On the identification and integration of countermeasures to cope with data intensiveness in collaboration settings

Manolis Tzagarakis
University of Patras &
Computer Technology Institute
and Press "Diophantus"
26504 Rio Patras, Greece
tzagara@upatras.gr

Spyros Christodoulou
University of Patras &
Computer Technology Institute
and Press "Diophantus"
26504 Rio Patras, Greece
shristod@cti.gr

Nikos Karacapilidis
University of Patras &
Computer Technology Institute
and Press "Diophantus"
26504 Rio Patras, Greece
nikos@mech.upatras.gr

ABSTRACT
This paper reports on an innovative approach that aims to facilitate and augment collaboration and decision making in data-intensive and cognitively-complex settings. Towards shaping our approach, we have first reviewed state-of-the-art Web 2.0 collaboration tools to identify cognitive overload issues that these tools are prone to and explore the countermeasures they adopt to suppress data intensiveness. Exploiting this analysis, our approach minimizes the causes of information overload by integrating appropriate countermeasures. The proposed solution brings together human and machine intelligence and enables an incremental formalization of the collaboration workspace. It incorporates a set of interoperable services that reduce the data-intensiveness and complexity overload to a manageable level, thus permitting stakeholders to be more productive and creative.

Author Keywords
Collaboration; decision making; sense making; data integration; human intelligence; machine intelligence

ACM Classification Keywords
K.4.3 Organizational Impacts: Computer-supported collaborative work; H.4.2 Types of Systems: Decision support (e.g., MIS); H.5.3 Group and Organization Interfaces: Computer-supported collaborative work.

General Terms
Design; Experimentation; Human Factors; Performance.

INTRODUCTION
Information overload is a major problem in today's organizations. While incoming data is rapidly increasing, making sense and filtering what is important for the problem in hand becomes more and more difficult and time consuming. In many cases, the raw information is so overwhelming that stakeholders are often at a loss to know even where to begin to make sense of it. "Big Data" [1] can negatively affect the effectiveness of decision making in an organization [2] and create stress and cognitive overload to its stakeholders [3].

In addition, this data may vary in terms of subjectivity and importance. Admittedly, it is nowadays easier to get the data in than out. The problems start when we want to consider and exploit the accumulated data and meaningfully analyze them towards making a decision. Thus, in complex settings, being able to pull only the relevant information and efficiently share, interpret and use this information for decision making becomes a challenge when the right tools and information systems are missing [4]. When things get complex, we need to identify, understand and exploit data patterns, aggregate large volumes of data from multiple sources, and then mine it for insights that would never emerge from manual inspection or analysis of any single data source. In other words, the pathologies of "big data" are primarily those of analysis.

Generally speaking, information management related tasks need to be streamlined and automated to make information work more productive. In the settings under consideration, the way that data will be structured for query and analysis, as well as the way that tools to handle them efficiently will be designed are of great importance and certainly set a big research challenge. Overall, such innovative solutions have to face two major imperatives: (i) they need to exploit the information growth by ensuring a flexible, adaptable and scalable information and computation infrastructure; new approaches, taking advantage of distributed computing and clusters of computational resources, are needed for structured and unstructured data search, data pre-processing, database analytics, resource pooling, and storage optimization [1]; (ii) they need to exploit the competences of all stakeholders and information workers to meaningfully confront various information management issues [5], such as information characterization, classification, presentation, retention, storage, disposal etc.; in other words, dealing with data-intensive and cognitively complex settings is not a technical problem alone.

CSCW'12, February 11–15, 2012, Seattle, Washington, USA.

To deal with this complex problem in contemporary collaboration and decision making settings, we first attempted to devise a roadmap of features to be incorporated in the suggested solution; towards this, we reviewed representative Web 2.0 collaborative tools. The purpose of our review was twofold: (i) to identify a number of issues that may cause information overload in the settings under consideration, and (ii) to explore the countermeasures each tool adopts to suppress the effects of information overload. Based on this review and adopting current technology advancements, we shaped an innovative approach that exploits the synergy of human and machine reasoning to facilitate collaboration and decision making in the settings under consideration. The proposed solution is being developed in the context of an FP7 EU project, namely Dicode (http://dicode-project.eu/).

