ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ
UNIVERSITY OF PATRAS

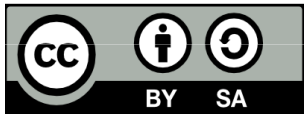
ΑΝΟΙΚΤΑ ακαδημαϊκά μαθήματα ΠΠ

Μελέτη Περιπτώσεων στη Λήψη Αποφάσεων

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Σημείωμα Αδειοδότησης

- Το παρόν εκπαιδευτικό υλικό υπόκειται σε άδειες χρήσης Creative Commons.
- Για εκπαιδευτικό υλικό, όπως εικόνες, που υπόκειται σε άλλου τύπου άδειας χρήσης, η άδεια χρήσης αναφέρεται ρητώς.



Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στο πλαίσιο του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Πατρών**» έχει χρηματοδοτήσει μόνο την αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.

Mastering Data-Intensive Collaboration and Decision Making

Nikos Karacapilidis
University of Patras & CTI
nikos@mech.upatras.gr

Lecture
May 15, 2015

Dicode (dicode-project.eu) is funded by the European Commission, Information Society and Media Directorate General, under the FP7 Cooperation Programme

collaborative decision making

Collaboration leads to transparency, openness and better decision making

Basic characteristics

- Information overload & cognitive overhead
- Diverse social behavior
  - structures, relationships and interactions
- Situational differences
  - diverse collaboration modes and paradigms
- Expression of tacit knowledge
- Integration of legacy resources

→ Data processing and decision making support

Issues to be addressed

- Use of communication and information processing technology to make collaboration more efficient and effective
- Work structuring in order to improve coordination
- (Semi-) Automation of data processing
  - especially in data intensive situations
- User/group modeling
- Visualization
- Argumentation & reasoning mechanisms
  - rules and procedures for achieving consistency

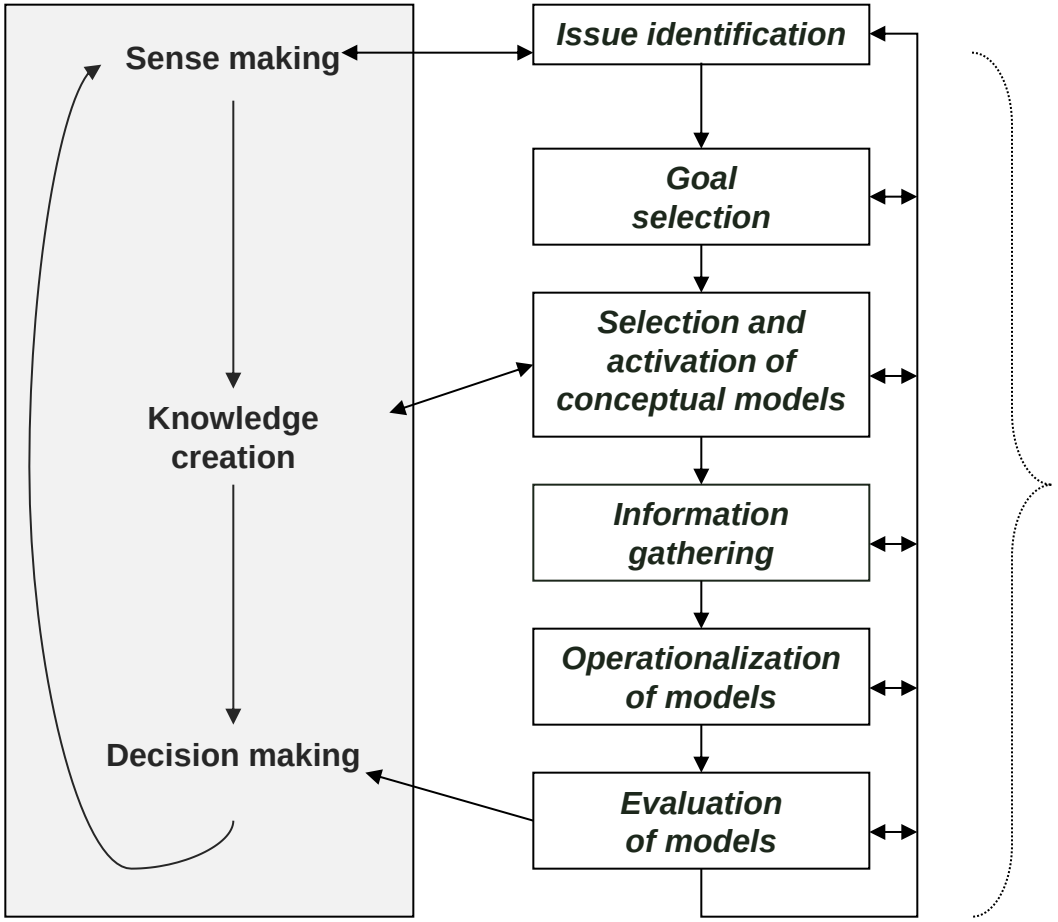
Services required

- Information services
  - Information search and retrieval, interoperability, transformation, data mining, ...
- Knowledge Management services
  - Knowledge management, metadata, ontologies, annotation & tagging, opinion mining, …
- Collaboration & DM services
  - Conducting of debates, argumentation, negotiations, handling of conflicts, sense-making, decision making, awareness, …

Communities of practice (1/2)

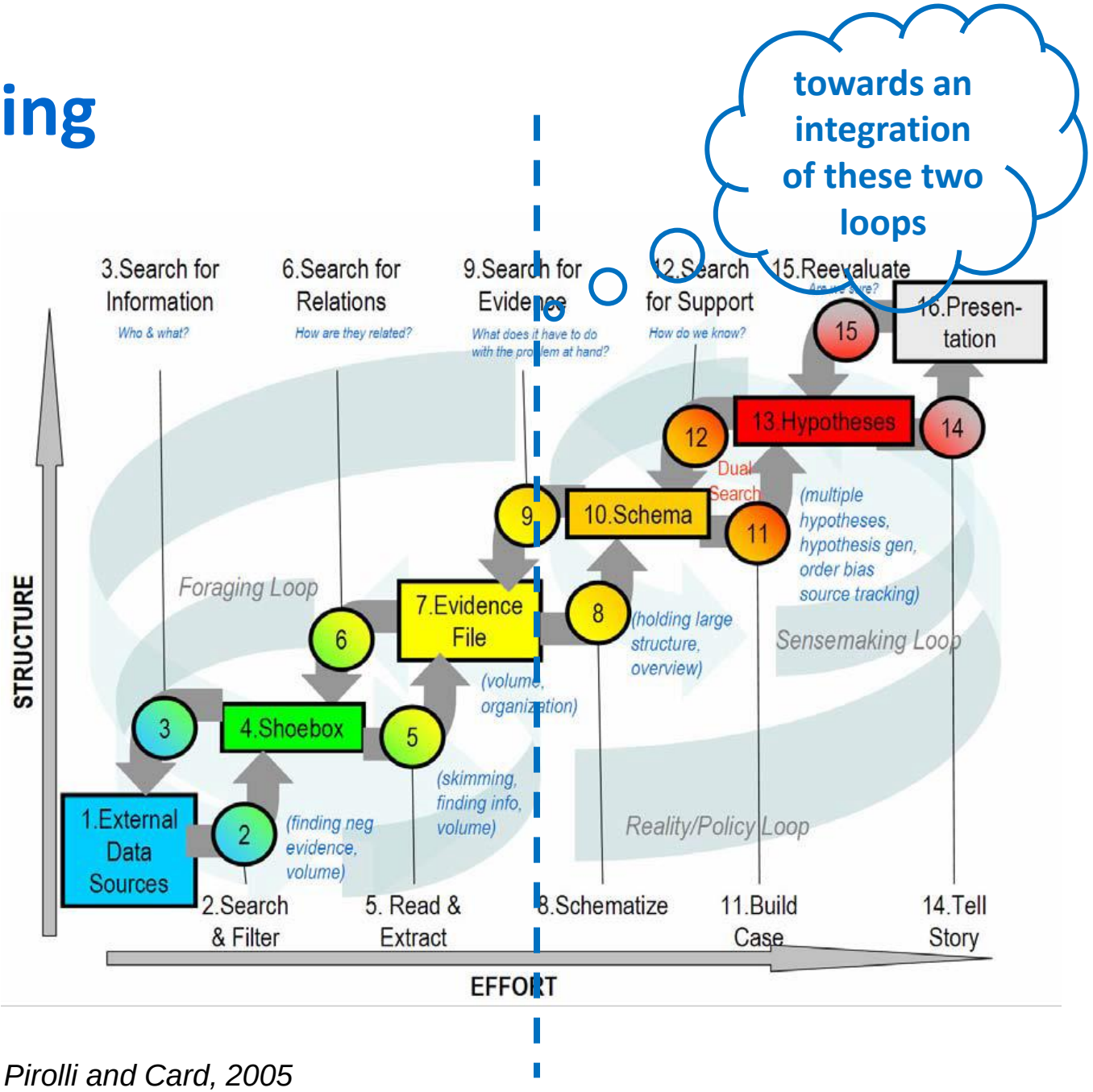
- Communities of practice (CoPs) are formed for resolving poorly structured problems, over which a plurality of views holds
  - Diverse group of specialists exchange views through elaborated discussion
  - The objective is to create and use problem-specific knowledge through the social interaction of different sources of codified and tacit knowledge

Communities of practice (2/2)



Sense-making

- A process of transformation of information into a knowledge product, including:
  - a **foraging loop** that involves seeking, filtering, and extracting information into schemas; and
  - a **sensemaking loop** that involves iterative development of a mental model from the schemas that best fit the evidence.



Pirolli and Card, 2005

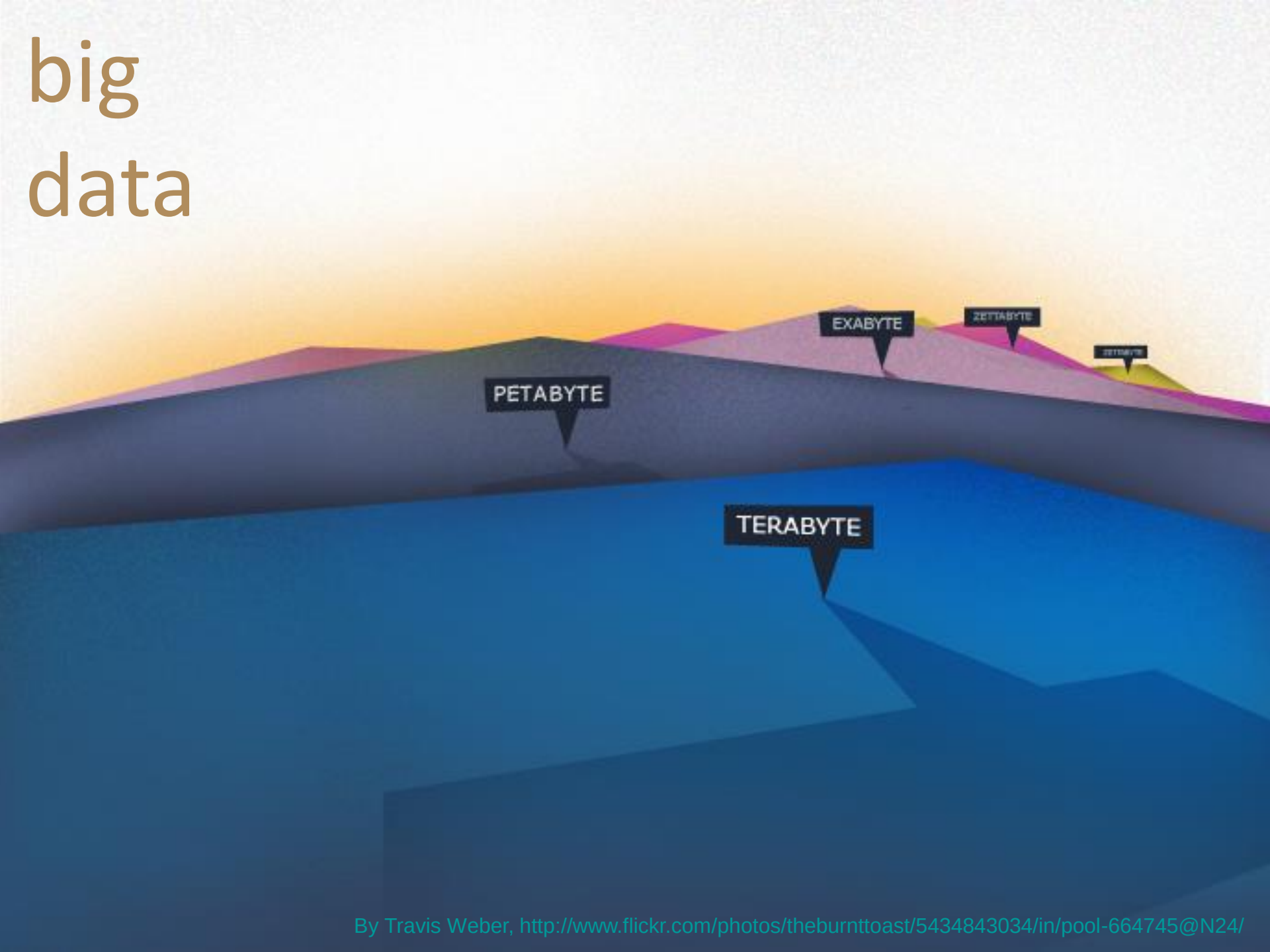
State-of-the-art and beyond (1/2)

