



# Mastering Data-Intensive Collaboration and Decision Making

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

*Keynote Lecture*  
*IC3K 2013 – Vilamoura, Portugal*  
*22 September 2013*



Dicode ([dicode-project.eu](http://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

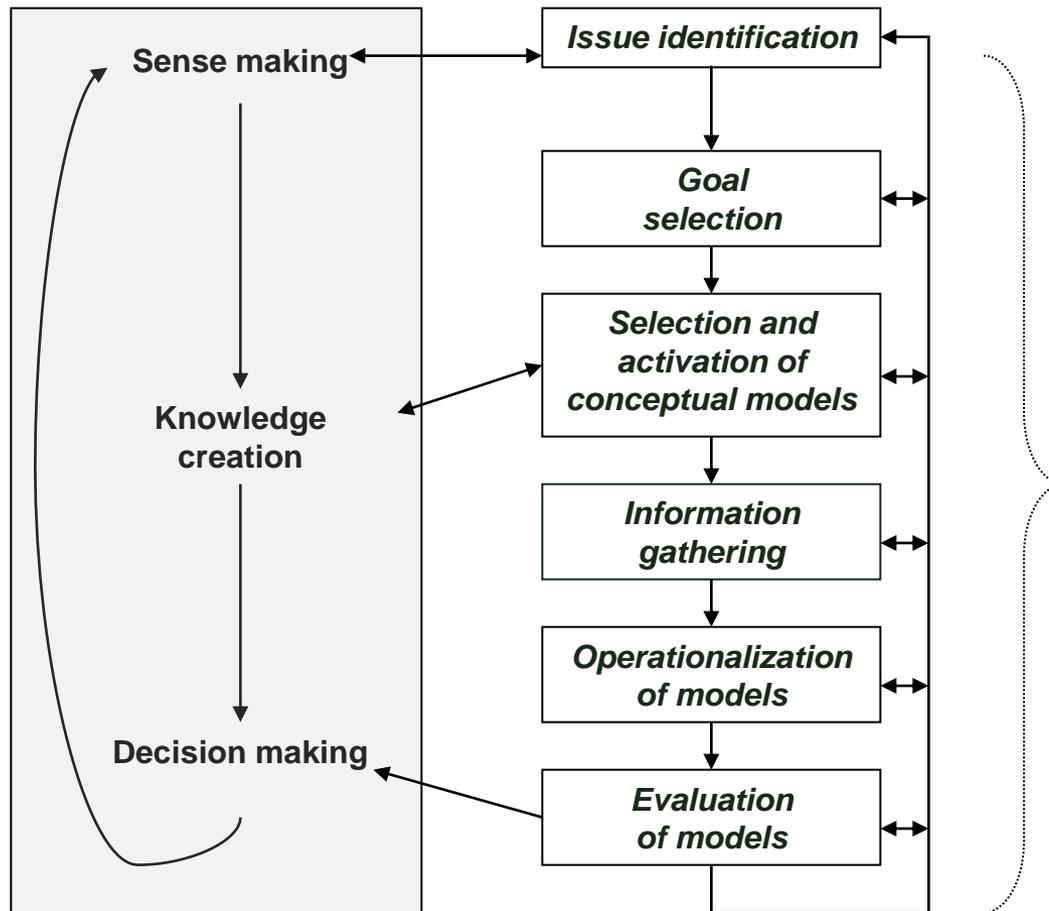
# Services required

- 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)

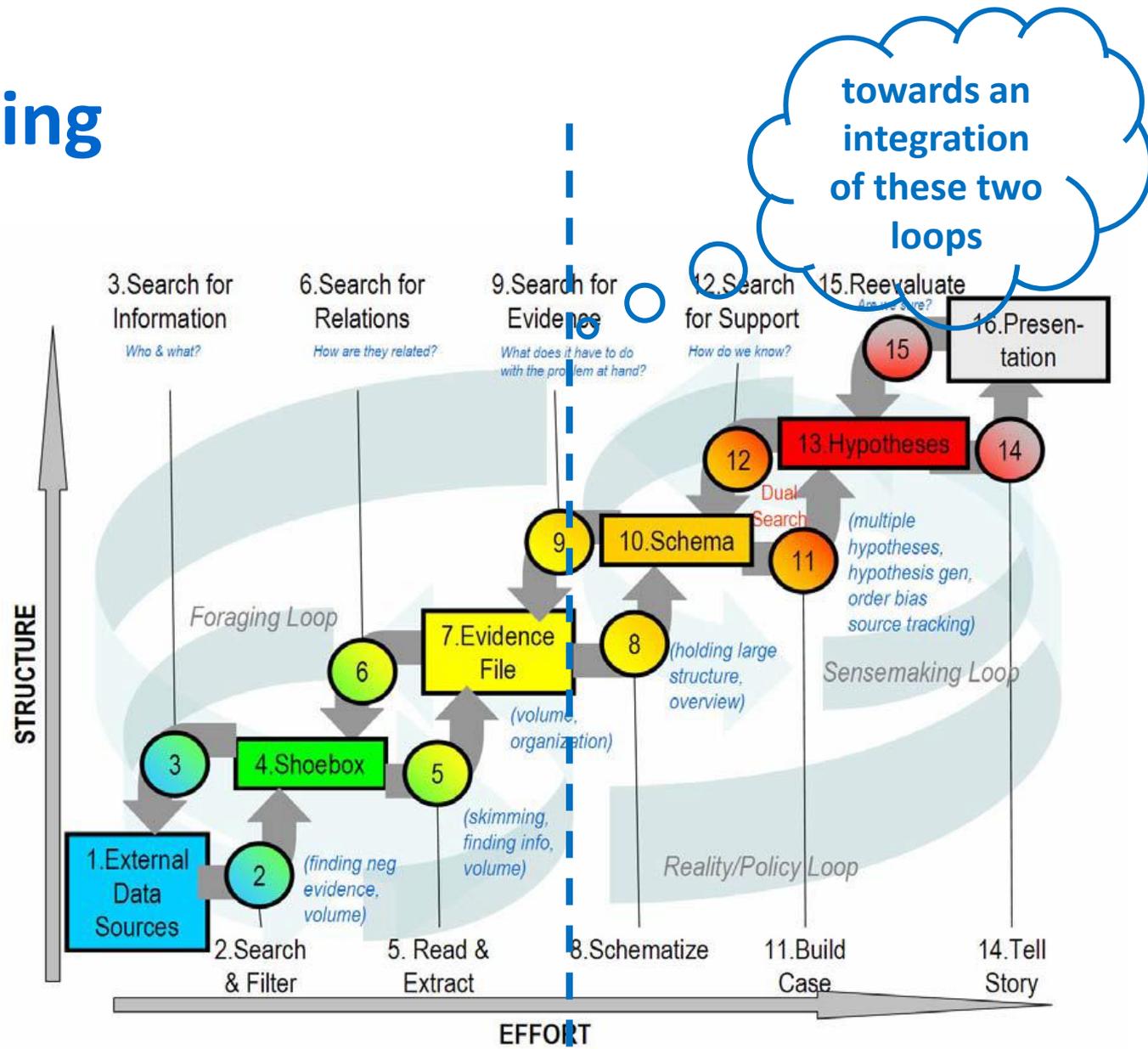
- Communities of practice (CoPs) are formed for resolving poorly structured problems, over which a plurality of views holds
  - Diverse group of specialists **exchange views** through elaborated discussion
  - The objective is to create and use problem-specific knowledge through the **social interaction**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



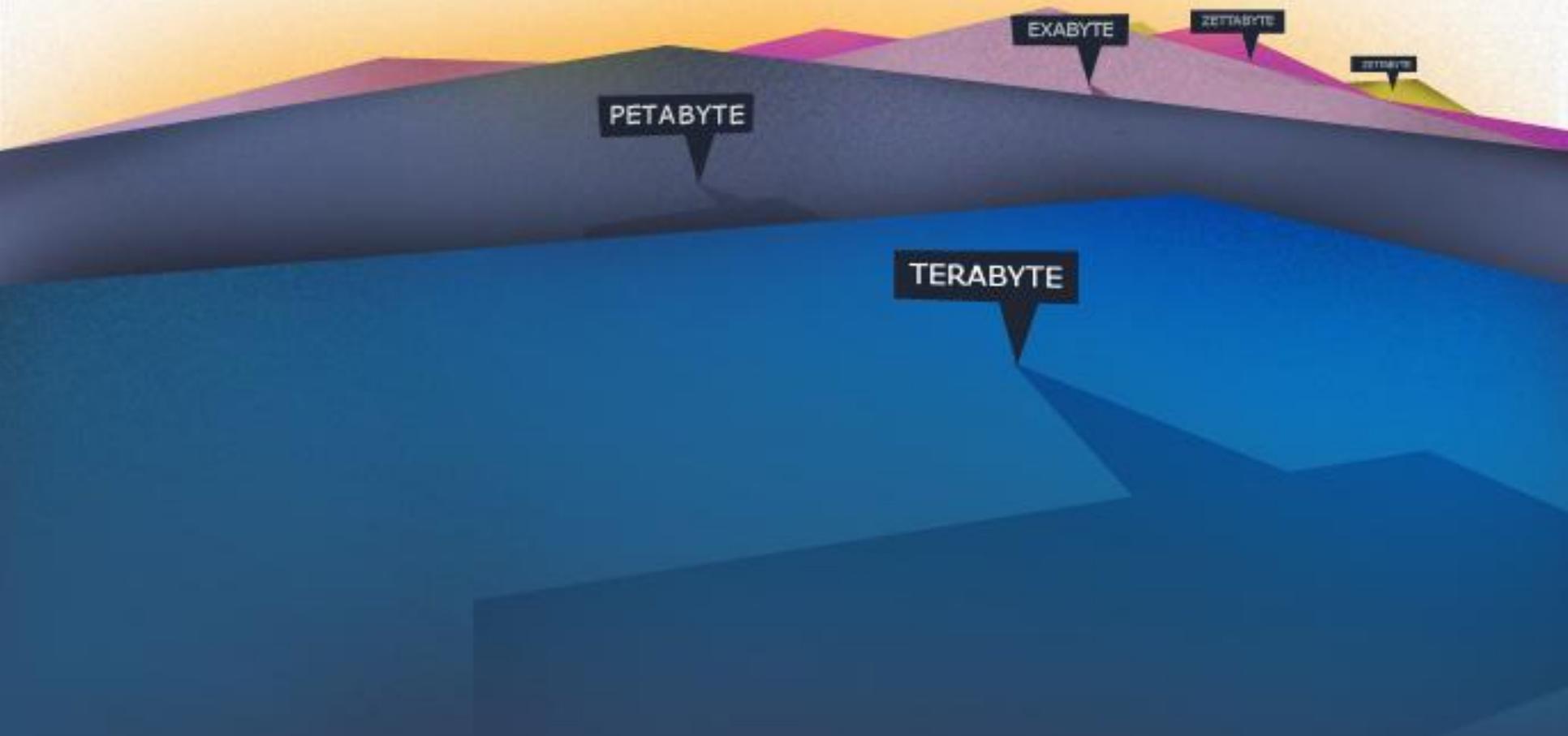
# State-of-the-art and beyond (1/2)