CURRENT APPROACHES

The emergence of the Web 2.0 era introduced a plethora of collaboration tools which provide engagement at a massive scale and feature novel paradigms. In an attempt to identify state-of-the-art functionalities that aim at dealing effectively with data-intensive situations in collaboration and decision making settings, we briefly review in this section some representative Web 2.0 tools. In this paper, what we are particularly interested in is not the features of the tools *per se*, but rather their functionalities designed to cope with data-intensive situations. Based on this fact and the extensive list of causes and countermeasures against information overload reported in the work of Eppler and Mengis [6], we analyze collaboration tools categories according to the sources of cognitive overload and the countermeasures taken by each tool. By the term 'source of cognitive overload' we refer to the characteristics of information that may lead to cognitive overload situations in each tool, while by 'countermeasures' to the solutions that the tools make available to remedy cognitive overload.

Web 2.0 tools overview

Classified upon their basic purpose, Web 2.0 tools deal with mind mapping, file sharing and collaborative editing, social networking, note taking and annotation, project/task management, and argumentative collaboration.

Mind mapping tools permit the representation of ideas and concepts that can be connected to form diagrams. Representative tools of the category include MindMeister, Mindomo, Bubbl.us and Xmind. MindMeister (http://www.mindmeister.com/) enhances collaboration in data intensive environments by providing awareness mechanisms (notifications via emails and SMS), the focus (zoom in/out) feature to browse and work on maps with a large number of topics, the feature of expanding/collapsing the subtopics of a topic, the filtering feature that isolates part of the map and the "history view" of a map displaying the entire history since the creation of a mind map. Mindomo (http://www.mindomo.com/) supports collapsing/expanding of user selected parts of the map, "tagging" on a topic's visualization by using a number of specific icons ("tags"), zooming in/out on a large map, the history of a mind map, the filtering tool to isolate part of the map by using various criteria and the search tool. Bubbl.us (http://www.bubbl.us/) provides a zoom in/out tool which may be efficiently used to help browsing and scrolling in maps that include many bubbles. Coediting of maps is also possible resulting in sharing maps among the members of a group of users. XMind (http://www.xmind.net/) supports a filtering mechanism (by selecting "markers" or "labels"). The extending/collapsing feature of the subtopics of a topic may also be useful when dealing with maps containing a large number of topics. Boundaries are used as a topic aggregation mechanism to which info concerning the state of the topic can be attached.

File sharing tools enable the distribution and facilitate access of digital information in the form of files. *Collaborative editing tools* (that permit the joint authoring of documents via individual contributions), such as wikis, also fall in this category. Popular tools include Dropbox, Humyo, Box.net, Google Docs, MediaWiki, Confluence and PbWorks. DropBox (http://www.dropbox.com/) handles information overload issues by providing version history, file recovery, online list to all events having taken place and awareness mechanisms. Humyo.com (http://www.humyo.com/) provides online teamspaces and a filtering mechanism. Box.net (http://www.box.net/) offers document versioning, file and folder tagging and filtering mechanisms. Google Docs (http:// docs.google.com/) supports history revision, searching, sorting based on file type and tagging for all saved documents. MediaWiki (http://www.mediawiki.org/wiki/MediaWiki) provides awareness features (watchlists), page history, versioning and access control mechanisms, hierarchical organization of the Wiki pages and a search mechanism. Confluence (http://www.atlassian.com/software/ confluence/) supports notifications, separate spaces, access control mechanism, page categorization and tagging, searching and a notification mechanism to avoid information overloading. PBworks (http://pbworks.com/) enhances Wiki versioning and notifications, page tagging, workspace exporting (to a zip file) and a keyword-based search mechanism.

Social networking tools aim at building social networks or social relationships through which users may share interests and activities. Representative tools of this category include Facebook, MySpace, LinkedIn and Twitter. Facebook (http://www.facebook.com/) deals with the enormous number of its members and the vast amount of information contributed by incorporating the "news feed" feature, notifications and Facebook groups which consist of people sharing common interests. MySpace (http://www.myspace.com) incorporates group management features, notifications by email upon other users posting comments or uploading content and a search mechanism. LinkedIn (http://www.linkedin.com) has also implemented group management features and a notification mechanism that makes users aware of recent events in their networks (such as new connections). Twitter (http://www.twitter.com/) displays messages in reverse chronological order and enables organizing messages via hashtags. In addition, many third party tools are available to organize and filter messages.