- Collaboration support
  - Current tools are "information islands" → increased interoperability and synergy with third party tools
  - Web 2.0 collaboration tools are rather passive media → intelligent reasoning services to actively and meaningfully support collaboration
  - Web 2.0 collaboration tools cope poorly with voluminous and complex data → advanced decision making support services; building on the synergy of human and machine reasoning

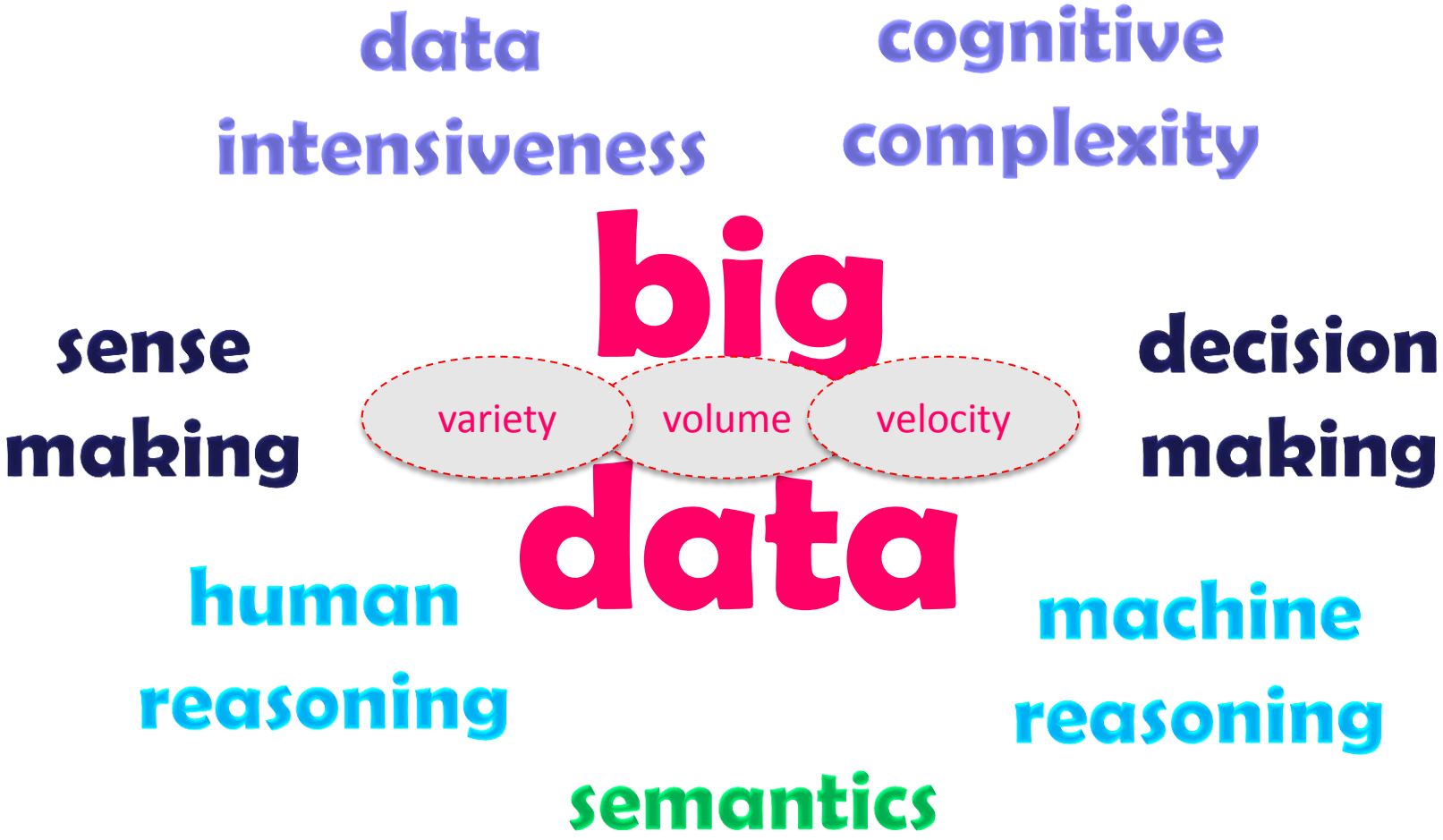
State-of-the-art and beyond (2/2)

- Decision Making support
  - Problem-centric view → emphasis on human-centric view
  - No thorough exploitation of underlying knowledge → knowledge-based decision-making view; building on the synergy of human and machine reasoning
  - Little attention to dialoguing and argumentation → argumentation-based reasoning mechanisms

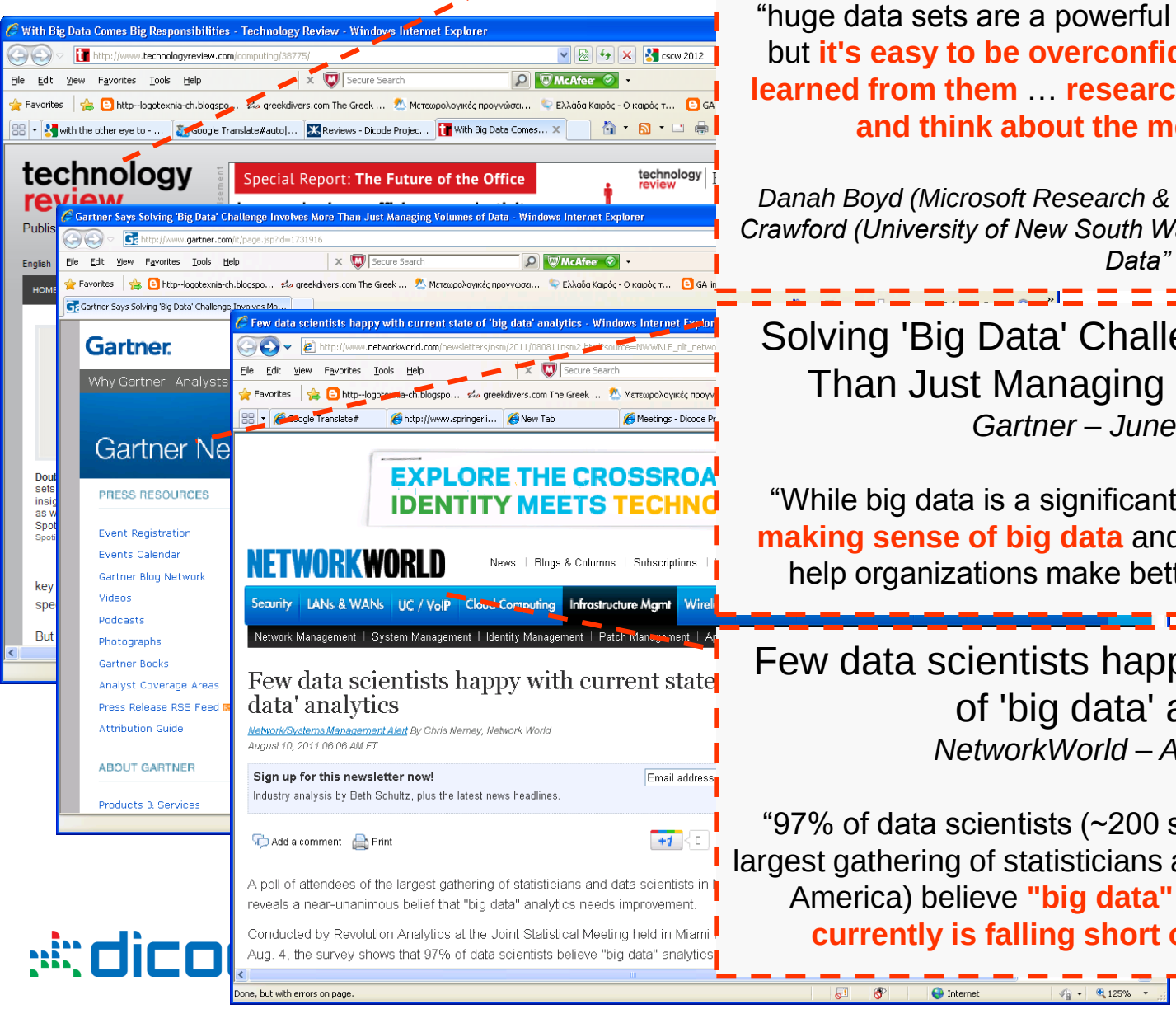
big data



By Travis Weber, http://www.flickr.com/photos/theburnttoast/5434843034/in/pool-664745@N24/


data intensiveness
cognitive complexity
big data
variety
volume
velocity
sense making
decision making
human reasoning
machine reasoning
semantics

Big Data criticism



With Big Data Comes Big Responsibilities

Technology Review – by MIT – Oct 5, 2011

"huge data sets are a powerful new tool for researchers, but **it's easy to be overconfident about what can be learned from them** … **researchers need to slow down and think about the methods they use**"

Danah Boyd (Microsoft Research & Harvard University) and Kate Crawford (University of New South Wales): "Six Provocations for Big Data"

Solving 'Big Data' Challenge Involves More Than Just Managing Volumes of Data

Gartner – June 27, 2011

"While big data is a significant issue, **the real issue is making sense of big data** and finding patterns in it that help organizations make better business decisions"

Few data scientists happy with current state of 'big data' analytics

NetworkWorld – Aug 10, 2011

"97% of data scientists (~200 scientists surveyed in the largest gathering of statisticians and data scientists in North America) believe **"big data" analytics technology currently is falling short of enterprise needs**"

The Big Data fallacy

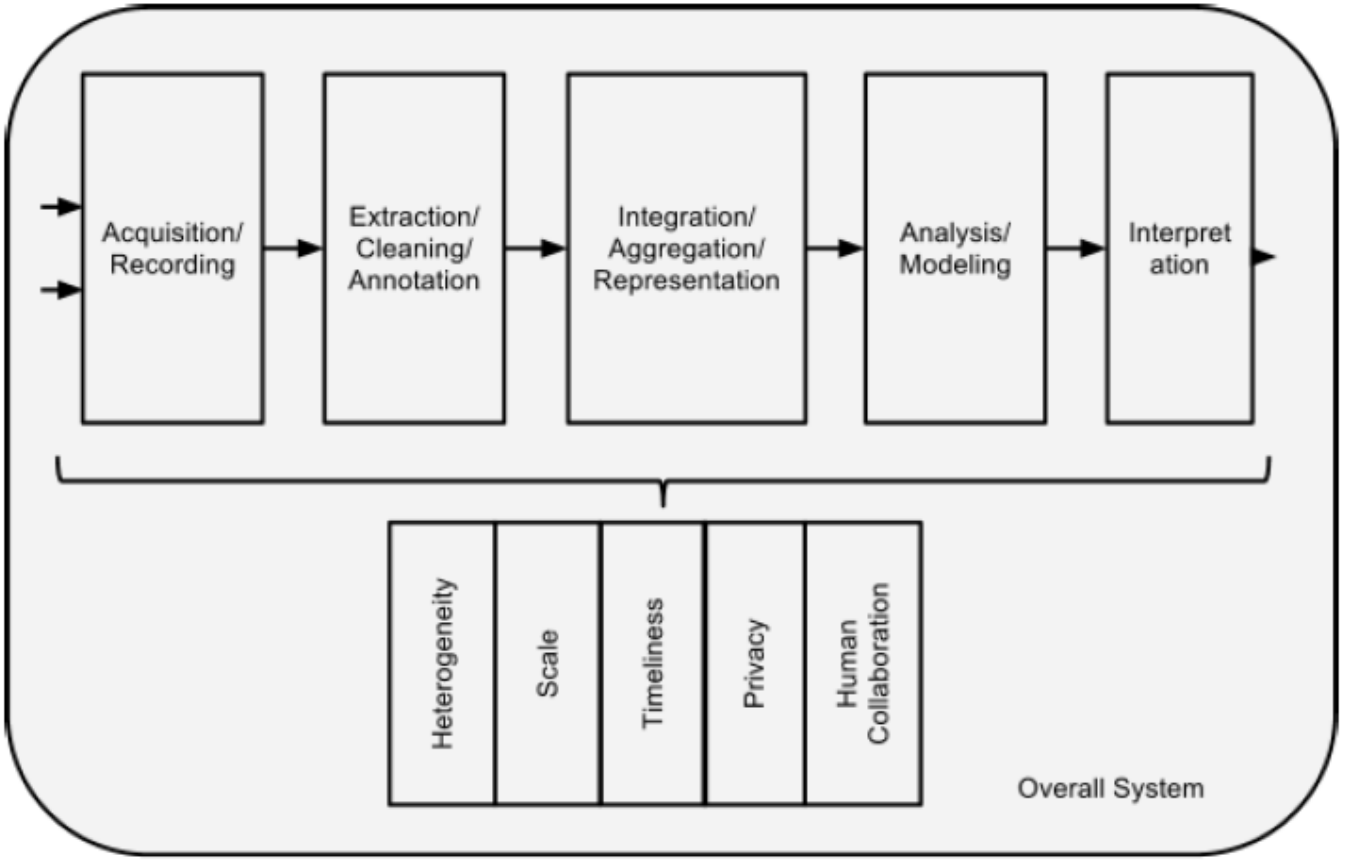
- More data doesn't mean you will get "*proportionately*" more information
  - In fact, the more data you have, the less information you gain as a proportion of
- The value of big data is often ov
  - Its value is in the information an provide
  - Information is only the n dant portions of the data, which is a ti minishing fraction of the overall data vo