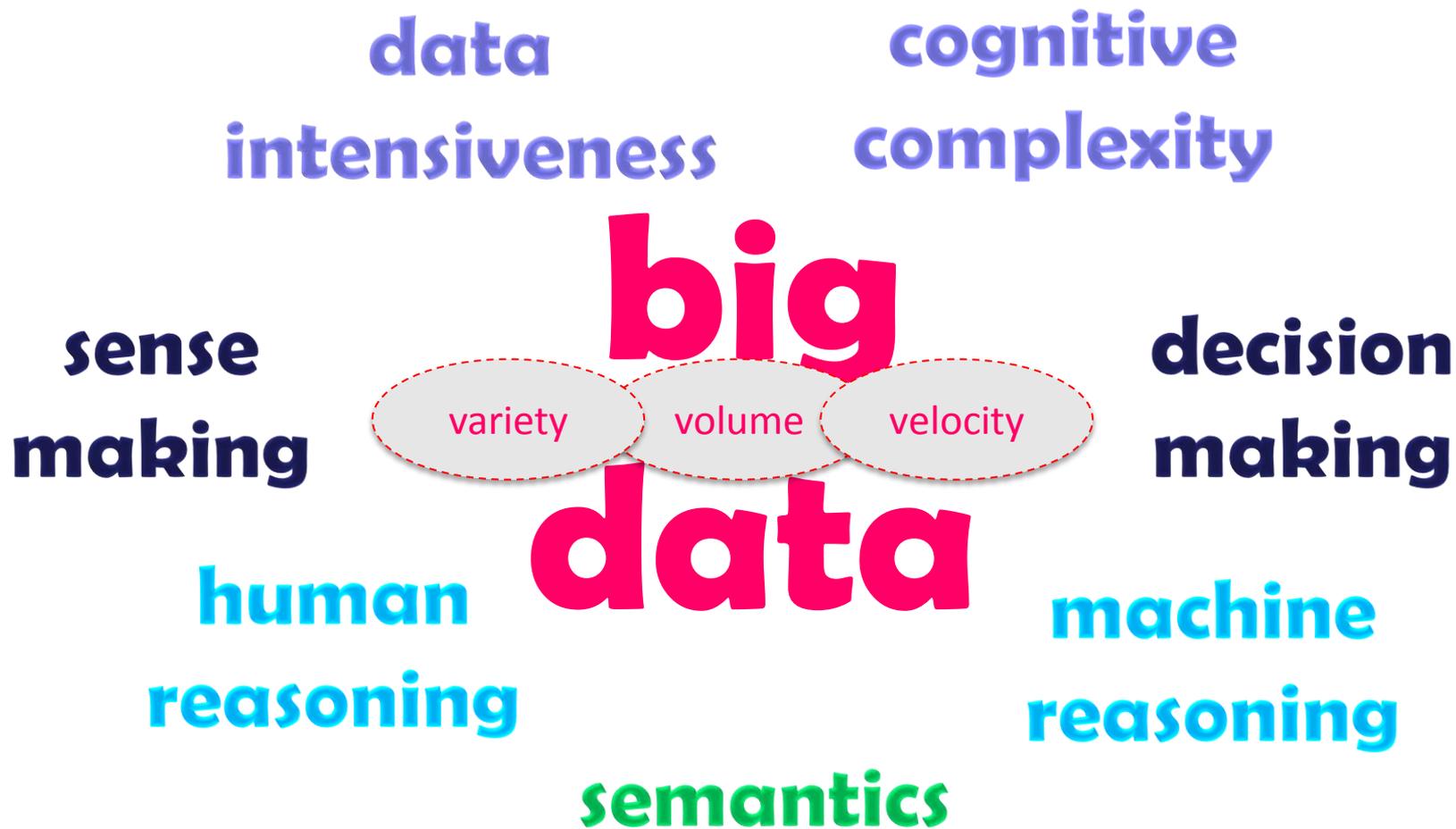
- Collaboration support
  - Current tools are “information islands” → increased interoperability and synergy with third party tools
  - Web 2.0 collaboration tools are rather passive media → intelligent reasoning services to actively and meaningfully support collaboration
  - Web 2.0 collaboration tools 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





# 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 collage consists of several overlapping screenshots from different websites:

- Technology Review:** A screenshot of a browser window showing the article "With Big Data Comes Big Responsibilities". The URL is <http://www.technologyreview.com/computing/38775/>. The page features a "Special Report: The Future of the Office" banner.
- Gartner:** A screenshot of the Gartner website with the headline "Gartner Says Solving 'Big Data' Challenge Involves More Than Just Managing Volumes of Data". The URL is <http://www.gartner.com/it/page.jsp?id=1731916>.
- NetworkWorld:** A screenshot of the NetworkWorld website with the headline "Few data scientists happy with current state of 'big data' analytics". The URL is [http://www.networkworld.com/newsletters/nem2\(2011/080811nem2\)source=NNW/NLE\\_nk\\_netwo](http://www.networkworld.com/newsletters/nem2(2011/080811nem2)source=NNW/NLE_nk_netwo). The page includes a "Sign up for this newsletter now!" form and a poll result stating that 97% of data scientists believe "big data" analytics needs improvement.
- Dico:** A logo for "dico" is visible in the bottom left corner of the collage.

# The Big Data fallacy

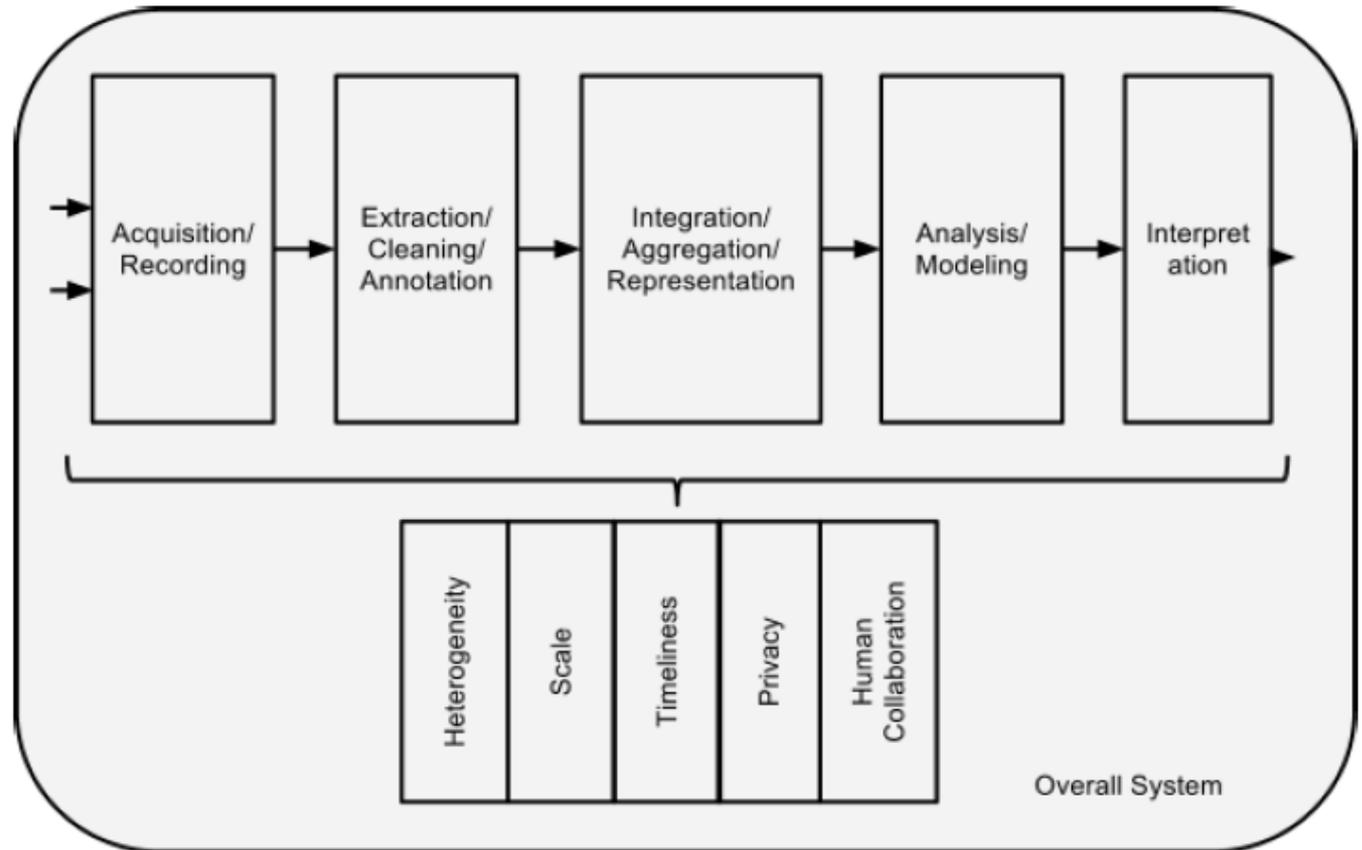
- More data doesn't mean you will get “*proportionately*” more information
  - In fact, the more data you have, the less information you gain as a proportion of the data
- The value of big data is often overestimated
  - Its value is in the information that can be extracted from it
  - Information is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