Note taking and annotation tools deal with the creation of shared comments, notes, explanations or other types of remarks that can be attached to any part of a resource. Representative tools of this category include Zoho Notebook, EverNote and SimpleNote. Zoho Notebook (http://notebook.zoho.com) supports collaboration in data intensive environments with Zoho NoteBook groups that may be used to allow sharing content among specific users. Versioning of shared notebook content allows keeping track of changes and modifications. In Evernote (http://www.evernote.com/), notes are organized in notebooks. Each note may be tagged and organized in folders. A search mechanism is also available for spotting a desired note. SimpleNote (http://simplenoteapp.com/) provides versioning on a user's notes, note searching, tagging, and prioritization based on the importance of a note.

Project and task management tools coordinate projects by managing all related resources and enforcing all necessary constraints. Representative tools in this category include: Basecamp (http://basecamphq.com/), which attempts to deal with data-intensive issues by providing features such as awareness mechanisms, filters and milestone "zoom-in"; ActiveCollab (http://www.activecollab.com/), which enables filtering of task assignments based on various criteria that include date, priority, assignees and project; Redmine (http://www.redmine.org/), which manages the complexity of projects by enabling the creation of sub-projects and automates functionalities such as the Gantt charts and calendars that are calculated based on the start and due time of issues; awareness services are also available by publishing project activities, news, issues, and issue changes either as email or as atom feeds.

Argumentative collaboration tools deal with the creation of argumentative discourses where a group of people exchange positions and arguments in order to achieve consensus on the issue discussed. Representative Argumentative collaboration tools include Araucaria, DebateGraph, Compendium, CoPe_it! And Cohere. Araucaria (http://araucaria.computing.dundee.ac.uk/doku.php) enables argument analysis through diagrams. Araucaria has been designed to enhance teaching and critical thinking [7] but does not include mechanisms to cope with data intensiveness. In DebateGraph (http://debategraph.org/), several mechanisms support large scale argumentation and collaboration: a "history" mechanism, a progressive visualization of the argumentation map, awareness and search mechanisms, ability to traverse through different views of a map. Compendium (http://compendium.open.ac.uk/) deals with data intensive environments by supporting multiple level maps and a zoom in/out tool. "Aerial" view, the search mechanism, multiple-dragging nodes and node bookmarking are helpful in large maps. CoPe_it! (http://copeit.cti.gr/) supports a number of features to enhance collaboration in data intensive cases. For instance, the "minimap" of a workspace provides an overview of its contents. Also, there is a "review/history" mechanism, through which one may follow the evolution of a workspace. In addition, multiple items may be grouped together. Finally, a filtering mechanism enables one to view argumentation items fulfilling specific criteria, such as the item's title, date, author and type. Finally, Cohere (http://cohere.open.ac.uk/) copes with data intensive situations through various filtering mechanisms (e.g. ideas and their connections can be filtered according to their type), the tagging of ideas and the search mechanism.

Causes and countermeasures of information overload in Web 2.0 tools

In order to devise a roadmap towards the identification of the appropriate collaboration technologies and functionalities, we categorize the tools discussed in the previous sections according to (i) the cognitive overload issues they are prone to, and (ii) the countermeasures that these tools introduce to overcome cognitive overload issues. These dimensions are important for our approach in the context of the Dicode project, as they outline the solution space not only in terms of objectives and functionalities, but also in terms of technologies to remedy cognitive overload and data-intensiveness issues.

Based on the list of causes and countermeasures concerning information overload as reported in the work of [6], we identified in our review the following sources of cognitive overload that need to be addressed in a collaboration tool: *(i) rising number of information:* the information items brought into collaboration increase as the collaboration proceeds; such increase may not be gradual but may appear in bursts; *(ii) uncertainty of information:* the inability to assess quickly the relevance of the available information; *(iii) information diversity and increasing number of alternatives:* the situation in which diverse types of information exist and the number of solutions increases as the collaboration proceeds; *(iv) ambiguity of information:* the situation where information can be interpreted in several ways; *(v) complexity of information:* the degree of interrelationships of information; *(vi) intensity of information:* the importance of particular information items; *(vii) increase of information dimensions:* the situation in which the way the available information brought in during collaboration can be combined with an increasing number of other items or can be considered along different aspects and dimensions; *(viii) information quality and value:* the degree of worth of information, and *(ix) overabundance of* *irrelevant information:* the excessive amount of irrelevant information which leads to a low signal/noise ratio of the items in the collaboration space.