The data-information inequality:
information << data

Source: http://lithosphere.lithium.com/t5/Science-of-Social-blog/The-Big-Data-Fallacy-Data-Information/ba-p/59250

Big Data white paper - Feb 2012 (1/2)

Big Data analysis pipeline (major steps and needs that make them challenging)



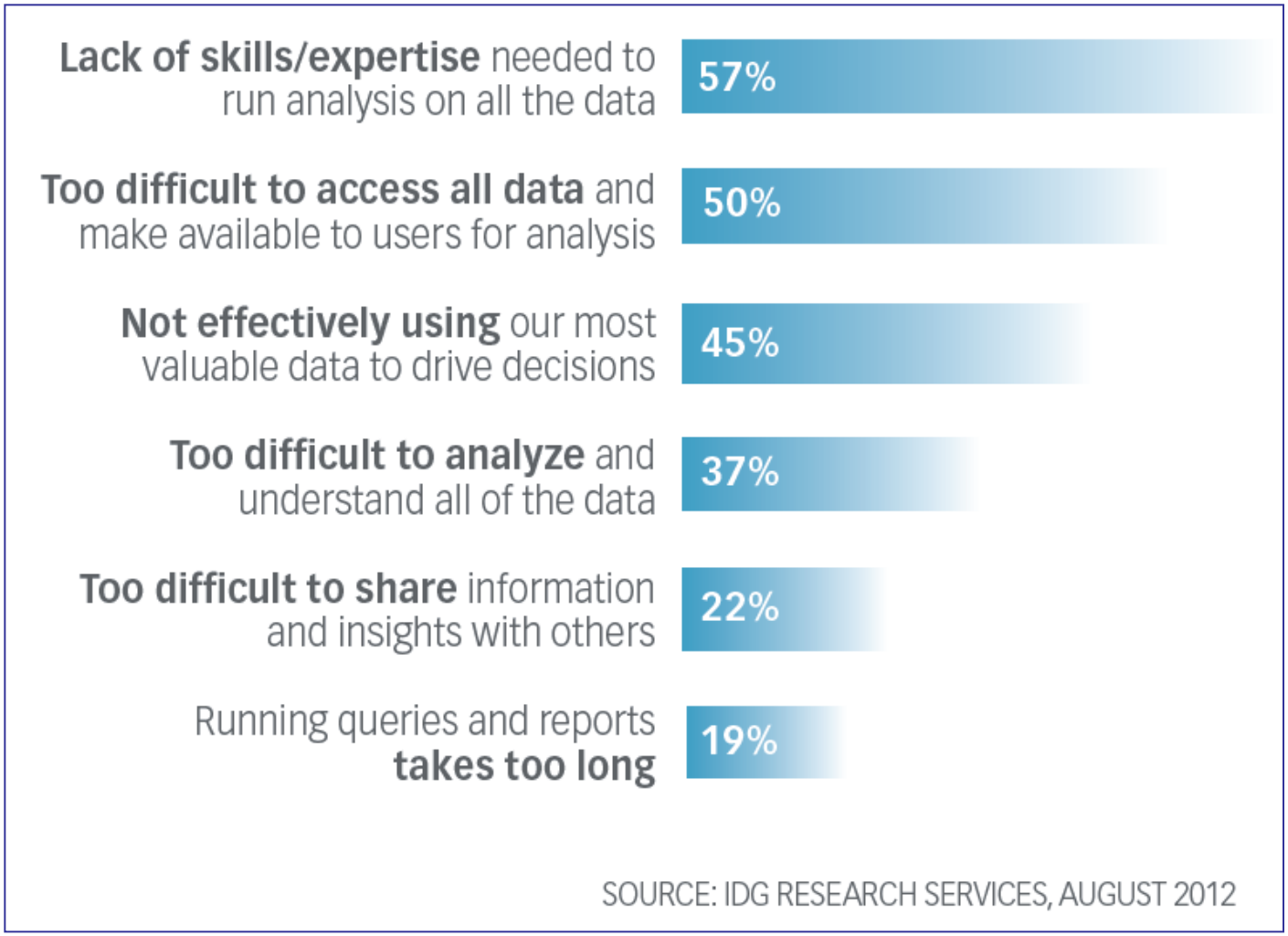
Source: http://cra.org/ccc/docs/init/bigdatawhitepaper.pdf

Big Data white paper - Feb 2012 (2/2)

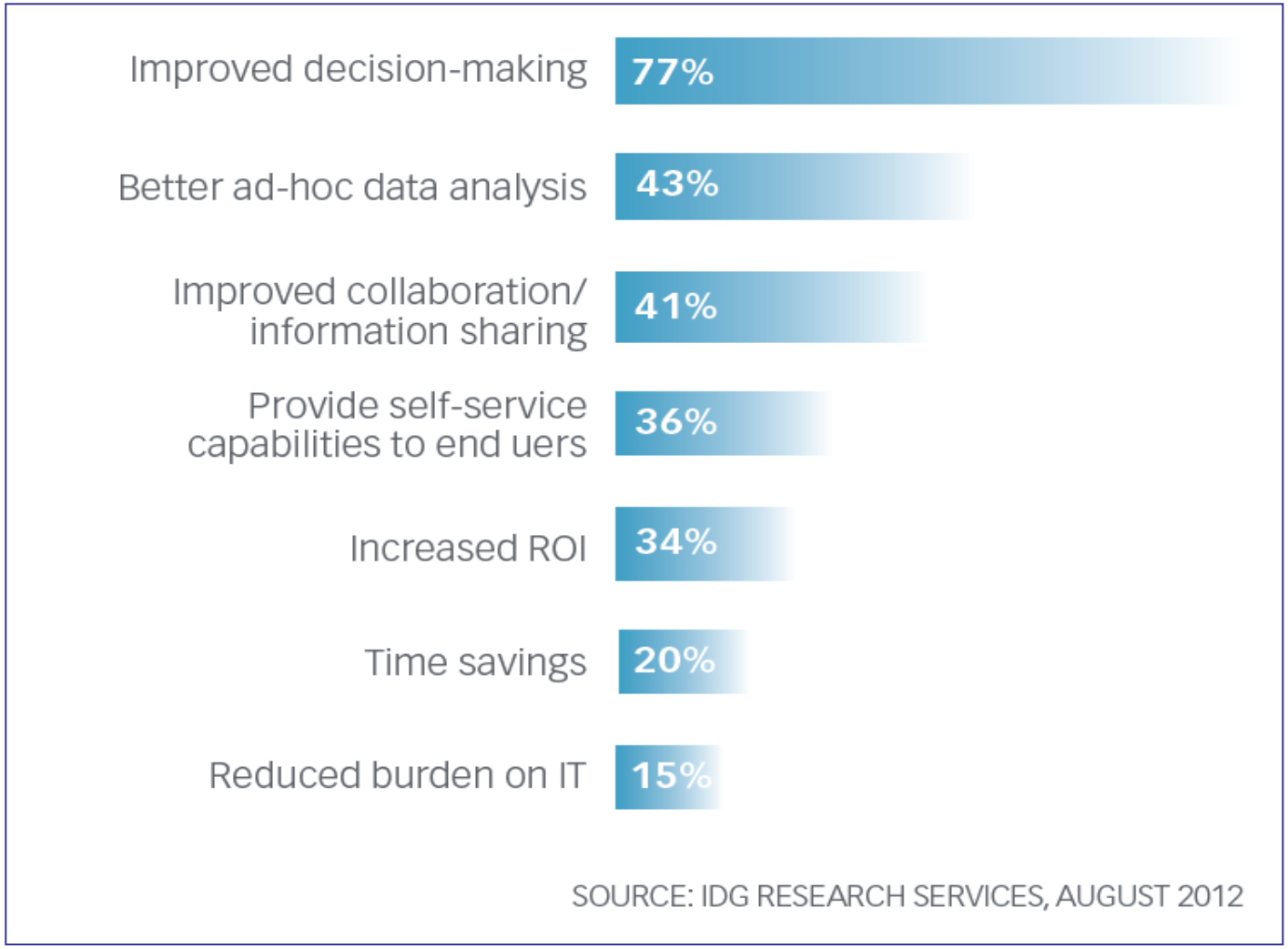
- "In spite of the tremendous advances made in computational analysis, there remain many patterns that humans can easily detect but computer algorithms have a hard time finding"
- "Ideally, analytics for Big Data will not be all computational – rather it will be designed explicitly to have a human in the loop"
- "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".
- "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. Such supplementary information is called the provenance of the (result) data".
- "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".

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 challenges



Top benefits of data visualization



the Dicode roadmap

http://www.flickr.com/photos/piero/151739521/

Dicode's main goal

- What
  - facilitate and augment collaboration and decision making in data-intensive and cognitively-complex settings
- How
  - by exploiting and building on the synergy of human and machine reasoning
  - by deepening our insights on the proper exploitation of Big Data and related technologies