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)

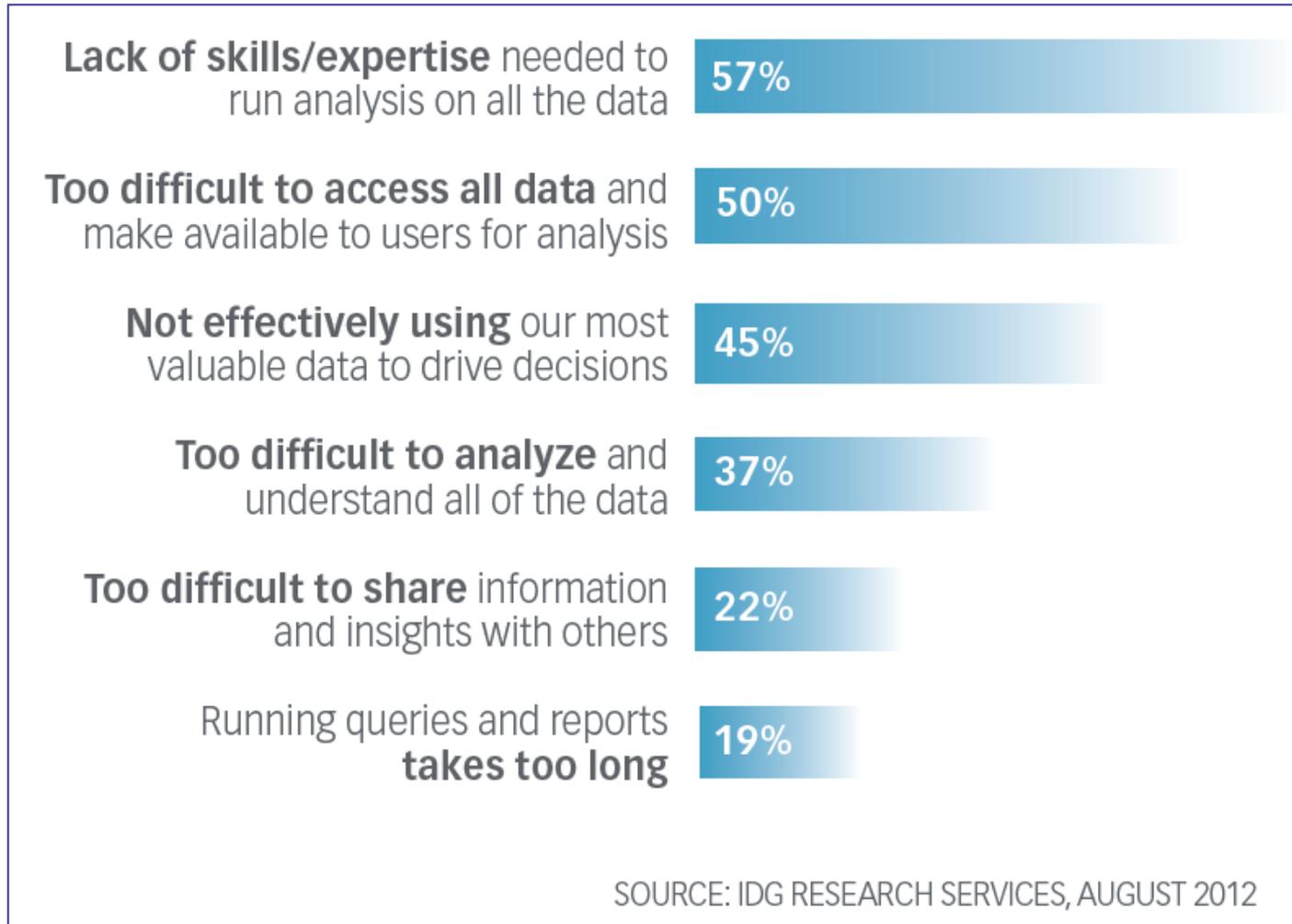


# Big Data white paper - Feb 2012 (2/2)

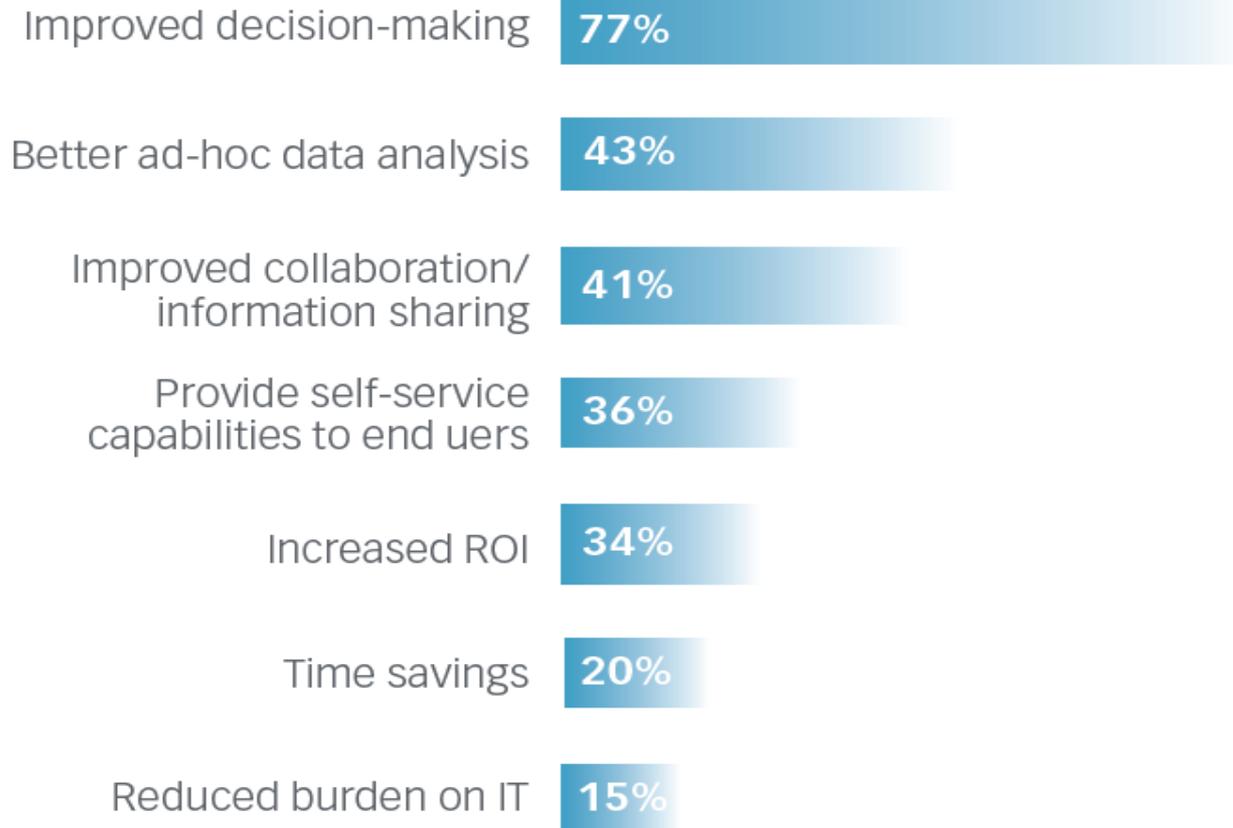
- “In spite of the tremendous advances made in computational analysis, there remain many **patterns that humans can easily detect but computer algorithms have a hard time finding**”
- “Ideally, analytics for Big Data will not be all computational – **rather it will be designed explicitly to have a human 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 Suci, Univ. of Washington  
Shiv Vaithyanathan, IBM  
Jennifer Widom, Stanford Univ.

# Big Data challenges



# Top benefits of data visualization



SOURCE: IDG RESEARCH SERVICES, AUGUST 2012

# the Dicode roadmap



# Dicode's main goal

## ■ What

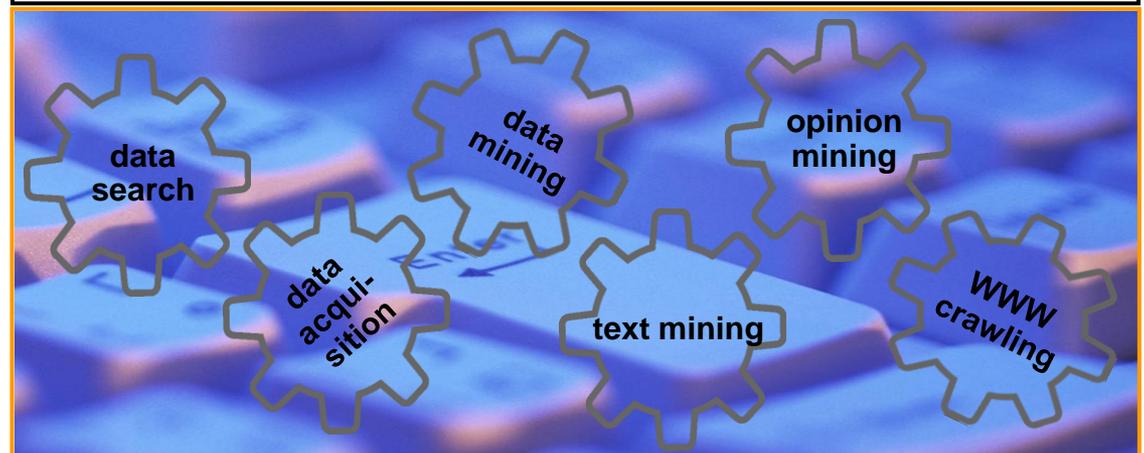
- facilitate and augment collaboration and decision making in data-intensive and cognitively-complex settings