Causes of information overload for each category of collaboration tools are summarized in Table 1. Concerning the features that each tool provides to suppress data-intensiveness, we focus on the following countermeasures [6]: *(i) structuring information:* features concerning interlinking, aggregating, annotating of information items; *(ii)visualization:* refers to tools providing visual representations of information such as graphs; *(iii)formalization:* to allow formality to vary within the representations of information items; *(iv) simplicity*; *(v)customization and personalization:* provide user-defined ability to depict information items; *(vi) levels of detail / summaries*; *(vii) awareness:* mechanisms such as notification, history, and versioning; *(viii) search and filtering:* based on criteria or features of information items; *(ix) quality filters:* intelligent agents or Decision Support Systems. Countermeasures taken by collaboration tools to suppress information overload are summarized in Table 2.

Implications for Dicode

From the analysis of Table 1, it results that a plethora of collaboration technologies is available, each of them aiming to support different objectives. Analyzing the tools with respect to the services they provide, it is evident that although tools belonging in the same category provide a common core set of services, they integrate services that are not typical for their category. While elaborating on the issue of sources of cognitive overload, similar concerns were revealed. However, the analysis also showed that each tool attempts to address the related data intensiveness and cognitive overload issues by introducing particular services or approaches, which aim at alleviating the severe consequences. In the same line, each category of tools favours particular cognitive overload countermeasures (Table 2), which are explicitly designed to address the problems that occur in a particular collaboration context. When each tool is used independently, the available countermeasures may provide the required support to address information overload issues. However, when an integrated approach must be considered, i.e. when two or more tools have to be deployed to address collaboration needs, the countermeasures may be insufficient and of limited use. This is mainly due to the fact that the countermeasures of each tool have a particular scope which is derived from the collaboration objective. Hence, tools which belong to different categories but exhibit common countermeasures conceive them in different terms, thus raising concerns on how to consider them when these tools have to be jointly used.

In Dicode, signals are strong that such an integrated approach to collaboration is required. In particular, argumentative collaboration, collaborative editing, note taking and mind mapping tools look promising to address the foreseen collaboration needs. Yet, these tools must be considered in an integrated manner. In our approach, we envisage collaboration tools grafted with effective cognitive overload countermeasures, which do not limit their focus to particular collaboration objectives but provide their services in situations where heterogeneous collaboration tools must interoperate.

Concerning the countermeasures taken by the Web 2.0 tools, an “ideal” tool to perform collaboration and decision making should first adopt as many of them as possible, in an attempt to restrict the causes of information overload. One would expect such a tool to provide information structuring mechanisms, visual information representation, information formalization, customization and personalization schemas, awareness mechanisms, searching, quantity and quality filtering as well as intelligent information management systems for fostering an easier prioritization of information [8].

THE PROPOSED APPROACH

According to the above analysis, current Web 2.0 collaboration tools exhibit two important shortcomings making them prone to the problem of information overload. First, these tools are “information islands”, thus providing only limited support for interoperation, integration and synergy with third party tools. Second, Web 2.0 collaboration tools are rather passive media; they lack reasoning services with which they could meaningfully support the collaboration.

As far as existing decision making support technologies are concerned, data warehouses and online analytical processing have been broadly recognized as technologies playing a prominent role in the development of current and future DSS [9]. However, there is still room for further developing the conceptual, methodological and application-oriented aspects of the problem. One critical point that is still missing is a holistic perspective on the issue of decision making. This originates out of the growing need to develop applications by following a more human-centric (not problem-centric) view, in order to appropriately address the requirements of the contemporary, knowledge-intensive organization’s employees.

The proposed approach will advance decision making support technologies by adopting a knowledge-based decision-making view, enabled by the meaningful accommodation of the results of the data mining processes. In such a way, the decision making process is able to produce new knowledge, such as evidence justifying or challenging an alternative or practices to be followed or avoided after the evaluation of a decision. Knowledge management activities such as knowledge elicitation, representation and distribution influence the creation of the decision models to be adopted, thus enhancing the decision making process [10].

	Rising number of information items	Uncertainty of information	Info diversity & increasing no of alternatives	Ambiguity of information	Complexity of information	Intensity of information	Increasing dimensions of information	Information quality, value	Overabundance of irrelevant information
MindMaps	x	x		x	x		x		x
Collaborative editing	x	x		x		x		x	
Social Networking	x					x			x
Note taking	x							x	
Project/task management	x				x				
Argumentative collaboration	x	x	x	x	x	x	x	x	x

Table 1: Causes of information overload for each category of collaboration tools.