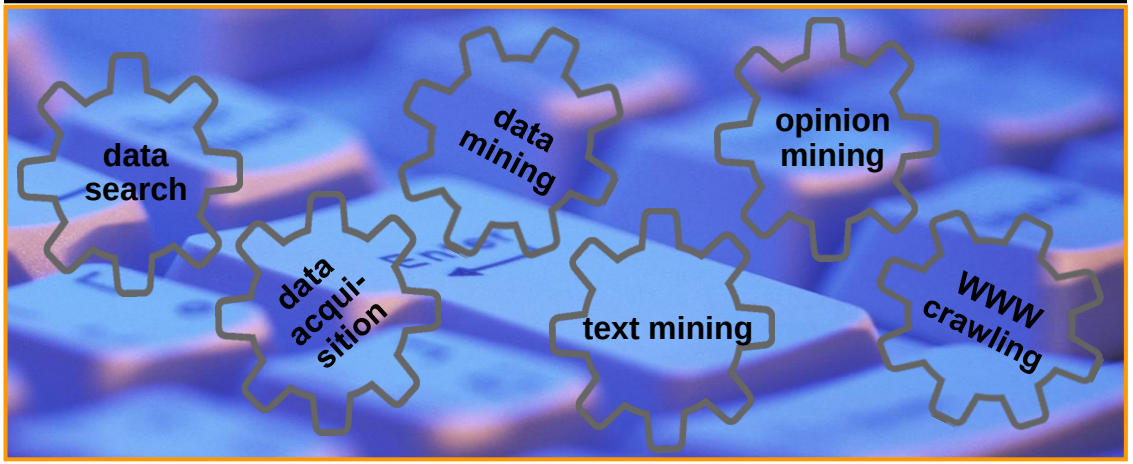
Synergy of human & machine reasoning

proper orchestration and exploitation of each side's strengths

Collaboration Support

Decision Making Support

Scalable High-Performance Data Mining

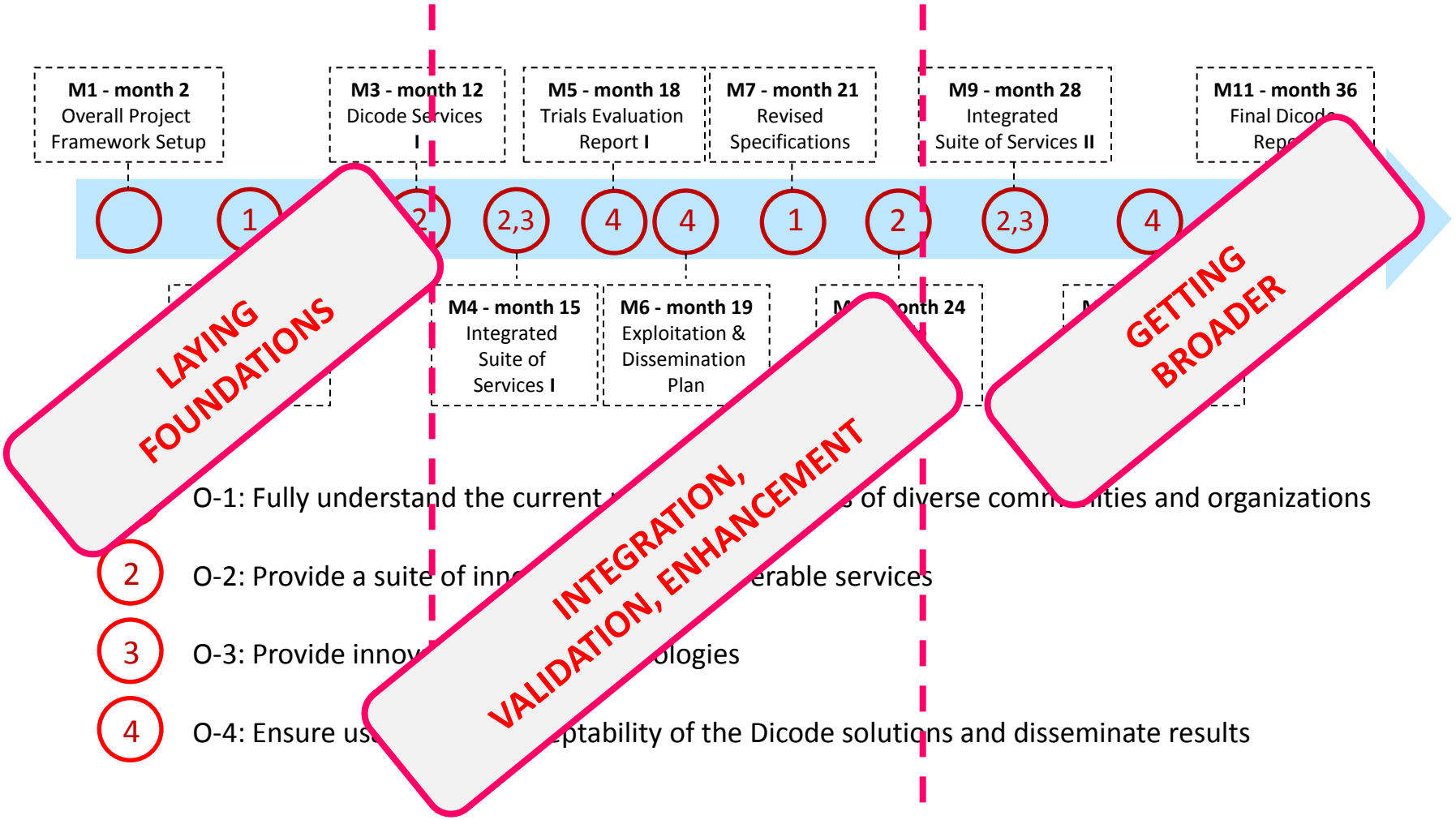


dicode

Project objectives

- O-1: To fully understand the current practices and needs of diverse communities and organizations
- O-2: To provide a suite of innovative, adaptive and interoperable services
  - O-2.1: Data acquisition services
  - O-2.2: Data pre-processing services
  - O-2.3: Data mining services
  - O-2.4: Collaboration support services
  - O-2.5: Decision making support services
- O-3: To provide innovative work methodologies
- O-4: To ensure usability and acceptability of the above services & work methodologies and disseminate the project's results

Objectives & milestones



Overall strategy

- Evolutionary approach
  - Stakeholders are actively engaged throughout the project;
  - Incremental development (operational prototype versions in month 12; enhanced versions in month 24; final versions in month 33);
  - User requirements to be refined through testing (from all use cases);
  - An operational integrated suite of services to be early available for trials and proof-of concept purposes.

- Two main phases
  - Phase I (months 1-18): reqs and specs are produced, operational versions of the Dicode services are developed and integrated, innovative work methodologies are sketched, and first feedback is collected;
  - Phase II (months 19-36): specs and overall conceptual framework is revised, Dicode services and integrated suite offer advanced capabilities, work methodologies turn to best practices and innovative work guidelines, Dicode outcomes continue to be thoroughly tested, while the final evaluation via use cases and the overall project's evaluation take place.

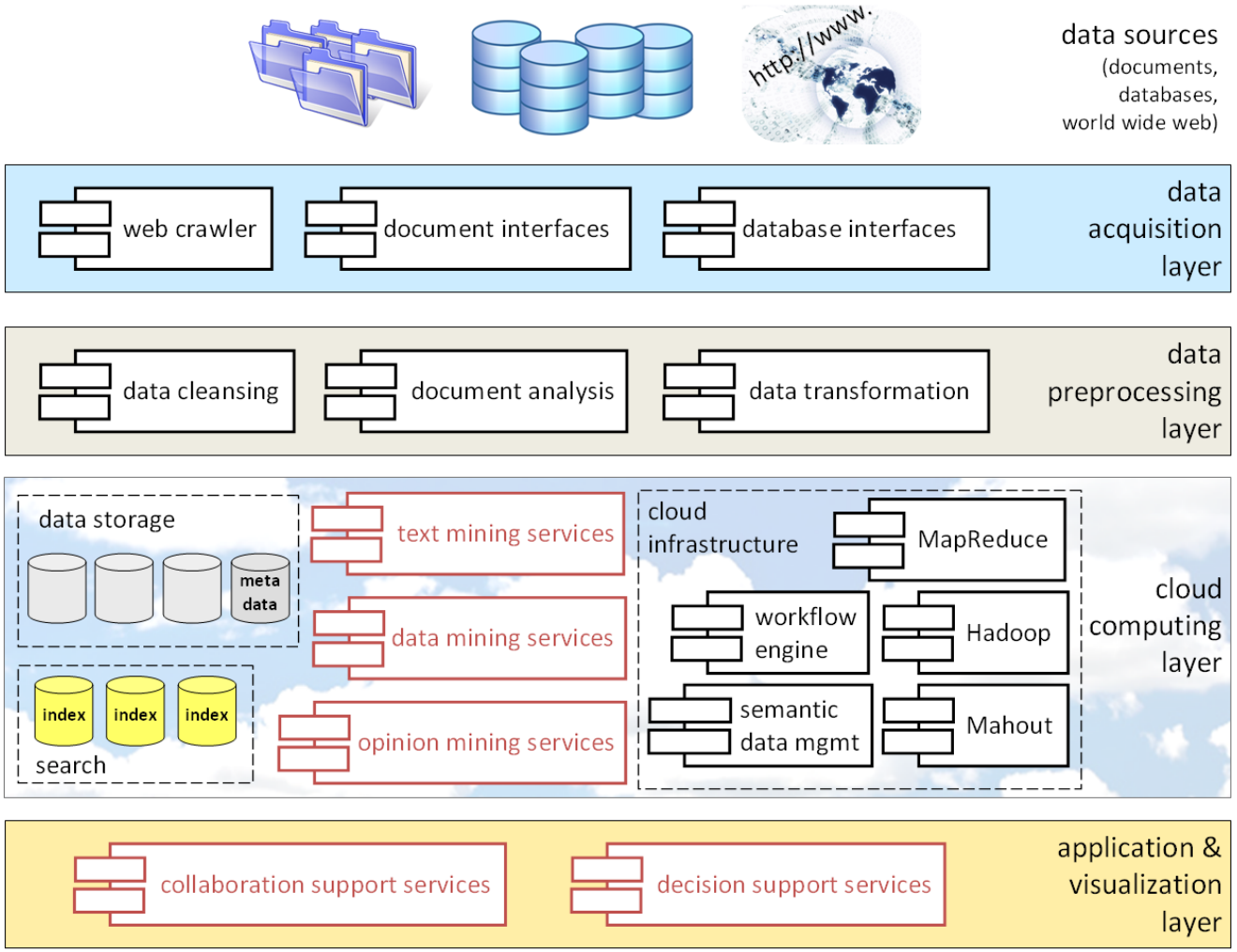
Use cases

- Carefully chosen to:
  - address clearly established problems, widely recognized in industry and academia;
  - cover the full range of features and functionalities of the project, while representing alternative collaboration and decision making paradigms
- UC1: Clinico-Genomic Research Assimilator
  - collaboratively explore, evaluate, disseminate and diffuse scientific findings and results
- UC2: Trial of Clinical Treatment Effects
  - making clinical decisions in drug trials by combining datasets from patient results and different scan modalities to reveal the effectiveness of a drug within a trial
- UC3: Opinion mining from Unstructured Web 2.0 Data
  - analysis of the voluminous amount of unstructured information existing on the Web; data primarily obtained from spidering the most popular social Web sites