## ■ 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

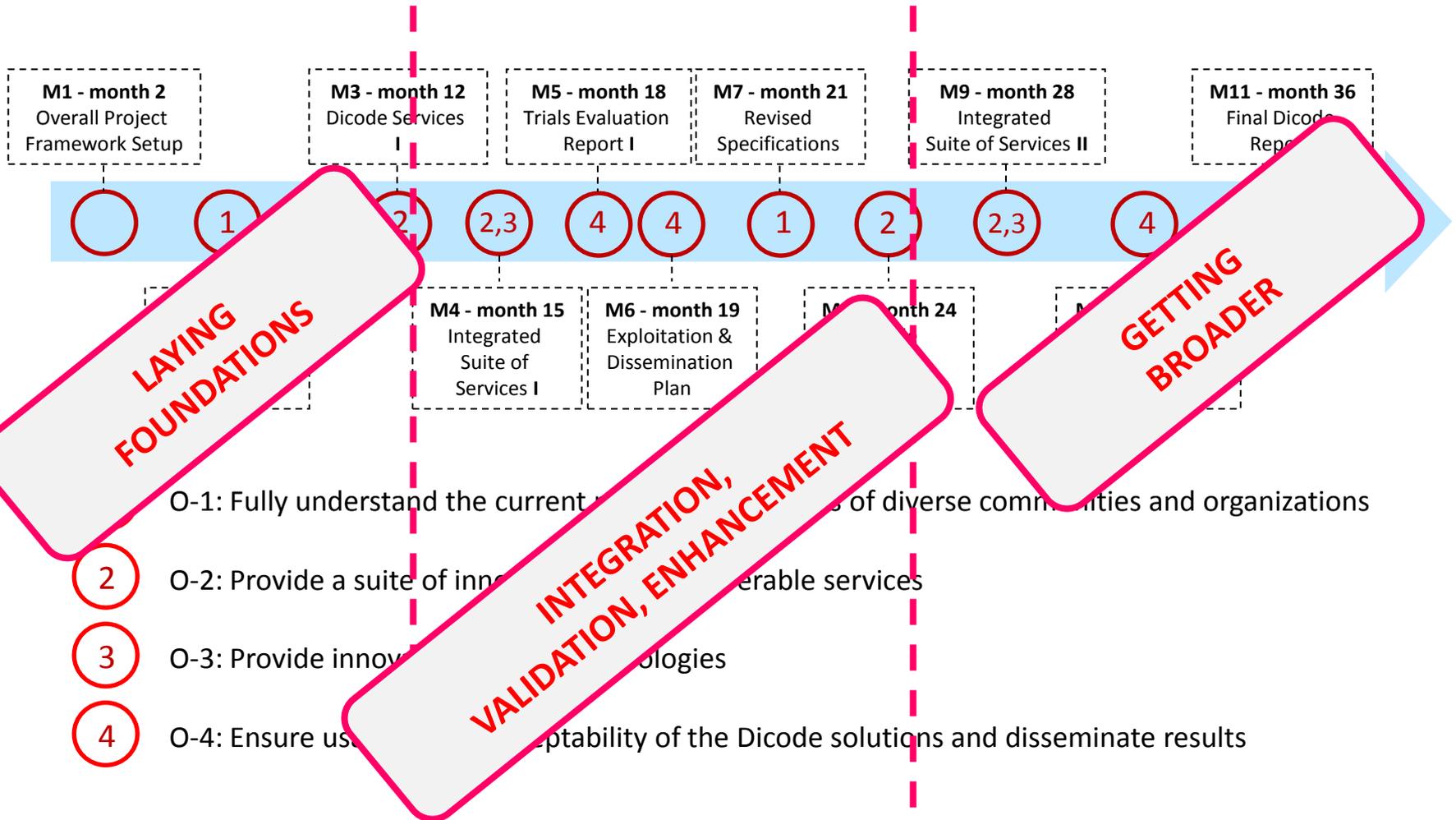
proper  
orchestration  
and exploitation  
of each side's  
strengths



# 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

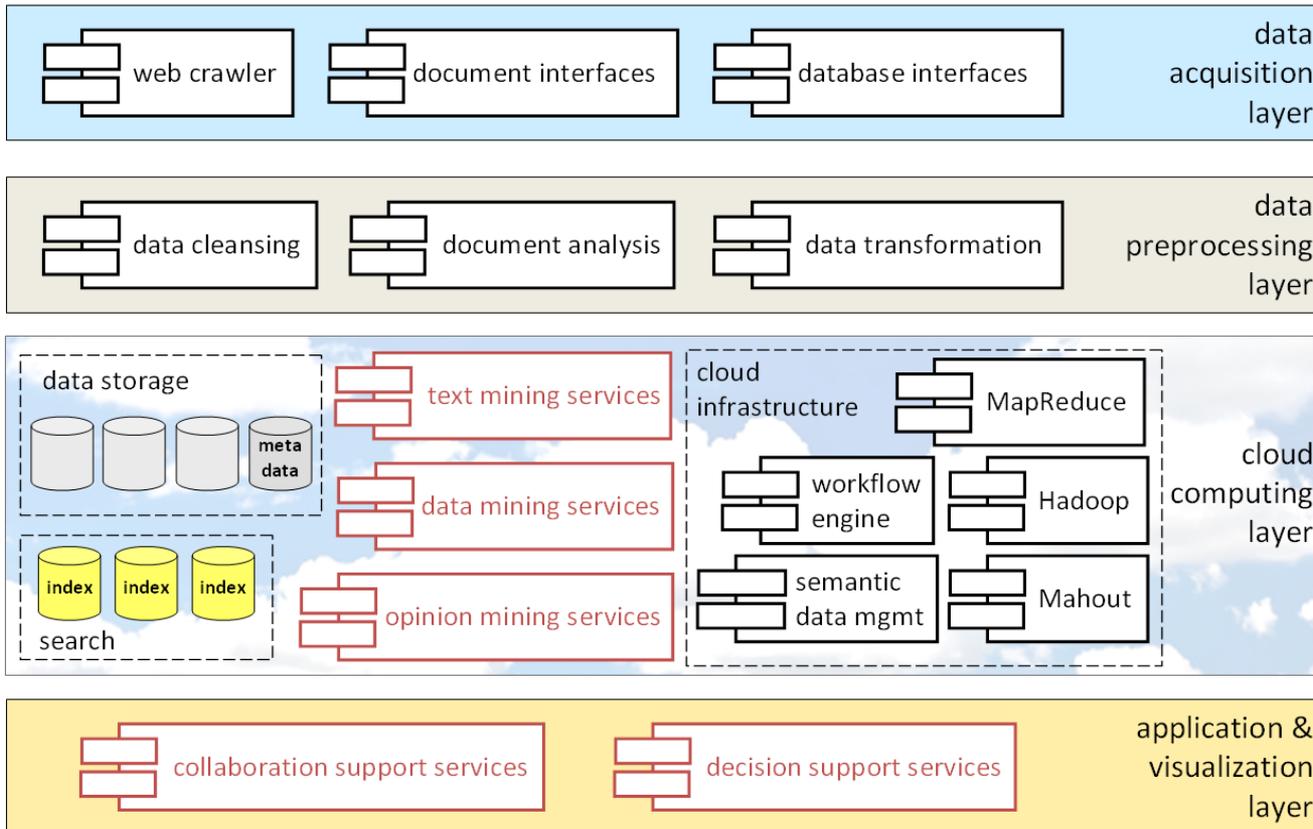
A close-up, low-angle shot of a red brick wall. The bricks are arranged in a traditional pattern, and the mortar is a light grey color. The wall recedes into the distance, creating a sense of depth. The word "Achievements" is overlaid in white, sans-serif font on the left side of the image.

# Achievements

# The Dicode architecture



data sources  
(documents,  
databases,  
world wide web)



# The Dicode Workbench

The screenshot displays the Dicode Workbench interface, which is a collaborative workspace for analyzing forum discussions. The main workspace is titled "Prostate cancer: Alternative treatment" and is divided into four colored panels, each representing a different forum topic:

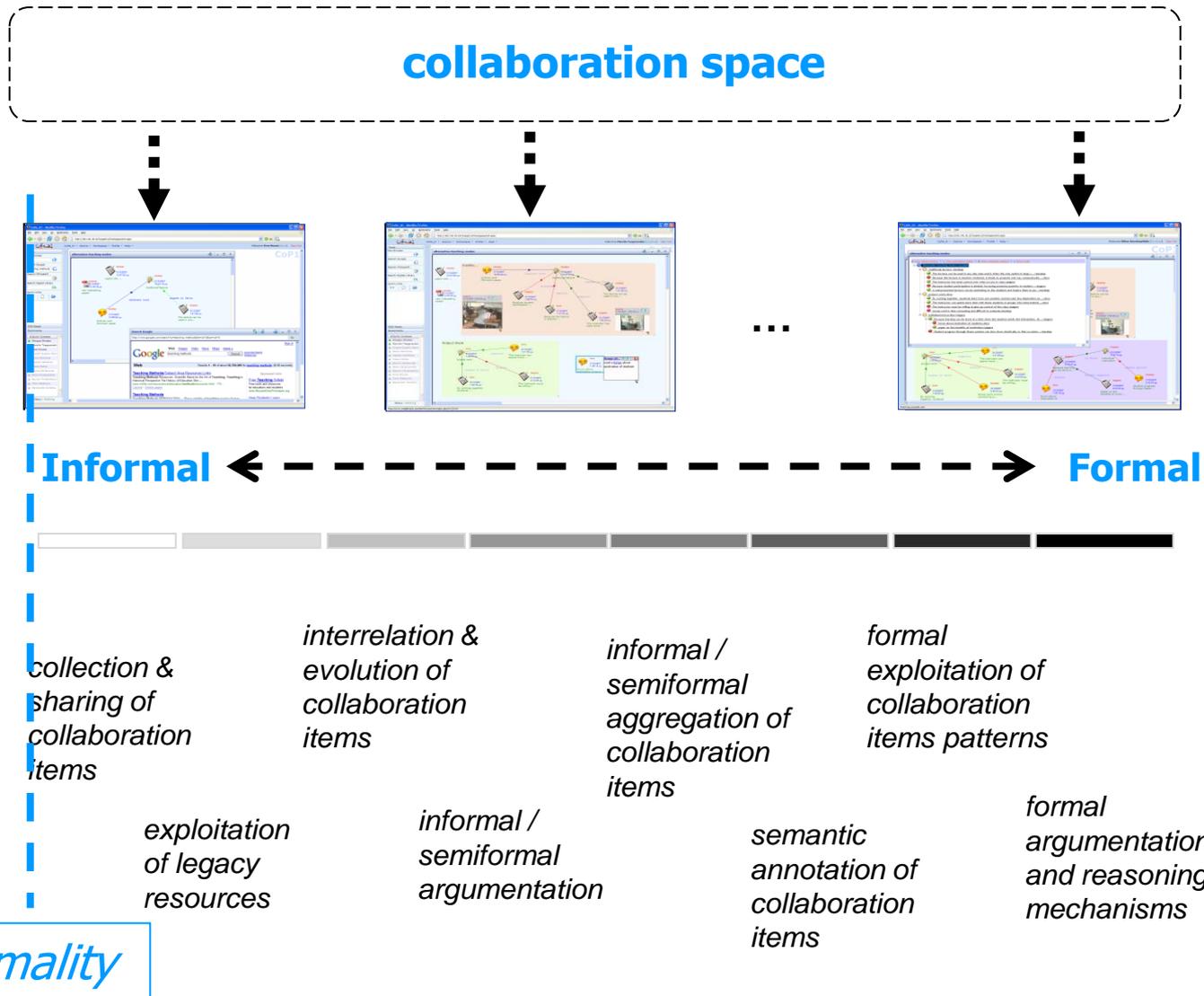
- Relevant resources (Pink):** Contains two PDF documents: "RadicalProstatectomy-Survey" and "RandomizedTrialsForActiveSu...".
- Active Surveillance (Orange):** A mind map showing a central node "Active Surveillance" connected to "Urologist 23 Oct 2012" and "Physician 23 Oct 2012". A sub-node "Requires close monitoring" is connected to "Physician 23 Oct 2012".
- Radical Prostatectomy (Yellow):** A mind map with a central node "Radical prostatectomy" connected to "Jane Monro... 23 Oct 2012" (Erectile dysfunction), "Urologist 23 Oct 2012" (Prevent solution), "Urologist 23 Oct 2012" (Accurate staging), and "Urologist 23 Oct 2012" (PSA levels predict recurren...).
- Brachytherapy (Blue):** A mind map with a central node "Brachytherapy for prostate..." connected to "Urologist 23 Oct 2012" (No post-treatment) and "Urologist 23 Oct 2012" (Ja...).

Surrounding the main workspace are several utility panels:

- Storage Service:** Includes "Upload", "Config...", and "About" buttons, and a "Users files" section.
- Locations:** A map titled "Locations of Twitter users" for the date "2012-10-28".
- Forum Summariz:** Contains "Main Menu", "All Questions", and "Cross Validated Statistics Forum - Topic Clouds".
- Pubmed:** A search window with a search bar and a "Search" button.
- Doc. Viewer:** A document viewer with a "Drag the document you want to view here or Open Document" instruction and a downward arrow.
- Entity Prominence:** A "Prominence Graph" showing entities like "Mercedes-Benz", "Audi", and "BMW".

At the bottom of the interface, there is a copyright notice: "Copyright © 2010 Dicode project \* About DICODE \* Contact" and a "VSC" logo.

# Collaboration workspaces



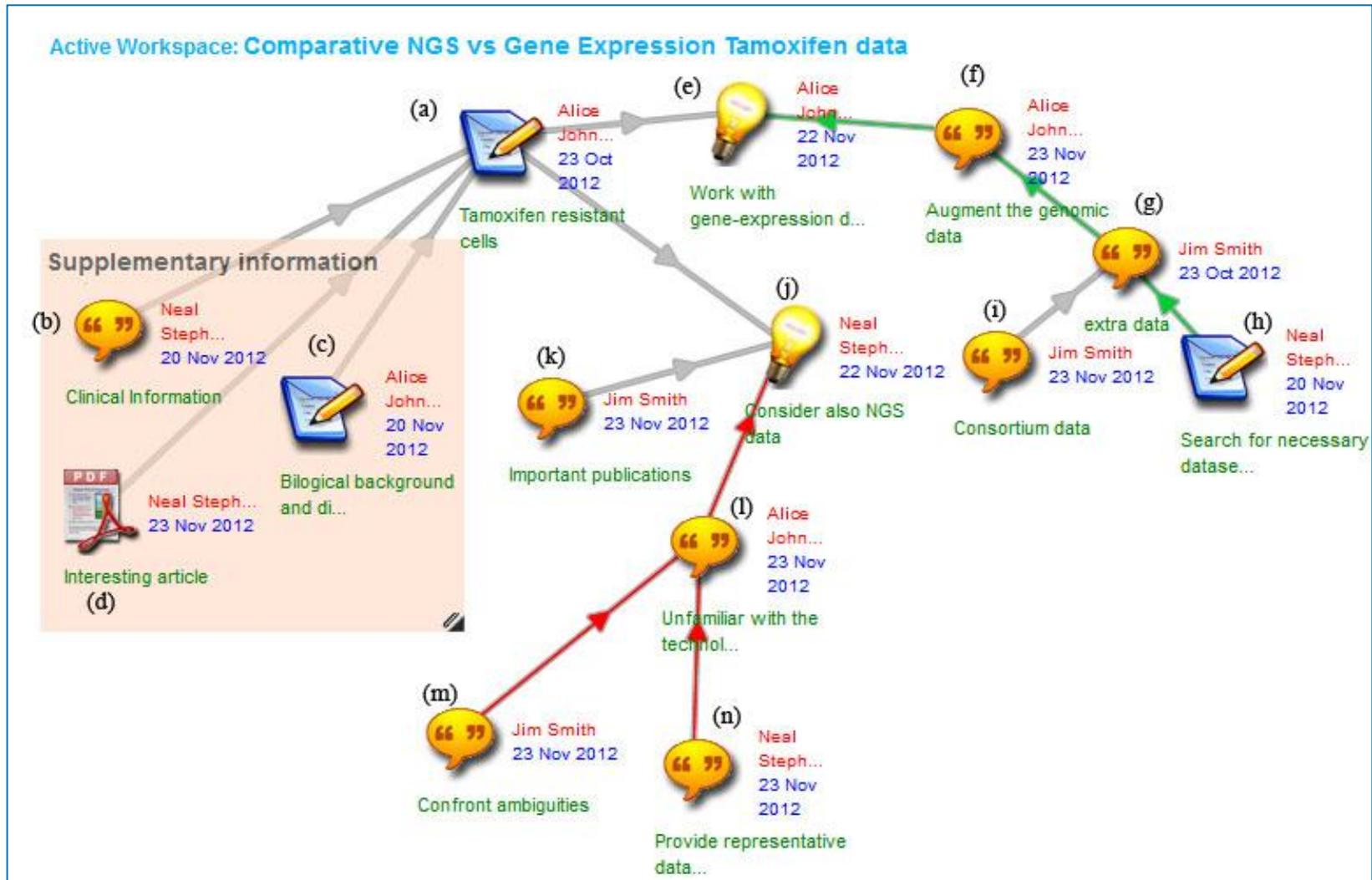
# Collaboration workspace: Forum view

Forum view of workspace: **SD service > Prostate cancer: Alternative treatment**

[Post Reply](#) [Relevant discussions](#)

| Author   | Message  |
|--|--|
| <br><b>Physician</b><br>Physician   | <h3>Avoids radiation side effects from radiation</h3> <p>Posted on 10/23/2012 5:09 PM</p> <p><b>QUOTE</b></p> <p>Avoids site effects from radiation therapy or prostatectomy</p>   |
| <br><b>jane</b><br>Jane Monroe      | <h3>Requires close monitoring</h3> <p>Posted on 10/23/2012 5:11 PM</p> <p><b>QUOTE</b></p> <p>Requires close monitoring (regular digital rectal exams, PSA tests, and prostate biopsy) to monitor for signs of progression</p> |
| <br><b>Urologist</b><br>Urologist  | <h3>Proven solution</h3> <p>Posted on 10/23/2012 5:16 PM</p> <p><b>QUOTE</b></p> <p>Proven to reduce prostate cancer death rates</p>   |
| <br><b>Urologist</b><br>Urologist | <h3>Accurate staging</h3> <p>Posted on 10/23/2012 5:17 PM</p> <p><b>QUOTE</b></p> <p>Removed tissue allows accurate staging. This is very important.</p>   |

# Collaboration workspace: Mind-map view



# Collaboration workspace: Filtering items

Get items created by users:

| <input type="checkbox"/> | Login    | Full name            |
|--------------------------|----------|----------------------|
| <input type="checkbox"/> | sysadmin | Dicode Administrator |

