITOR

- Services change and get outdated
- BioCatalogue monitors Services
- Monitors availability and reliability

[More info](#)

[RSS](#) Site Announcements [RSS](#)

[ELIXIR Tools registry 'community paper' published in IAR](#)

By [Niall Beard](#) (9 days ago)

[BiodiversityCatalogue presents development of Web Map Service support at TDWG](#)

By [Niall Beard](#) (about 1 month ago)

[Scheduled Maintenance](#)

By [Niall Beard](#) (over 1 year ago)

[BioCatalogue Rails 3 Upgrade](#)

By [Aleksandra Nenadic](#) (almost 2 years ago)

[Soaplab EMBOSS services end of life: 31st January 2013](#)

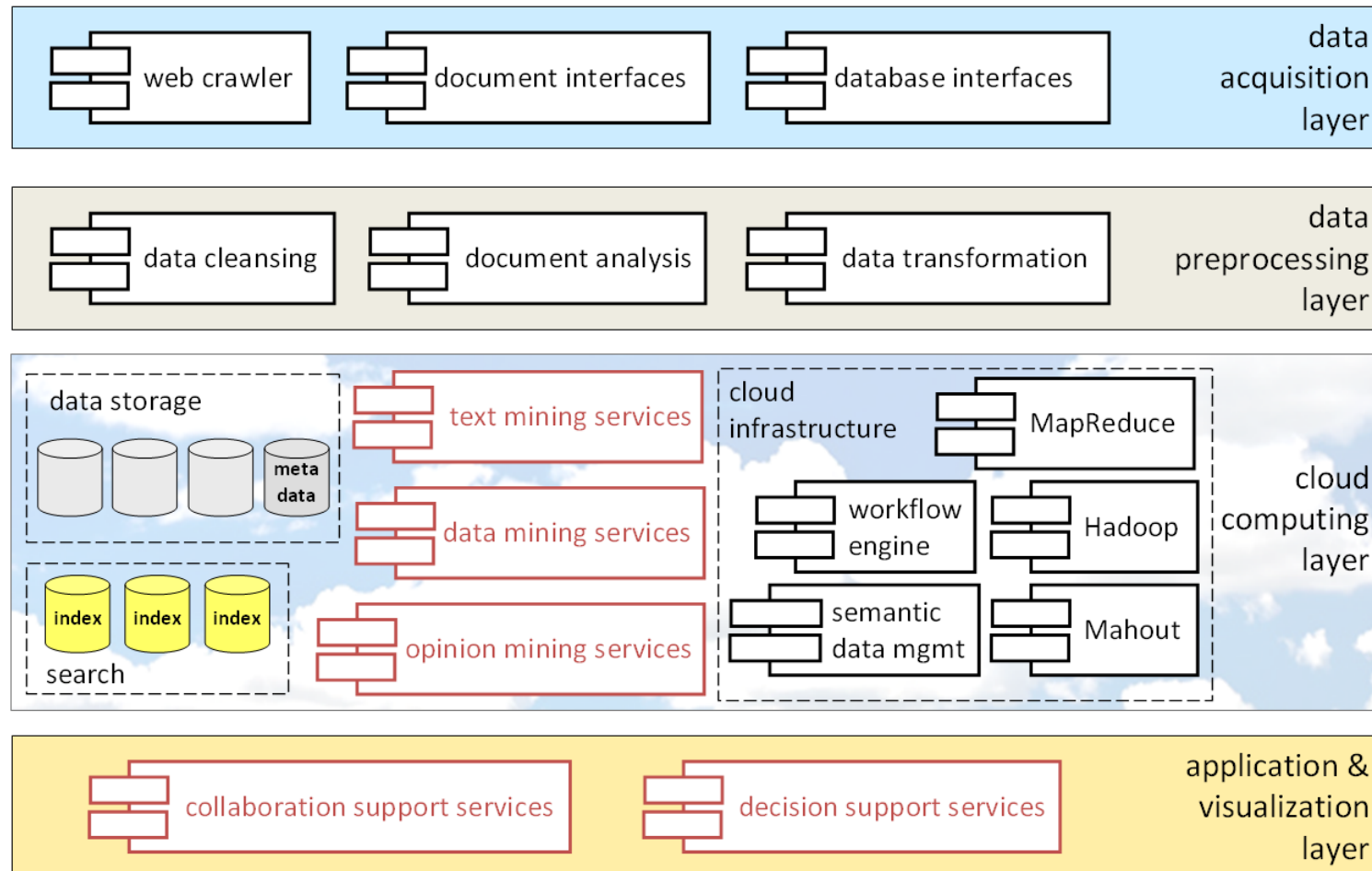
By [Robert Haines](#) (about 3 years ago)

[More](#)