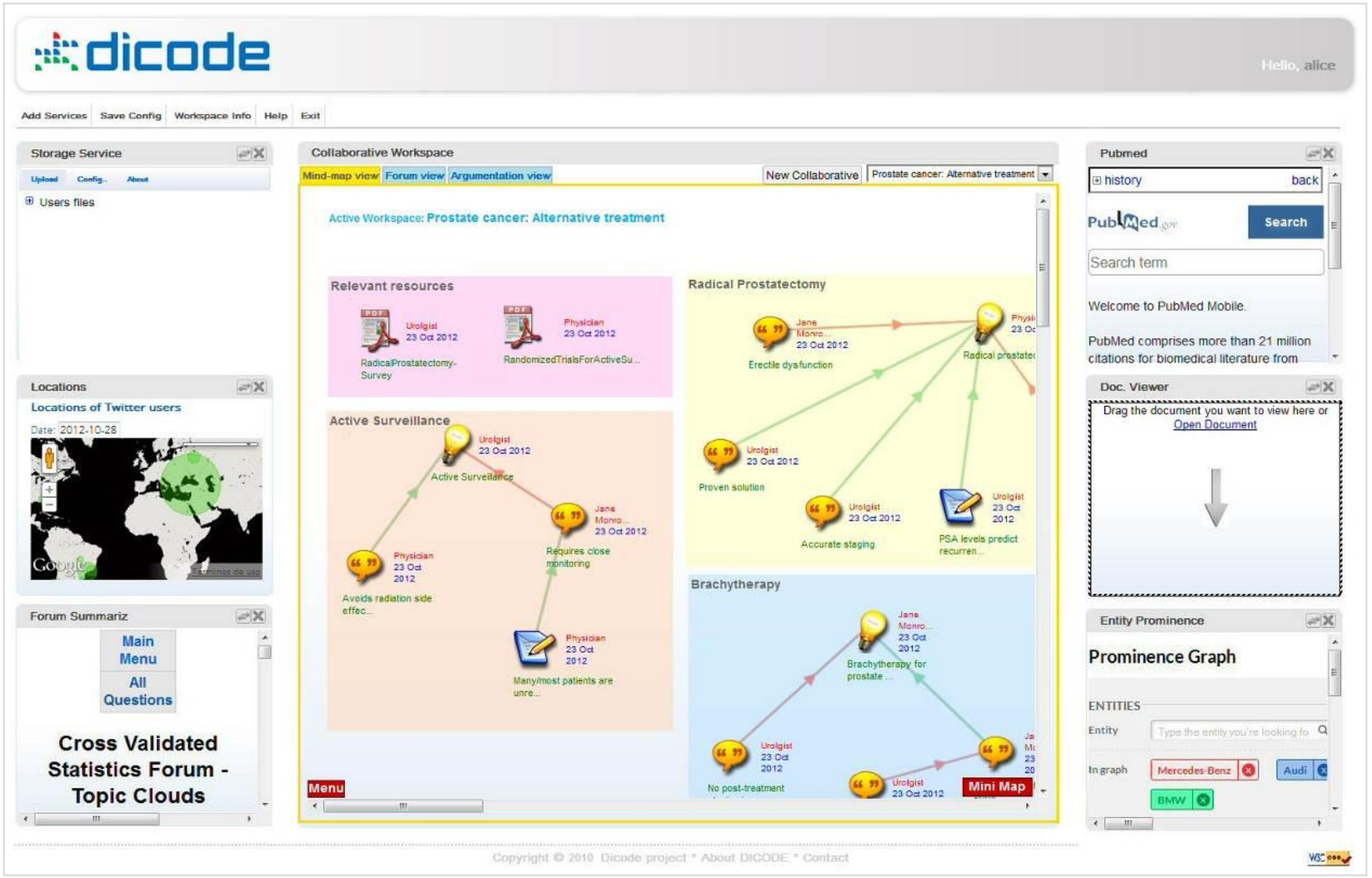
Achievements

http://www.flickr.com/photos/44704258@N06/5252531705/

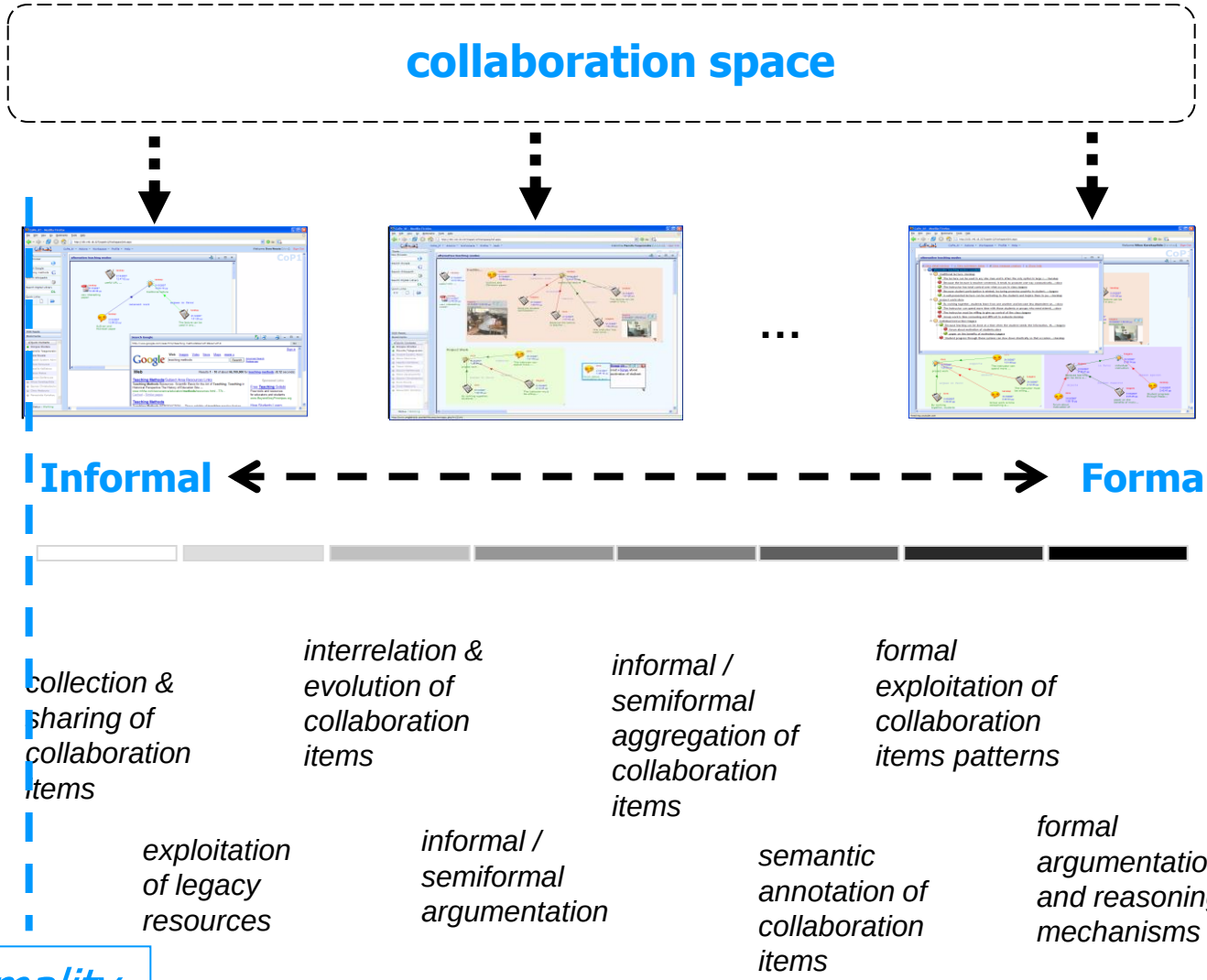
The Dicode architecture



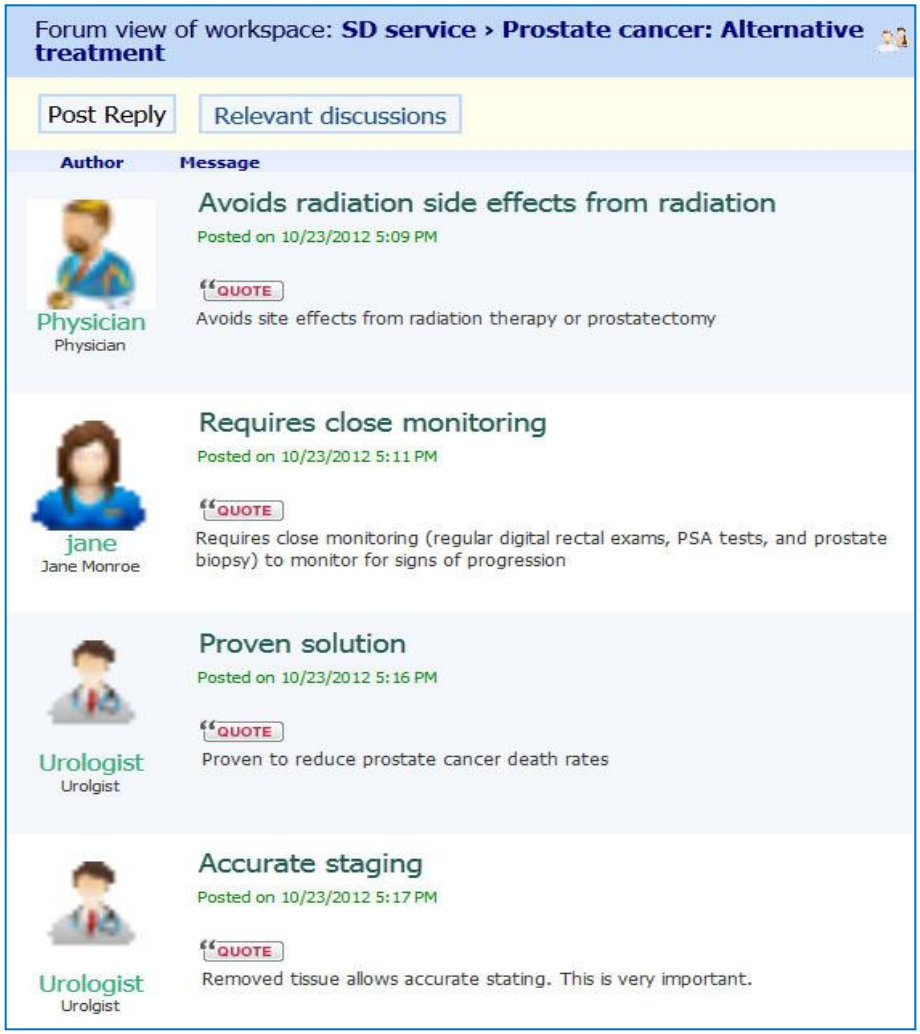
The Dicode Workbench



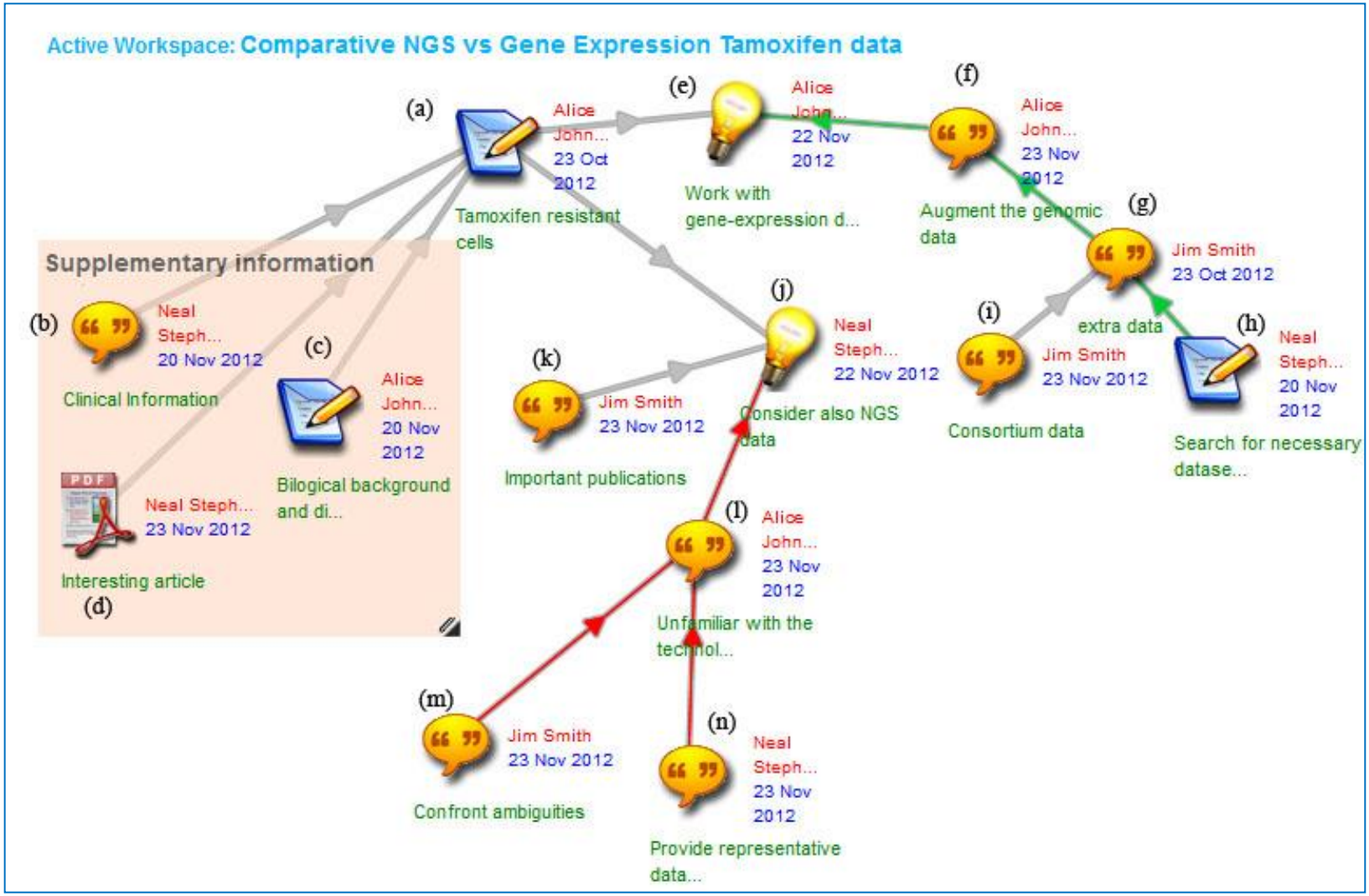