Get items modified by users:

| <input type="checkbox"/> | Login   | Full name          |
|--------------------------|---------|--------------------|
| <input type="checkbox"/> | neal    | Neal Stephenson    |
| <input type="checkbox"/> | tzagara | Manolis Tzagarakis |
| <input type="checkbox"/> | alice   | Alice Johnson      |
| <input type="checkbox"/> | jsmith  | Jim Smith          |
| <input type="checkbox"/> | goldin  |                    |

Select mime types:

| <input type="checkbox"/> | Mime type |
|--------------------------|-----------|
| <input type="checkbox"/> | html      |

Select ktypes:

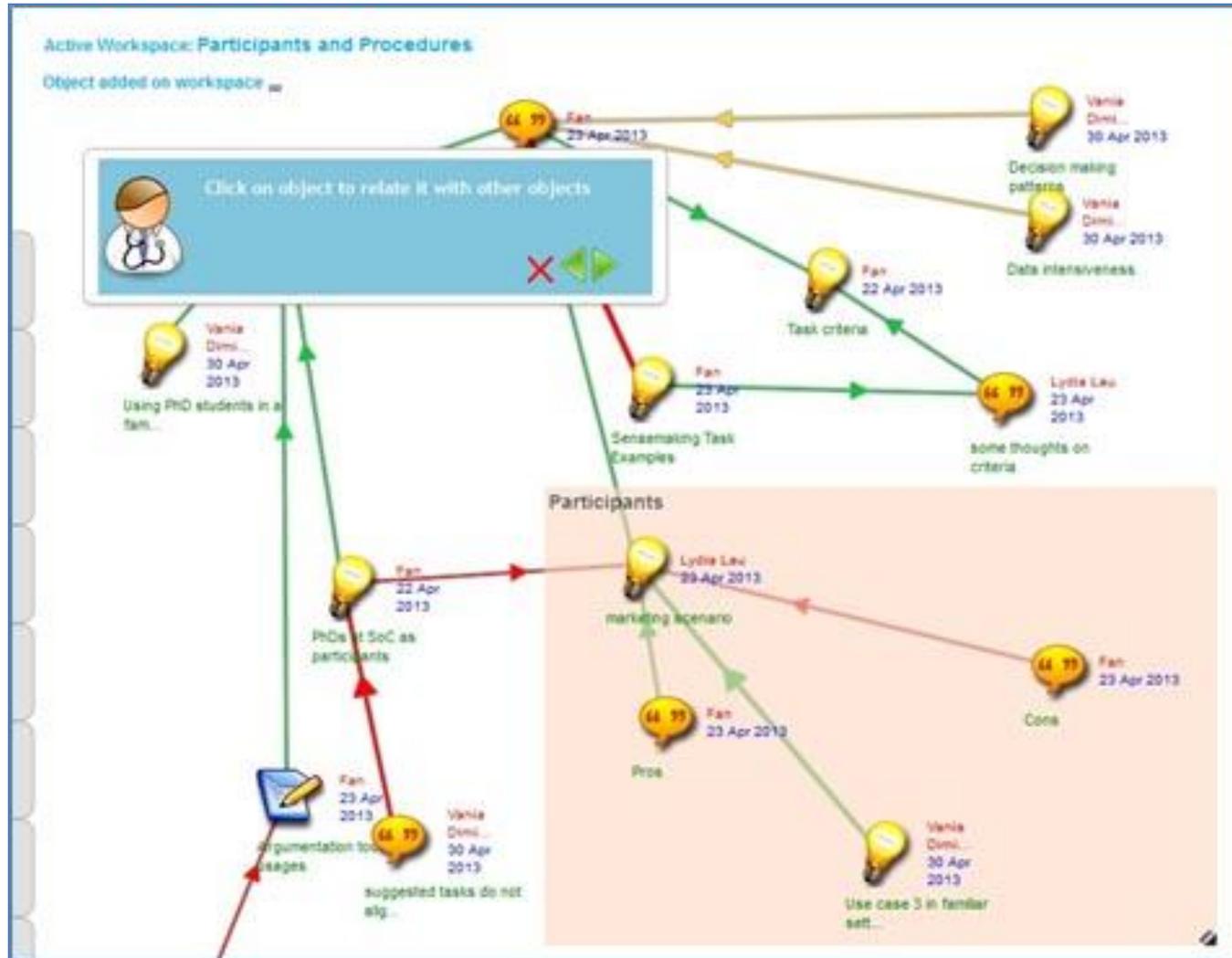
| <input type="checkbox"/> | Ktype   |
|--------------------------|---|
| <input type="checkbox"/> |  unknown     |
| <input type="checkbox"/> |  comment     |
| <input type="checkbox"/> |  note        |
| <input type="checkbox"/> |  data_mining |

Item creation date: 20/10/2011 - 14/2/2012

Argumentation level: 3

Title:

# Collaboration workspace: Proactive help



# Collaboration workspace: Formal view

The screenshot shows a Firefox browser window with the address bar displaying `dicodedev.cti.gr/workspace/formal/agora.aspx?ID=28224#`. The page title is "RecalculateDiscussion". The workspace is displayed in a formal view, showing a hierarchical tree structure of tasks and discussions. The root node is "Comparative NGS vs Gene Expression Tamoxifen data :: neal". It has several child nodes, including "Work with gene-expression data :: alice", "Augment the genomic data :: alice", "Consider also NGS data :: neal", "Unfamiliar with the technology :: alice", "Provide representative dataset :: neal", "Confront ambiguities :: ismith", "Notable agreement between NGS and GE analyses :: alice", "SD for Tam data :: alice", "SD for Tam RNA-Seq :: neal", and "Important publications :: ismith". The workspace is displayed in a purple header bar with navigation options: "View detail window", "View submission dates", "View message creators", and "Generate Report".

Firefox

http://dicodedev....tml?SpaceID=28224 x RecalculateDiscussion x +

dicodedev.cti.gr/workspace/formal/agora.aspx?ID=28224# Google

View detail window |  View submission dates |  View message creators [Generate Report](#)

- Comparative NGS vs Gene Expression Tamoxifen data :: neal
  - Work with gene-expression data :: alice
    - Augment the genomic data :: alice
  - Consider also NGS data :: neal
    - Unfamiliar with the technology :: alice
      - Provide representative dataset :: neal
      - Confront ambiguities :: ismith
    - Notable agreement between NGS and GE analyses :: alice
      - SD for Tam data :: alice
      - SD for Tam RNA-Seq :: neal
    - Important publications :: ismith

# Collaboration workspace: Other views

Neighbourhood of object titled "Gene Expression Omnibus"

Gene Expression Omnibus  
0.25

SMD is another interesting source (http://www.ncbi.nlm.nih.gov/geo/) The website stores sorted gene expression datasets, as well as original Status and Platform records in the Gene Expression Omnibus (GEO) repository.

0.35

ArrayExpress

0.20

Microarrays

0.20

GEO is highly populated to other full-scale gene expression repositories. This repository will help to see why that this was being the reason.

0.20

GEO includes tools for a full analysis of the data. This might be very handy as we do not know just what tools will be needed.

Decision making view of workspace "Clinico-genomic research: which public repositories to use in order to augment our gene expression datasets?"

MAUT Results - Alternatives Ranking

1. Links to GEO-profile expression data
2. Gene Expression Omnibus (GEO)
3. Stanford Microarray Database (SMD)
4. ArrayExpress
5. Microarrays
6. GEO is highly populated
7. Gene Description
8. Supplements for datasets
9. Support for R
10. Data analysis tools

Set weights  
Details  
Plot

Scoring based on weights:  
0.25 (Item Rating)  
0.35 (Relations)  
0.20 (User Score)  
0.20 (Likes/Dislikes)

AHP Results - Alternatives Ranking

1. Gene Description
2. Microarrays
3. Data analysis tools
4. Links to GEO-profile expression data
5. Support for R
6. Supplements for datasets
7. GEO is highly populated
8. Stanford Microarray Database (SMD)
9. Gene Expression Omnibus (GEO)
10. ArrayExpress

Set weights  
Details  
Plot

LEXIC Results - Alternatives Ranking

1. Links to GEO-profile expression data
2. Gene Expression Omnibus (GEO)
3. Stanford Microarray Database (SMD)
4. Microarrays
5. Gene Description
6. Supplements for datasets
7. GEO is highly populated
8. Support for R
9. ArrayExpress
10. Data analysis tools

Set weights  
Details  
Plot

# Decision Making view

WSM Results - Alternatives Ranking

- 1. Radical prostatectomy
- 2. Active Surveillance
- 3. Brachytherapy for prostate cancer

Set weights

Details

Plot

**Scoring based on weights:**

- 0.25 (Item Rating)
- 0.35 (Relations)
- 0.20 (User Score)
- 0.20 (Likes/Dislikes)

### Lexicographic - Setting priorities

Drag to move most important factors on top

Rating

Relations

UserRating

LikesDislikes

Calculate

Close

### AHP - Setting relative weights

- 1 Features comparison  
Comparing the features
- 2 Alternative comparison  
wrt 'Rating'
- 3 Alternative comparison  
wrt 'Relations'
- 4 Alternative comparison  
wrt 'User Rating'
- 5 Alternative comparison  
wrt 'Likes/Dislikes'

#### Step 1 - Features comparison

| AHP factors comparison     |                     |
|----------------------------|---------------------|
| Items compared             | Relative Importance |
| 1 Score/Relations          | 1/5                 |
| 2 Score/UserRating         | 1/9                 |
| 3 Score/LikesDislikes      | 7                   |
| 4 Relations/UserRating     | 3                   |
| 5 Relations/LikesDislikes  | 9                   |
| 6 UserRating/LikesDislikes | 1/7                 |

Page 1 of 1 1000

Previous

Next

Calculate

Cancel



# The Dicode Workbench

The screenshot displays the Dicode Workbench interface, which is a collaborative workspace for analyzing and discussing medical literature. The main workspace is titled "Prostate cancer: Alternative treatment" and is divided into several thematic panels:

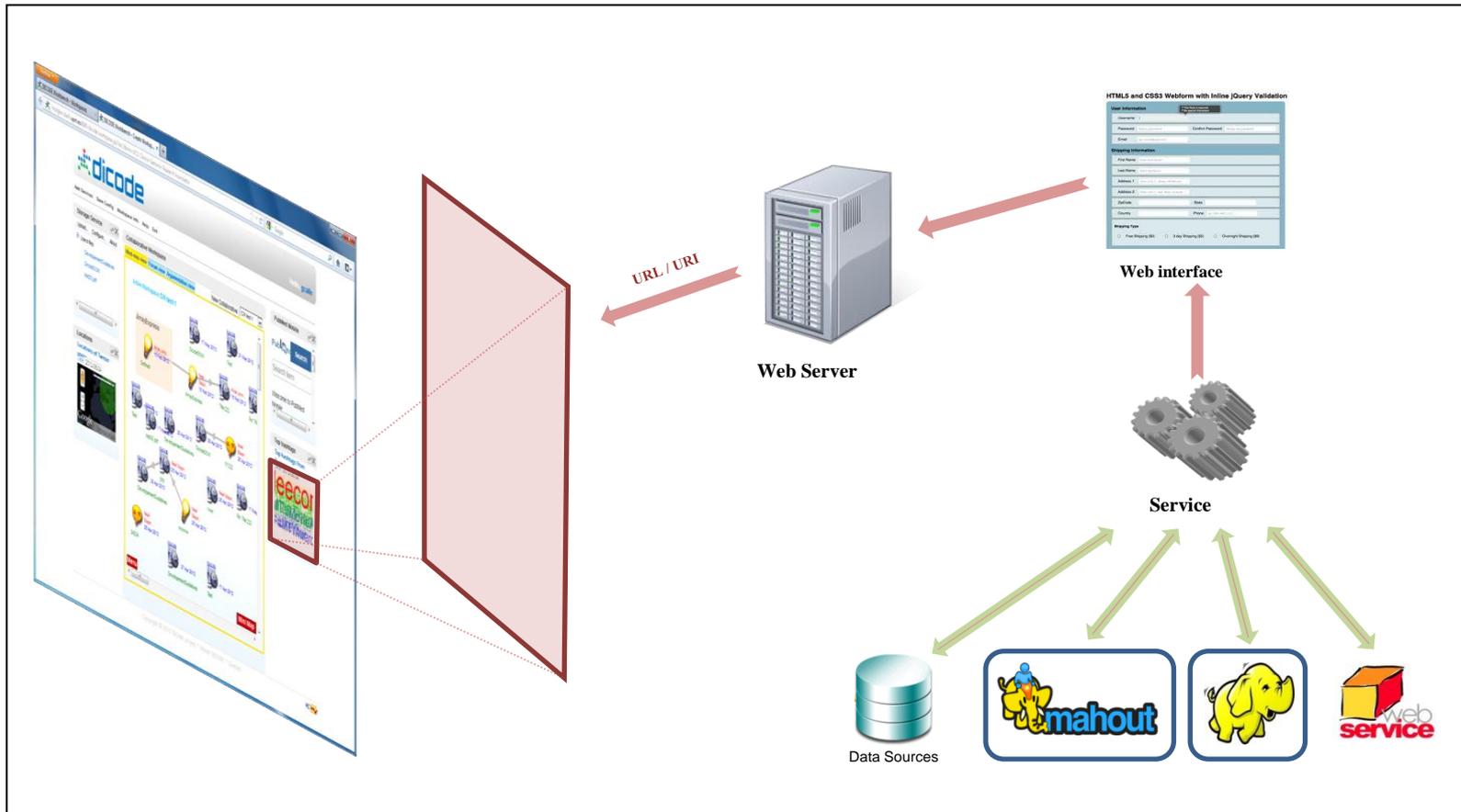
- Relevant resources:** A pink panel containing two PDF documents: "Radical Prostatectomy: Survey" and "Randomized Trials For Active Su...".
- Active Surveillance:** An orange panel showing a network of concepts. A central node "Active Surveillance" is connected to "Urologist" (23 Oct 2012), "Physician" (23 Oct 2012), and "Requires close monitoring". Other nodes include "Avoids radiation side effec..." and "Many/most patients are unre...".
- Radical Prostatectomy:** A yellow panel showing a network of concepts. A central node "Radical prostatectomy" is connected to "Jane Monro..." (23 Oct 2012), "Urologist" (23 Oct 2012), "Accurate staging", and "PSA levels predict recurren...". Other nodes include "Erectile dysfunction" and "Preven solution".
- Brachytherapy:** A blue panel showing a network of concepts. A central node "Brachytherapy for prostate..." is connected to "Jane Monro..." (23 Oct 2012), "Urologist" (23 Oct 2012), and "Urologist" (23 Oct 2012). Other nodes include "No post-treatment".

Surrounding the main workspace are several utility panels:

- Storage Service:** A panel for file management with options for "Upload", "Config...", and "About".
- Locations:** A panel titled "Locations of Twitter users" with a date filter set to "2012-10-28" and a map showing user locations.
- Forum Summariz:** A panel with a "Main Menu" and "All Questions" section, and a "Cross Validated Statistics Forum - Topic Clouds" section.
- Pubmed:** A search window with a search bar, a "Search" button, and a "Doc. Viewer" section for document preview.
- Entity Prominence:** A panel titled "Prominence Graph" with an "ENTITIES" list and a search bar for entities.

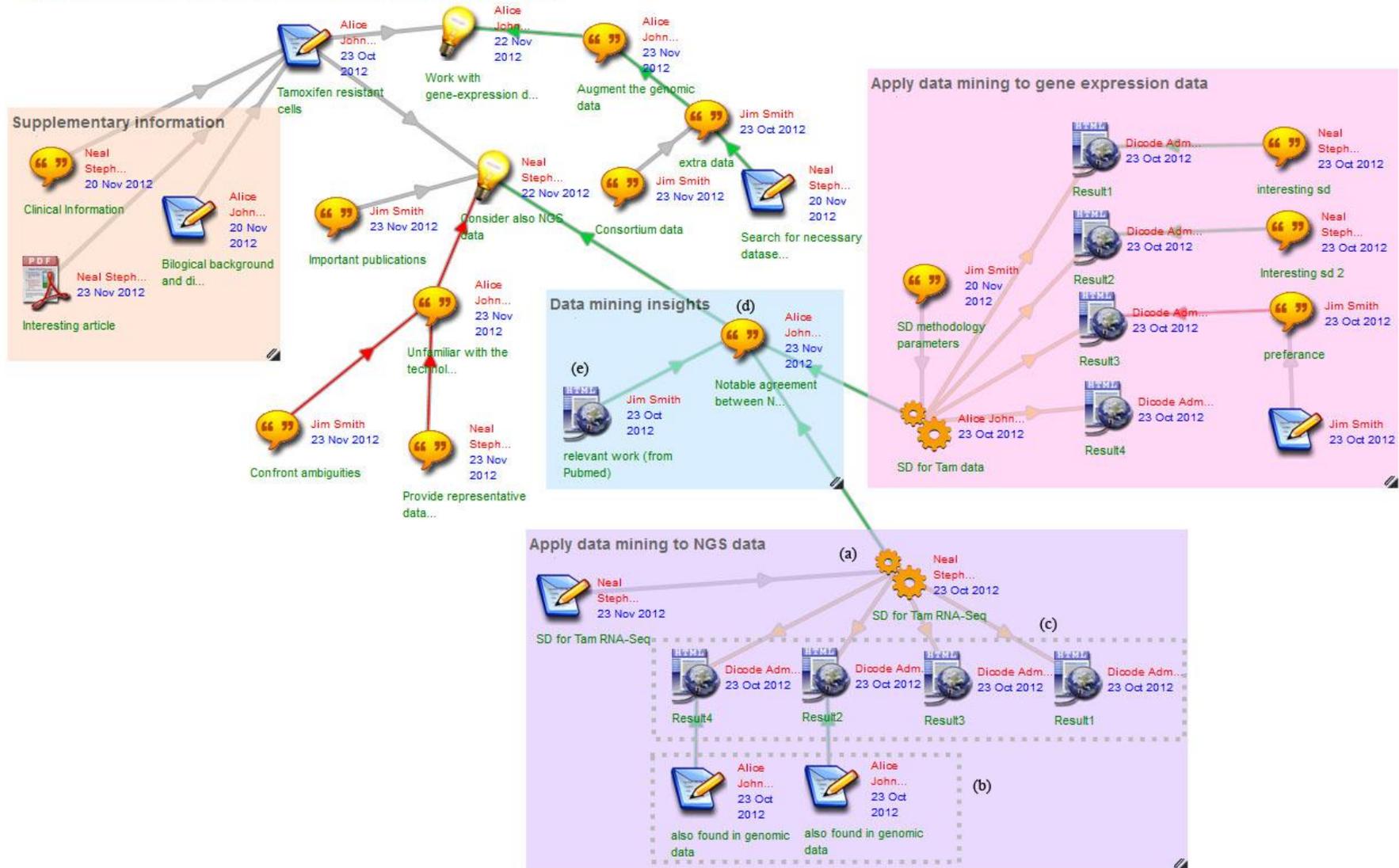
The interface includes a top navigation bar with "Add Services", "Save Config", "Workspace Info", "Help", and "Exit" options. A greeting "Hello, alice" is visible in the top right corner. The footer contains copyright information: "Copyright © 2010 Dicode project \* About DICODE \* Contact" and a "VSC" logo.