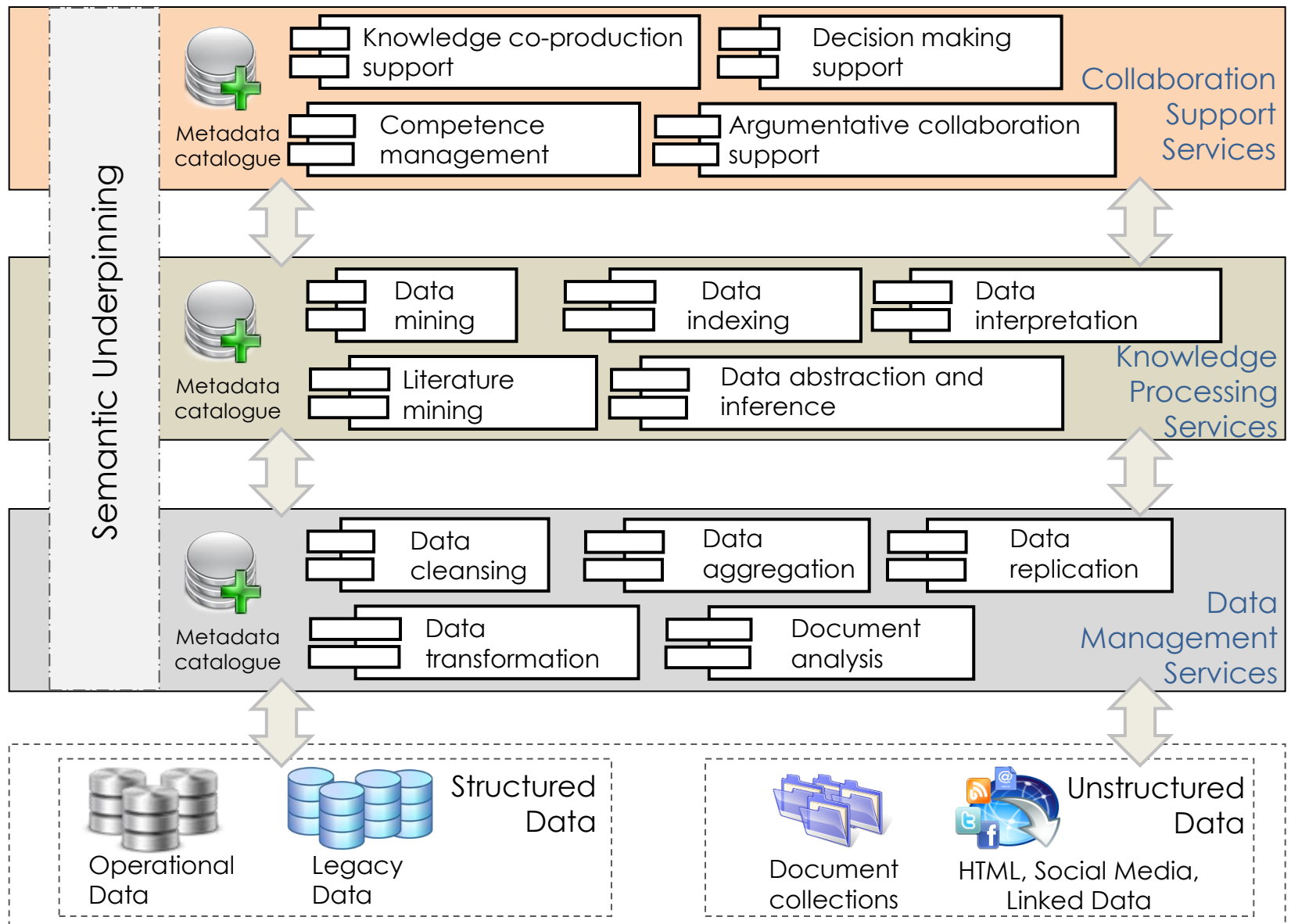
The Dicode architecture



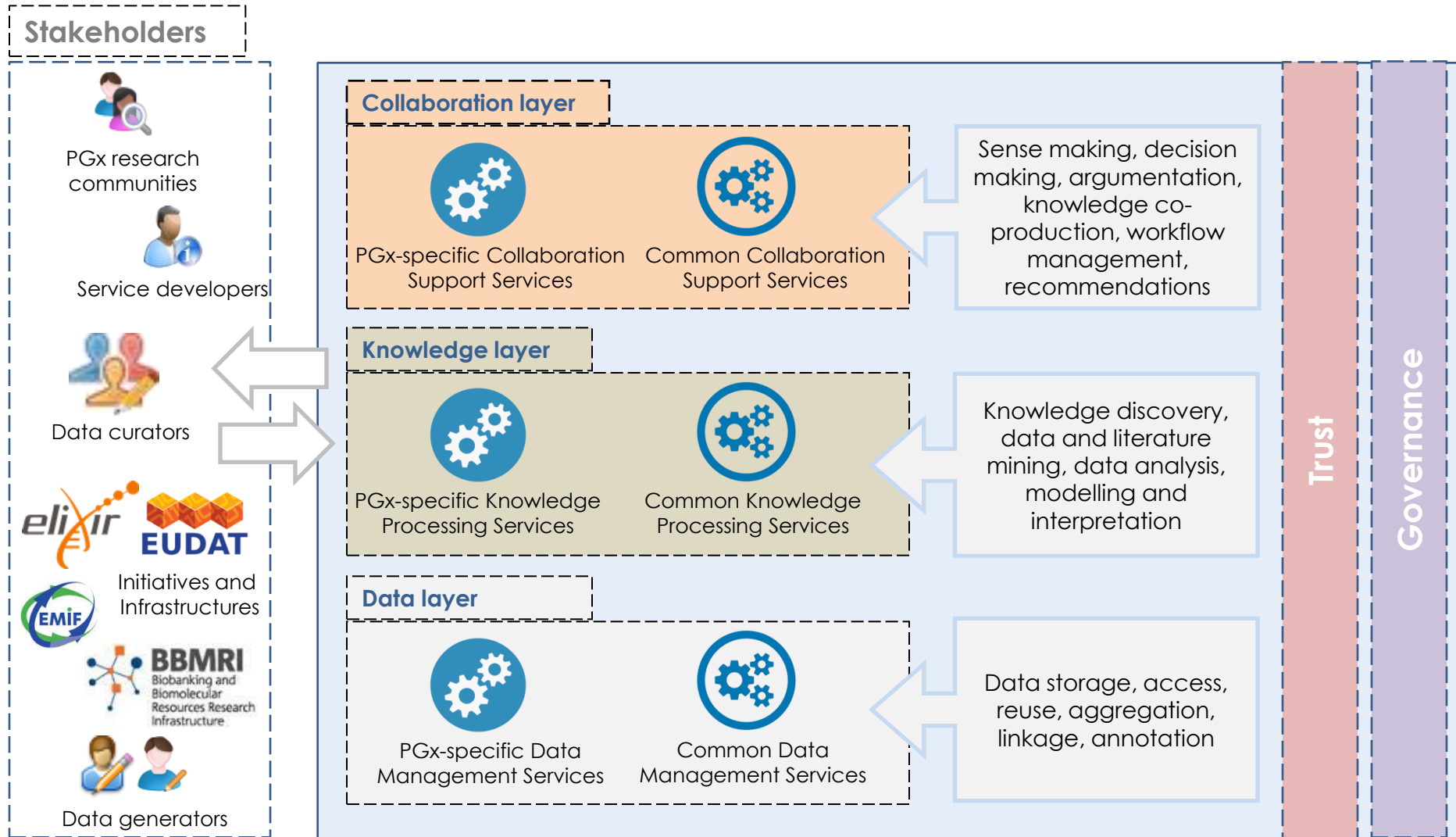
data sources
(documents,
databases,
world wide web)



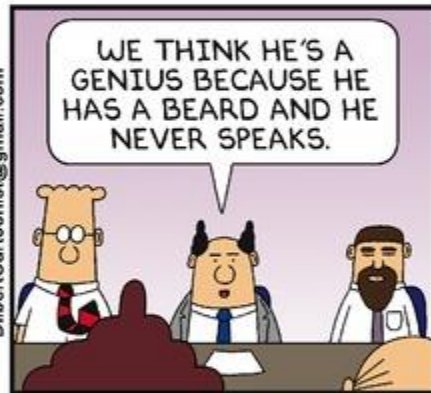
Towards an Open Collaboration Platform



An Open Working Environment for PGx research



ΑΝΤΙ ΕΠΙΛΟΓΟΥ



Βιβλιογραφία

- Etienne and Beverly Wenger-Trayner, “Introduction to communities of practice: A brief overview of the concept and its uses” → <http://wenger-trayner.com/introduction-to-communities-of-practice/>
- The Royal Society, “Science as an open enterprise”, Royal Society Science Policy Centre report, June 2012 → http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf
- Computing Community Consortium - Computing Research Association, “Challenges and Opportunities with Big Data: A community white paper developed by leading researchers across the United States”, White Paper, Feb 2012 → <http://www.cra.org/ccc/files/docs/init/bigdatawhitepaper.pdf>
- Gartner Group, “Pattern-Based Strategy: Getting Value from Big Data”, press release, July 2011 → <http://www.gartner.com/it/page.jsp?id=1731916>
- T. Heath and C. Bizer, “Linked Data: Evolving the Web into a Global Data Space” (1st ed.), Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136, Morgan & Claypool, 2011 → <http://linkeddatabook.com/book>
- J. Lehmann, R. Isele, M. Jakob, A. Jentzsch, D. Kontokostas, P. Mendes, S. Hellmann, M. Morsey, P. van Kleef, S. Auer and C. Bizer, “Dbpedia: A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia”, *Semantic Web Journal*, 6(2), pp. 167–195, 2015 → http://svn.aksw.org/papers/2013/SWJ_DBpedia/public.pdf

Σημείωμα αδειοδότησης

Το παρόν υλικό διατίθεται με τους όρους της άδειας χρήσης Creative Commons «Αναφορά Δημιουργού - Μη Εμπορική Χρήση - Παρόμοια Διανομή» 4.0 ή μεταγενέστερη, Διεθνής Έκδοση [<http://creativecommons.org/licenses/by-nc-sa/4.0/>]