Collaboration workspaces



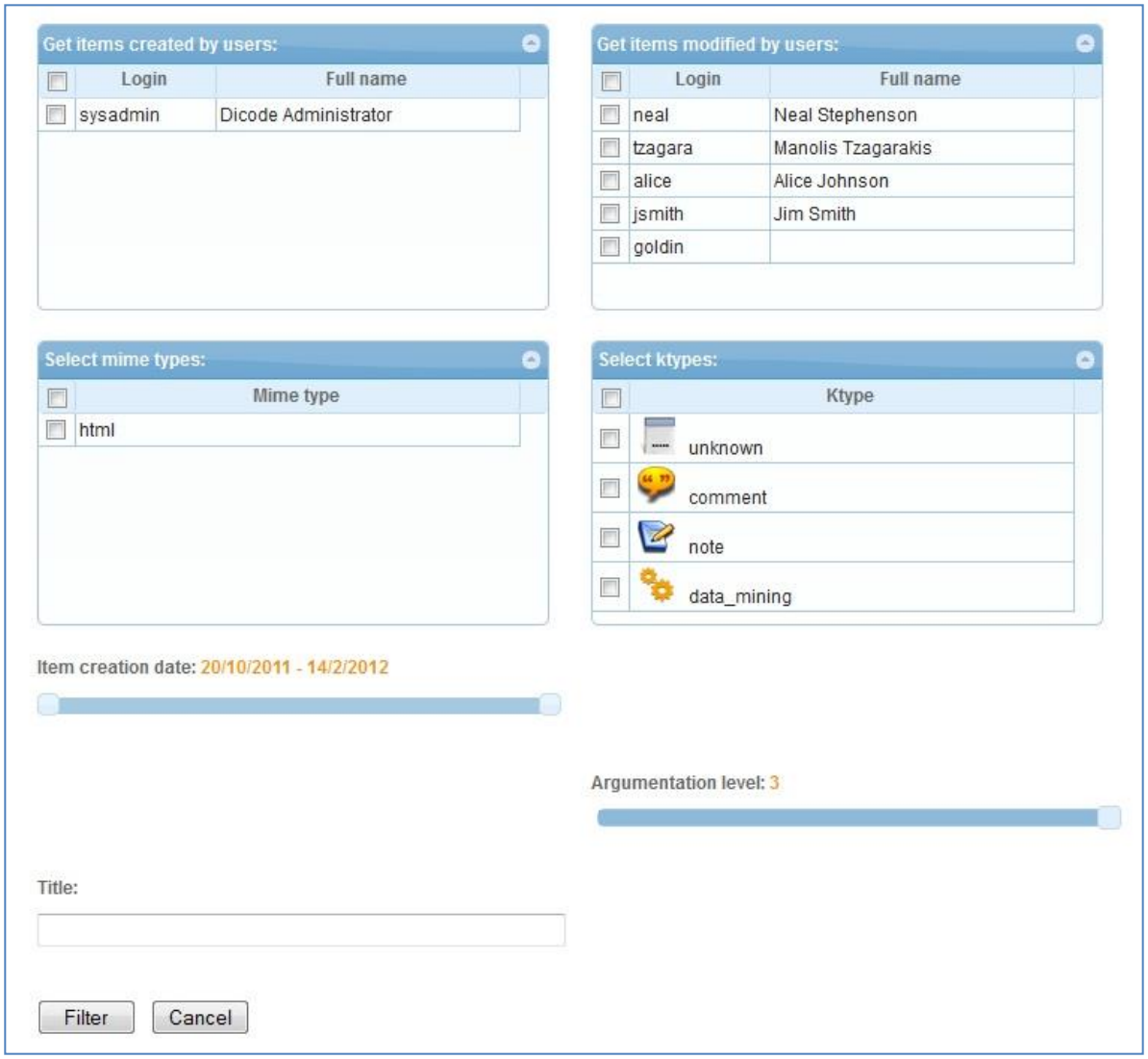
Collaboration workspace: Forum view



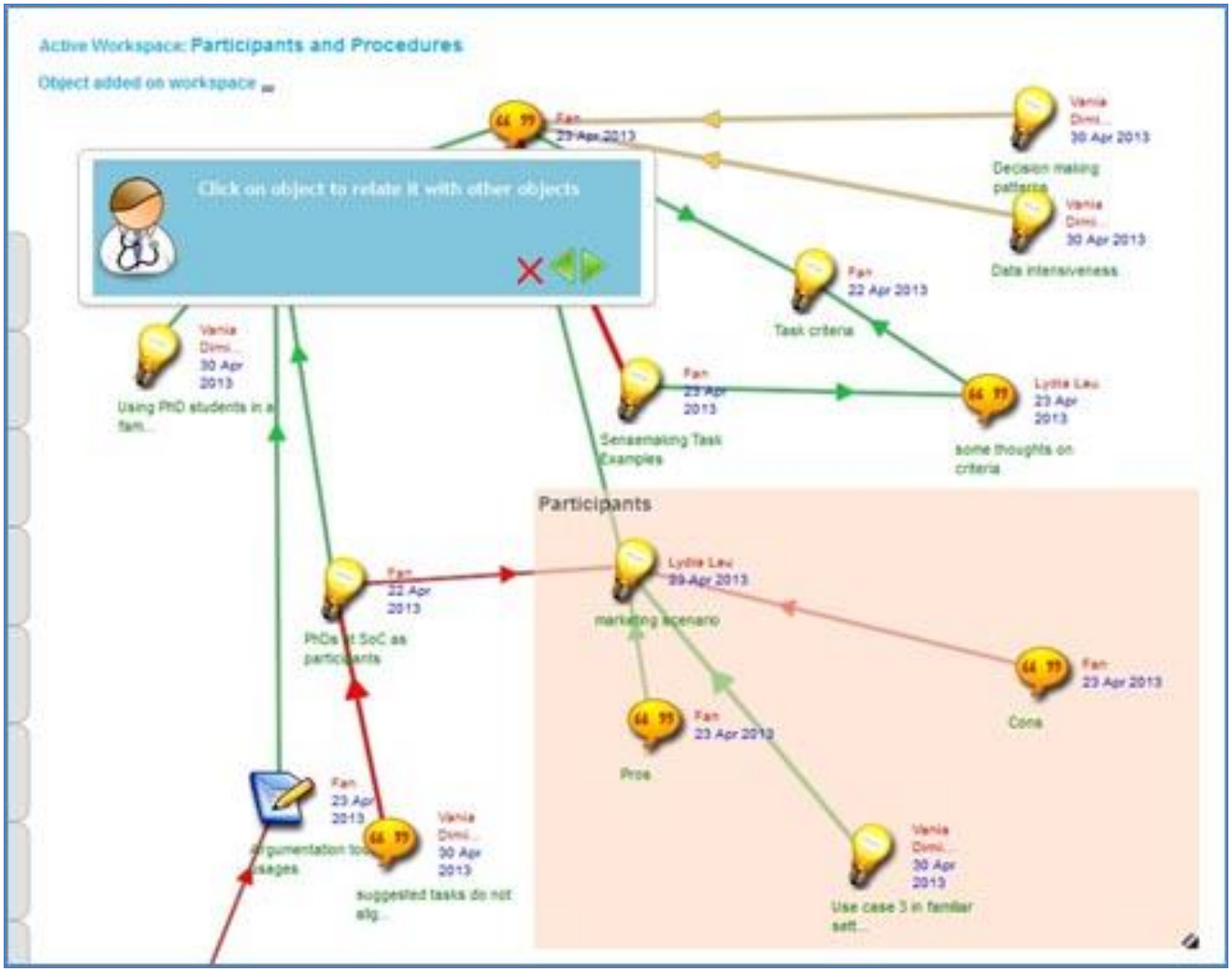
Collaboration workspace: Mind-map view



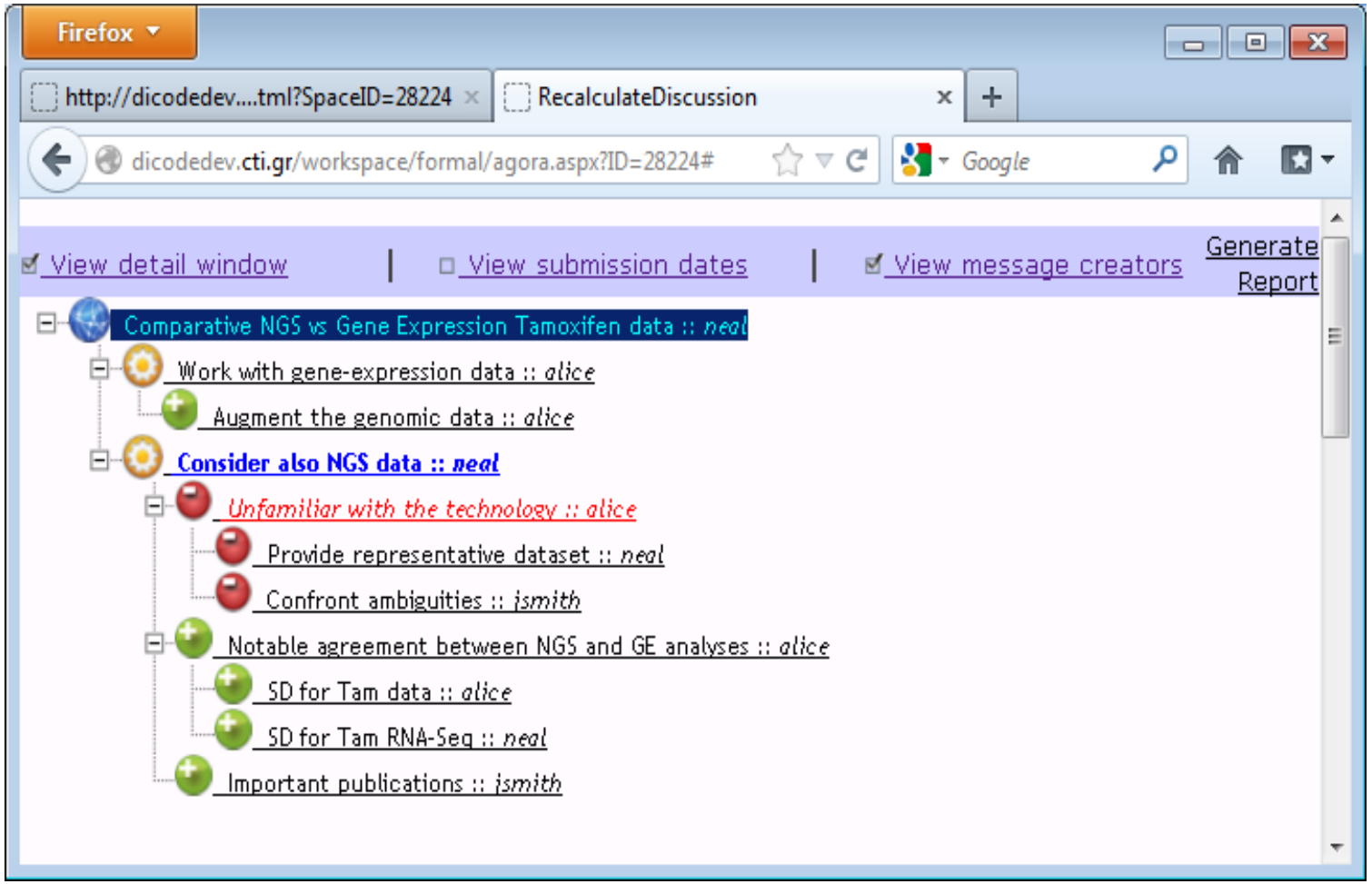
Collaboration workspace: Filtering items