# Integrating services (1/2)



# Integrating services (2/2)

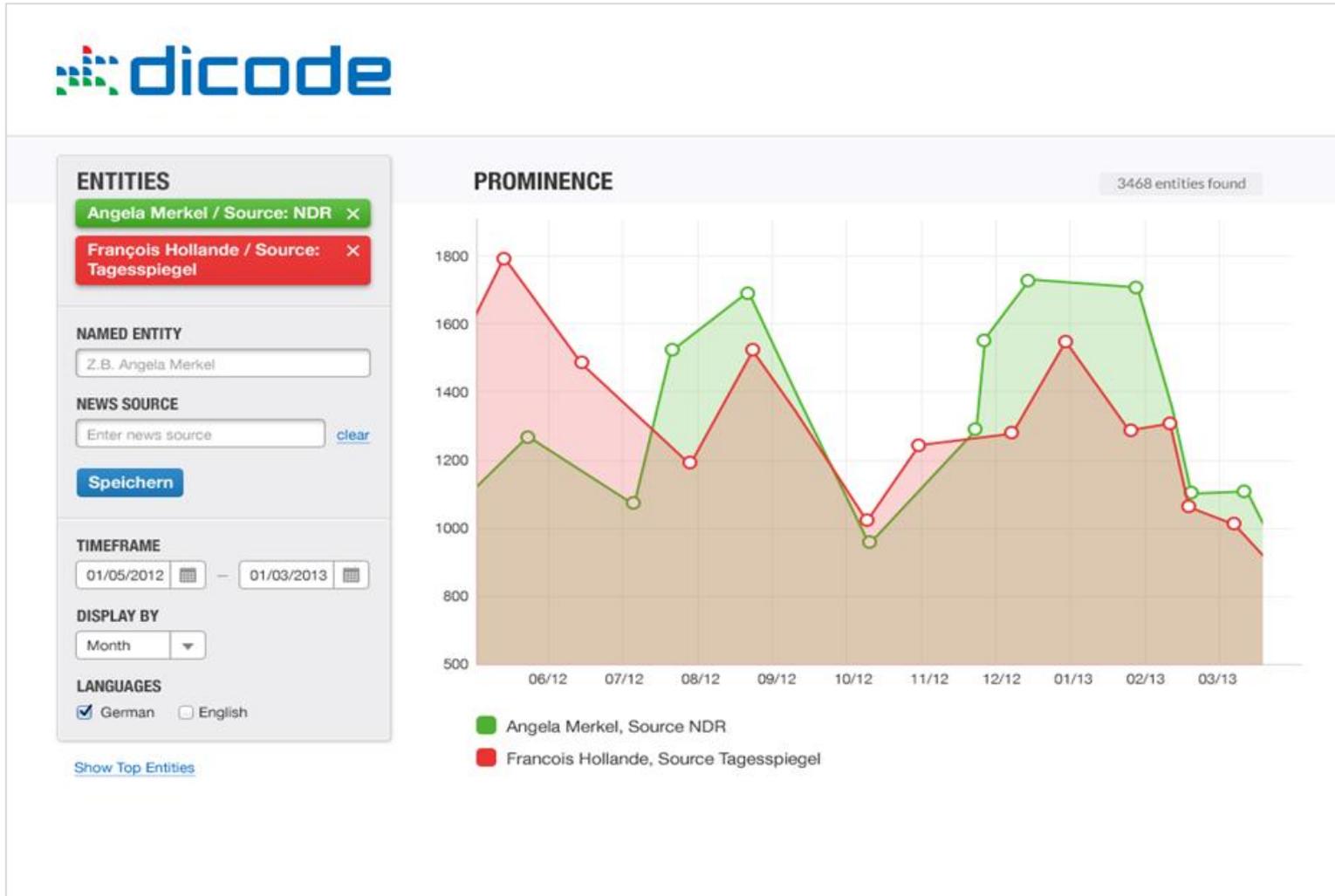
Active Workspace: Comparative NGS vs Gene Expression Tamoxifen data



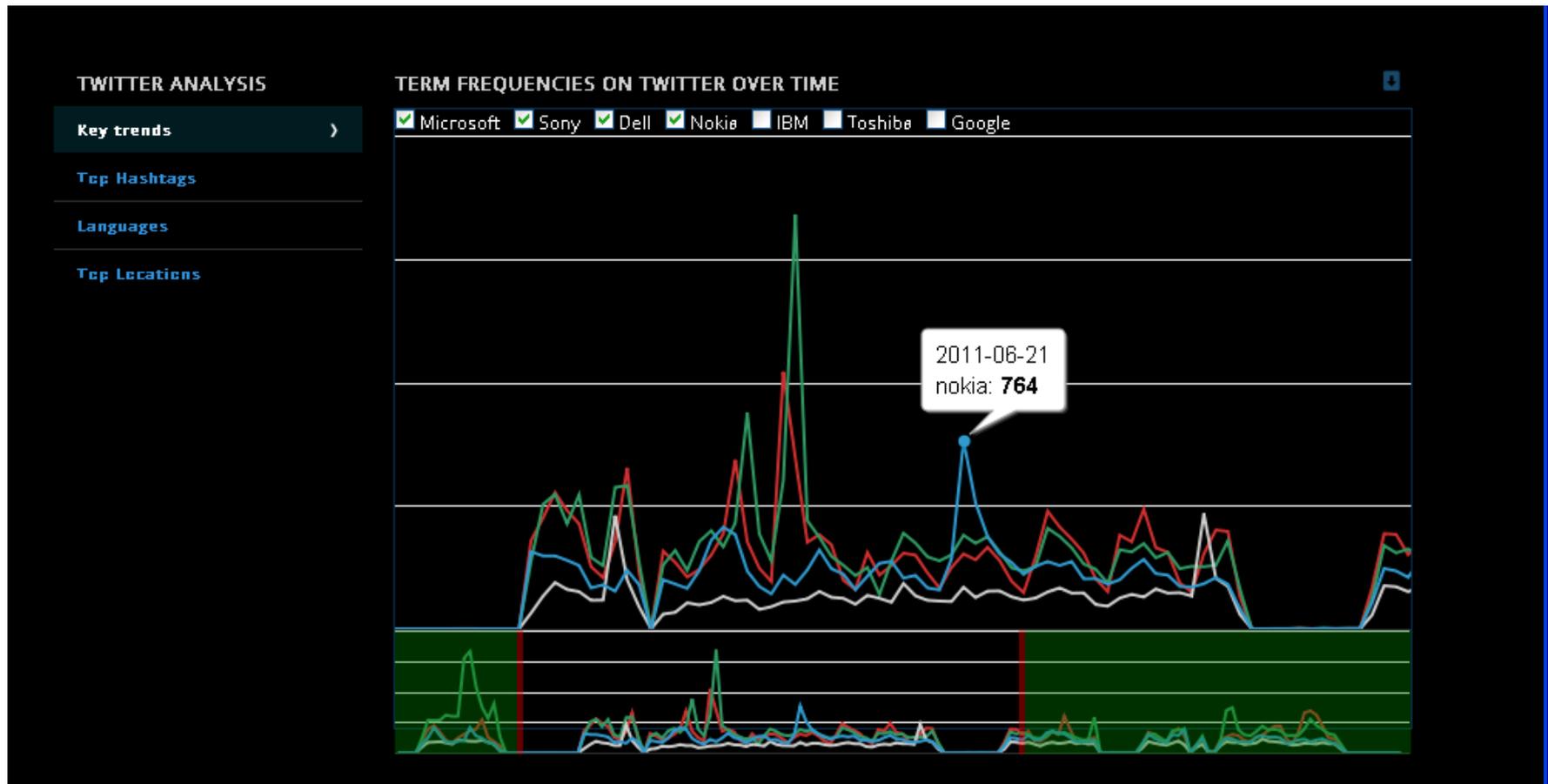
# Data mining services (1/3)

- **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



# 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

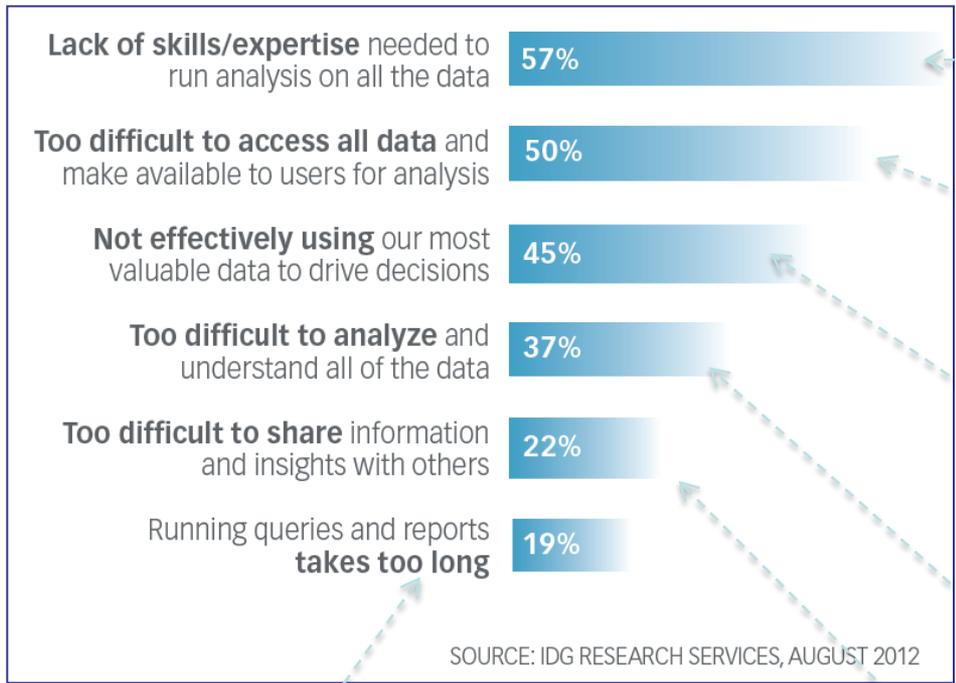
concluding ...



# 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!