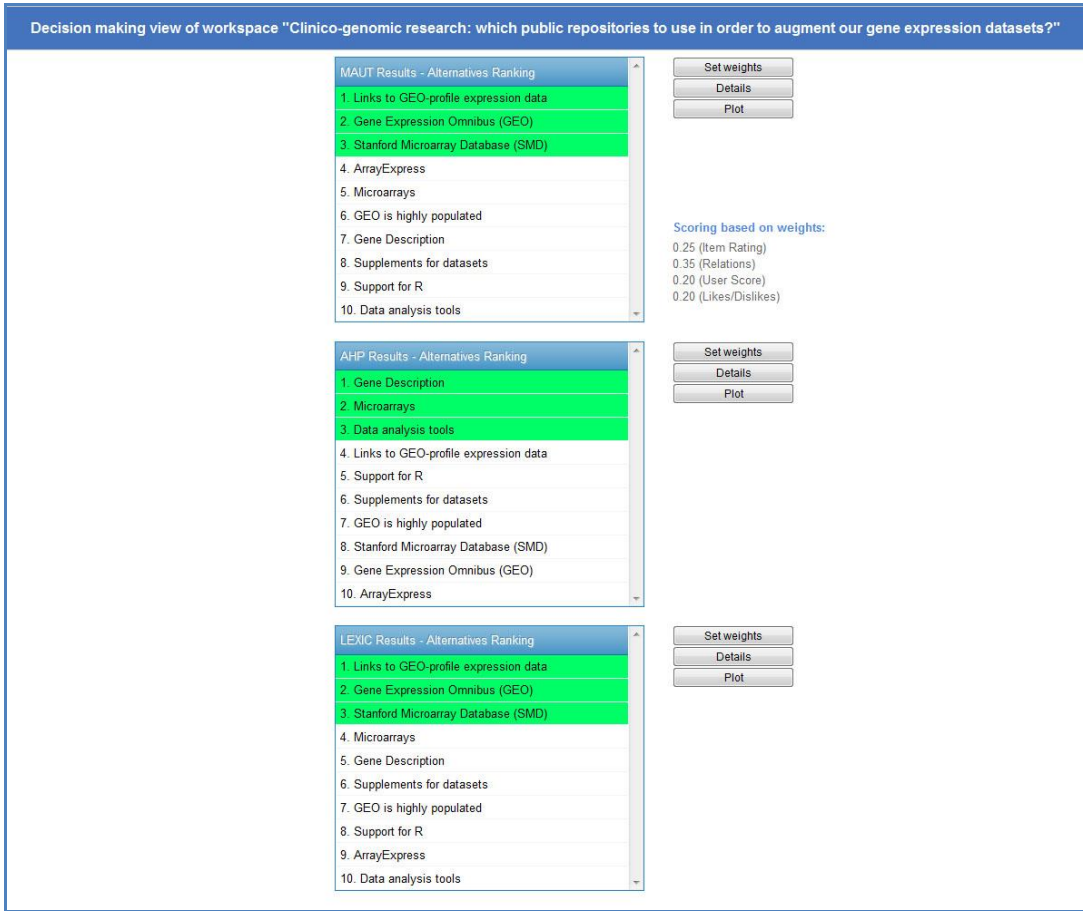
Collaboration workspace: Proactive help



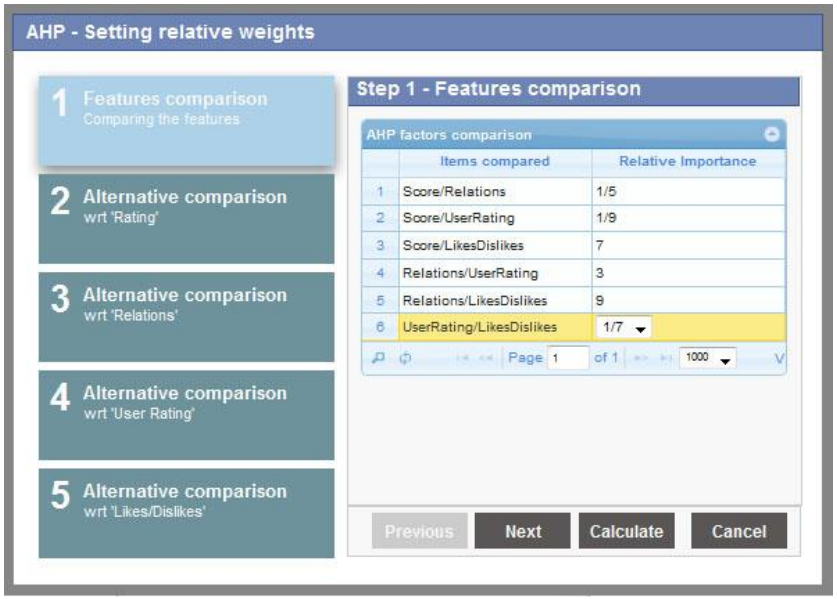
Collaboration workspace: Formal view



Collaboration workspace: Other views

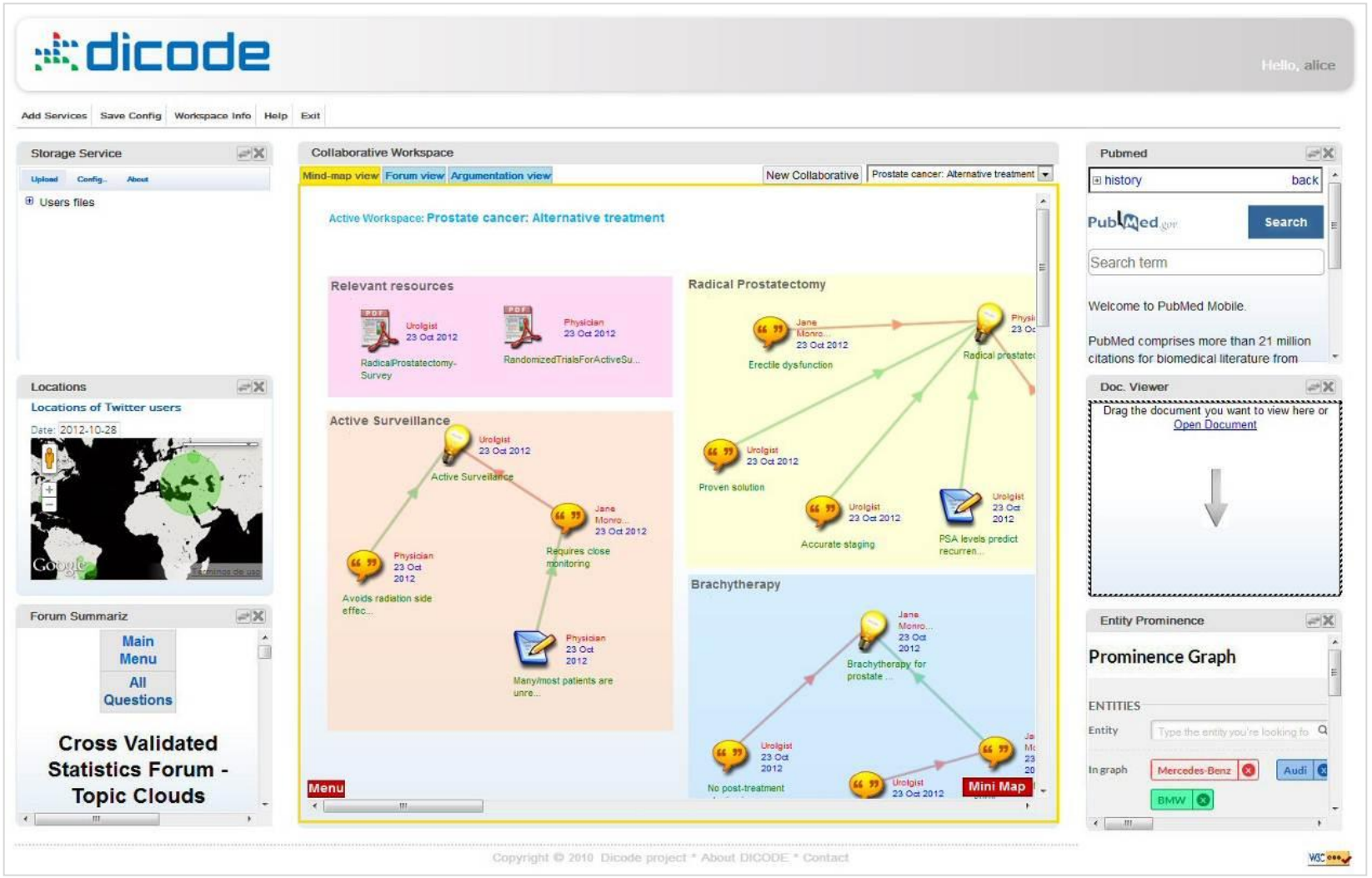


Decision Making view

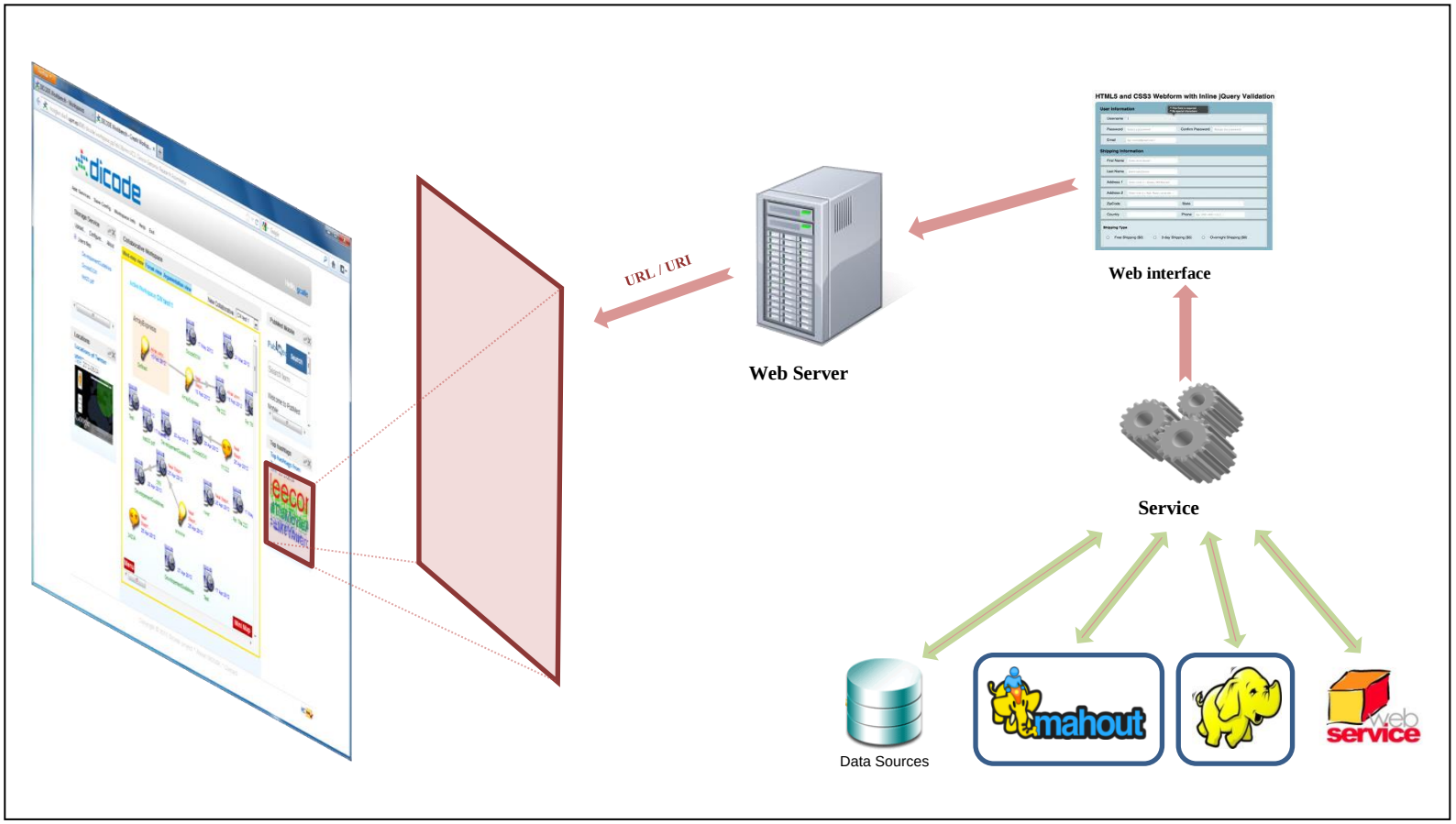


Workspace analytics view

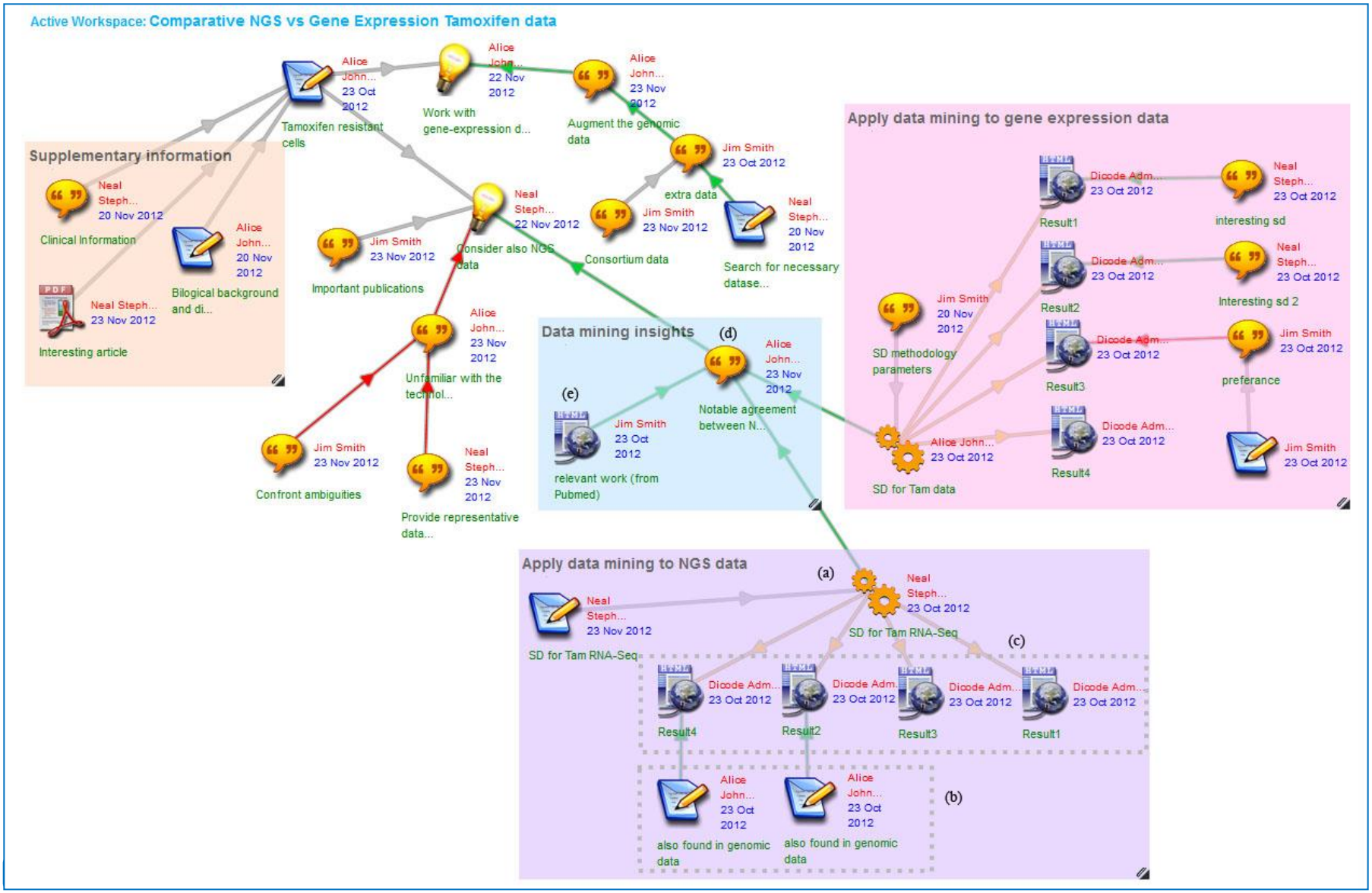
The Dicode Workbench



Integrating services (1/2)



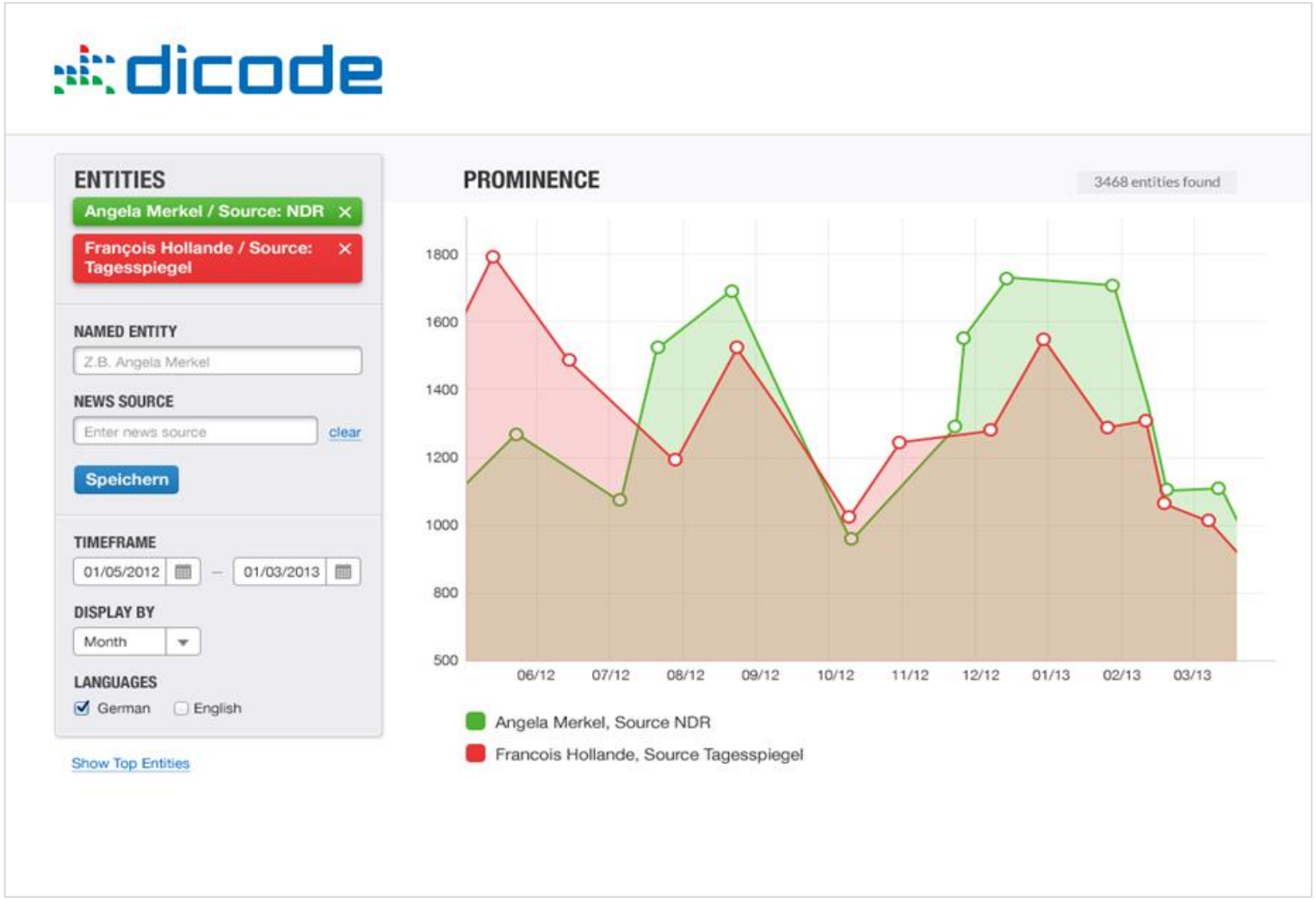
Integrating services (2/2)



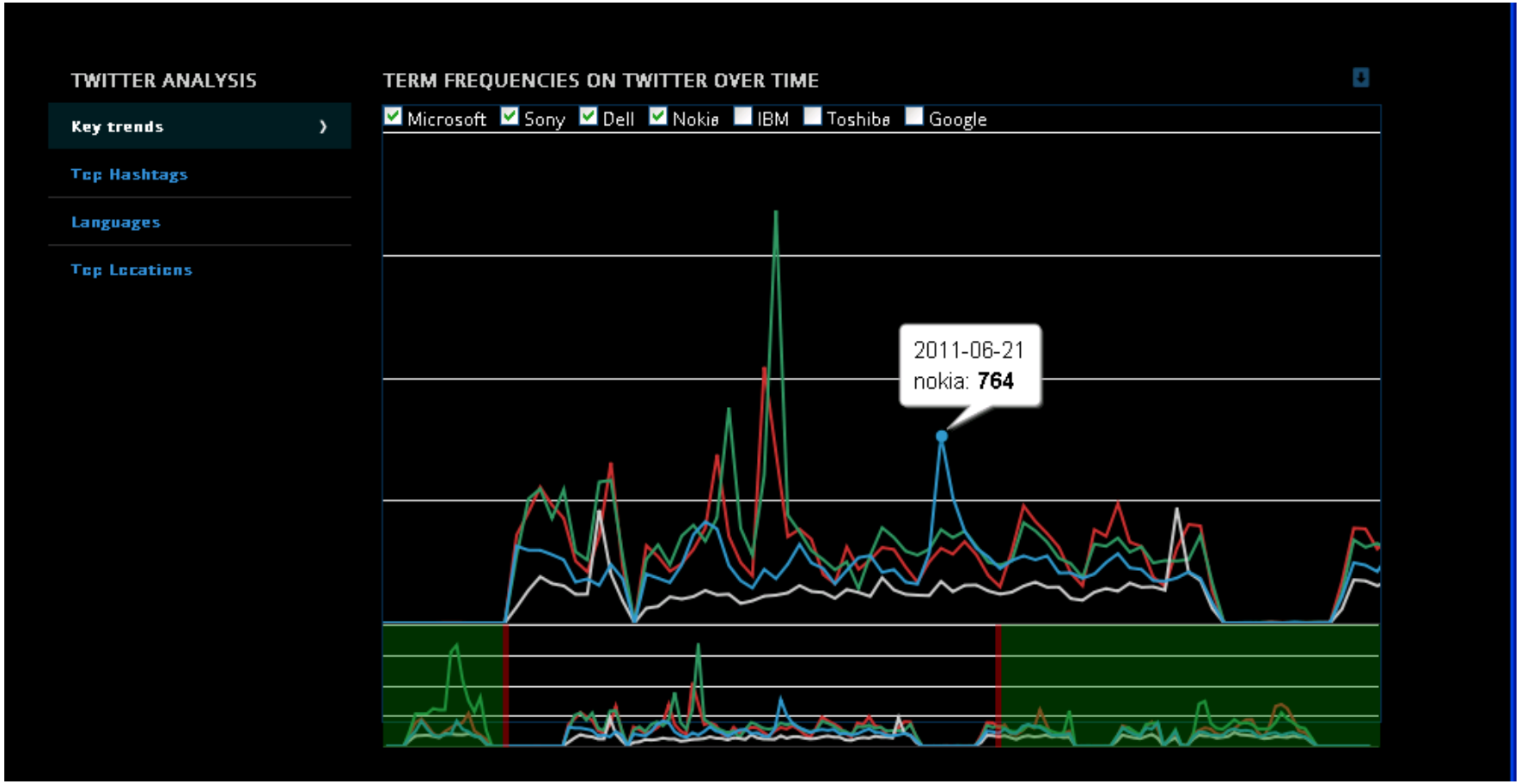
Data mining services (1/3)

- Forum Summarization service
  - takes a cluster of discussion threads (from various web forum types) as input and identifies the most prominent terms (topics)
- Subgroup Discovery service
  - searches for subgroups in any user provided data by searching the rules that cover many target value examples and few non-target value examples
- Recommendation service
  - recommends similar users or documents from different types of log file data based on similarity models learned by using appropriate algorithms
- Blog-preprocessing service
  - returns a condensed representation of weblog entries containing only significant nouns
- Topics service
  - aims to give the user a quick overview of the thematic content of a document collection
- Keytrends service
  - returns metadata about Tweets on a selected day; in each case, the top 100 (or less) values are returned.
- Twitter preprocessing service
  - returns a condensed representation of Tweets containing only significant nouns. Optionally, one can query for a search term (a regular expression).

Data mining services (2/3)



Data mining services (3/3)



lessons learned

http://www.flickr.com/photos/43343237@N04/5593973222/

Evaluation issues

- Two evaluation rounds
  - Questionnaires and interviews
  - 76 and 30 users involved
- Input from 3 distinct use cases
- Positive feedback
  - functionalities related to the evolution of collaboration and provenance of associated data
  - different manners of sharing or discussing data and results
  - ease of communication and strong data/information archiving features
  - integration of data mining results

Collaboration and decision making (1/2)

- Lesson 1: Alternative views of collaboration may significantly tame the complexity of data-intensive workspaces.
  - in such environments, formality in managing collaboration should not be considered as a predefined and rigid property, but rather as an adaptable aspect that can be modified to meet the needs at hand.
- Lesson 2: Collaboration and decision making services should not be regarded as 'application islands'.
  - Seamless interoperability is a crucial factor for their adoption and success.

Collaboration and decision making (2/2)

- Lesson 3: Effective collaboration and decision making requires appropriate mechanisms tailored to the needs of each use case.
- Lesson 4: Data analytics is an iterative exploratory task which requires multi-perspective view support.
- Lesson 5: Integrating data mining into collaboration support services makes the collaboration discourse more understandable and greatly facilitates collective sense and decision making in data-intensive environments.

Data Mining (1/2)

- Lesson 6: MapReduce is not always the best choice for Big Data.
  - A combination of batch and stream processing frameworks is suggested, which combines instant stream processing of incoming data and in-depth processing of batches of data for final results.
  - Lightweight stream-processing frameworks like Storm seem to fill a gap in Big Data scenarios and serve as an easier solution for large-scale text mining.

Data Mining (2/2)

- Lesson 7: Meaningful data visualization is highly important.
- Lesson 8: Knowledge extraction yields results which are often hard to interpret.
  - For instance, pattern interpretation is a time consuming task, since human experts must manually review the patterns.
  - Explore statistical characteristics without losing statistical descriptiveness.
    - E.g. the statistical quality of pattern to output the *k* top-quality patterns.
  - Take into account user feedback in the process.
    - E.g. enable them include/exclude certain attributes from the search

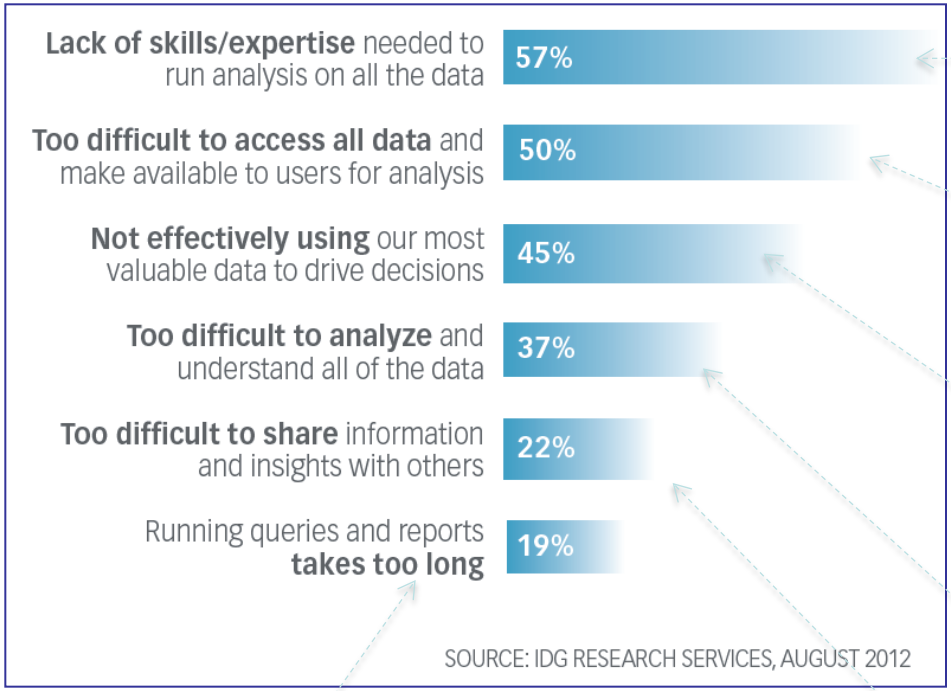
concluding ...

http://www.flickr.com/photos/37644759@N02/7128606893

Advancing the state-of-the-art

- Synergy of human and machine reasoning
  - Intelligent techniques for semantic data exploration
    - Summarization techniques for processing large & complex semantic data
  - Exploitation of collective intelligence in parameterizing data mining solutions, selecting the appropriate data sources, and interpreting data mining outputs
  - Preserving provenance of decision making
  - Appropriate visualizations of collaboration towards problem solving
- Practical success of data mining solutions
  - enabling users to guide and control the data mining process and include their domain knowledge
  - more compact and semantically-enriched data mining results
  - the overall usability of the data mining system, in particular the ability to re-use existing solutions and built upon documented decision provenance

Big Data challenges (revisited)



User-friendly environment; No data-mining expertise

Easy access through the Dicode Workbench and integrated services

Semantically-enriched data; Knowledge-based decision making

Multiple collaboration views; Argumentation-based reasoning

Intelligent data mining & collaboration support solutions;
Integrated reporting functionalities

Dicode's Collaboration Services; Exploiting the competences of all stakeholders to meaningfully confront various information management issues

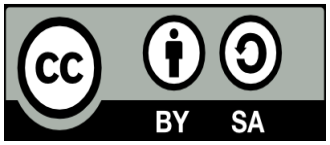
Last notes …

- Responding to technologies
  - Better understanding of big data …
  - Better understanding of cloud ...
  - No free lunch …
  - One size does not fit all …
  - Exploiting the information growth by ensuring a flexible, adaptable and scalable information and computation infrastructure;

- Getting broader
  - Generic Dicode framework
  - Dissemination and exploitation activities


Thanks!

Τέλος Ενότητας

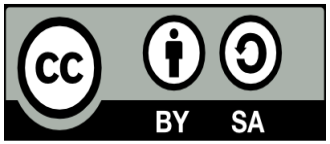


Σημείωμα Ιστορικού Εκδόσεων Έργου

Το παρόν έργο αποτελεί την έκδοση 1.0.

Σημείωμα Αναφοράς

Σημείωμα Αδειοδότησης



Σημείωμα Χρήσης Έργων Τρίτων

Το Έργο αυτό κάνει χρήση των ακόλουθων έργων:

Διαφάνεια 22:

https://www.flickr.com/photos/piero/151739521/

Διαφάνεια 24:

http://cdn2.hubspot.net/hub/175357/file-28809138-png/images/page_header_pbprofessional-resized-600.png?t=1432642333001

Διαφάνεια 24:

https://wce.wwu.edu/files/pages/EnterKey.jpg

Σημείωμα Χρήσης Έργων Τρίτων

Το Έργο αυτό κάνει χρήση των ακόλουθων έργων:

Διαφάνεια 29:

https://www.flickr.com/photos/44704258@N06/5252531705/

Διαφάνεια 47:

https://www.flickr.com/photos/43343237@N04/5593973222/

Διαφάνεια 53:

https://www.flickr.com/photos/37644759@N02/7128606893

Διατήρηση Σημειωμάτων

Οποιαδήποτε αναπαραγωγή ή διασκευή του υλικού θα πρέπει να συμπεριλαμβάνει:

- το Σημείωμα Αναφοράς
- το Σημείωμα Αδειοδότησης
- τη δήλωση Διατήρησης Σημειωμάτων
- το Σημείωμα Χρήσης Έργων Τρίτων (εφόσον υπάρχει) μαζί με τους συνοδευόμενους υπερσυνδέσμους.