	Structuring information	Visualization (graphs)	Formalization	Simplicity	Customization & personalization	Levels of detail / summaries	Awareness	Search & filtering	Quality filters (intelligent agents / DSS)
Mind mapping	1, 2, 3, 4	1, 2, 3, 4	1, 2, 3, 4	1, 2, 3, 4		1, 2, 3			
Collaborative editing	5, 6, 7, 8, 9, 10, 11			5, 6, 7, 9, 11	9, 10	5	5, 7, 8, 9, 10, 11	6, 7, 8, 9, 10, 11	
Social networking		12	12, 13, 14	12, 13, 14	12, 13, 14		12, 13, 14	12, 13, 14	
Note taking and annotation tools	16, 17	15						16, 17	
Project/task management	20		18, 19, 20			19	18, 19, 20	18, 19, 20	
Argumentative collaboration	22, 23, 24, 25	21, 22, 23, 24, 25	21, 22, 23, 24, 25			22, 23, 24, 25	22, 24	21, 23, 24	21, 22, 23, 24,25 25

Table 2: Countermeasures taken by collaboration tools.

Numbers correspond to tools according to the following mapping: 1:MindMeister, 2:MindDomo, 3:Bubbl.us, 4:XMind, 5:Drop box, 6:Humyo.com, 7:Box.net, 8:Google Docs, 9:MediaWiki, 10:Confluence, 11:PBWorks, 12:Facebook, 13:MySpace, 14:LinkedIn, 15:Zoho Notebook, 16:Evernote, 17:SimpleNote, 18:Basecamp, 19:ActiveCollab, 20:Redmine, 21:Araucaria, 22:DebateGraph, 23:Compendium, 24:CoPe_it!, 25:Cohere

Our approach builds on a conceptual framework where formality and the level of knowledge structuring during argumentative collaboration is not considered as a predefined and rigid property, but rather as an adaptable aspect that can be modified to meet the needs of the tasks at hand. By the term formality, we refer to the rules enforced by the system, with which all user actions must comply. Allowing formality to vary within the collaboration space, incremental formalization, i.e. a stepwise and controlled evolution from a mere collection of individual ideas and resources to the production of highly contextualized and interrelated knowledge artifacts, can be achieved.

Dicode views

In our approach, views enable the incremental formalization of collaboration. A view can be defined as a particular representation of the collaboration space, in which a consistent set of abstractions able to solve a particular organizational problem during collaboration is available. Our approach enables the switching from a view to another, during which abstractions of a certain formality level are transformed to the appropriate abstractions of another formality level. This transformation is rule-based; such rules can be defined by users and/or the facilitator of the collaboration and reflect the evolution of a community's collaboration needs. According to our approach, it is up to the community to exploit one or more views of a collaboration space (upon users' needs and expertise and/or the overall collaboration context). The foreseen collaboration views are described below:

Discussion-forum view: In this view, a collaboration space is displayed as a traditional web-based forum, where posts are displayed in ascending chronological order. Users will be able to post new messages to the collaboration space, which will appear at the end of the list of messages. Posts may also have attachments to enable the uploading of files. This view also provides support for sharing of resources.

Mind-map view: In this view, a collaboration space is displayed as a mind map, where users can "interact" with the items uploaded so far. The map deploys a spatial metaphor permitting the easy movement and arrangement of items on the collaboration space. In this view, information triage (i.e. the process of sorting and organizing through numerous relevant materials and organizing them to meet the task at hand [11]) is supported. During such a process, users can effortlessly scan, locate, browse, update and structure knowledge resources that may be incomplete, while the resulting structures may be subject to rapid and numerous changes. Items posted on the collaboration space in this view can be one of the following types: "idea", "comment", "note" and "generic". The type "generic" will be used in case the semantics of an item is unclear. Files of any content type (e.g. pdf, jpg) can be also uploaded on the collaboration space. The mind-map view will also provide a set of mechanisms through which: (i) items on the collaboration space can be related, and (ii) new abstractions can be created. Creation of relationships between items will be facilitated by drawing directed arrows between items on the collaboration space. Visual characteristics of the arrows connecting items, such as their color and thickness, may convey meaningful semantics. Mind-map view also provides abstraction mechanisms that allow the creation of new abstractions out of existing ones. Abstraction mechanisms include: (i) annotation and metadata (i.e. the ability to annotate instances of various knowledge items and add or modify metadata); (ii) aggregation (i.e. the ability to group a set of instances of knowledge items so as to be handled as a single conceptual entity; this may lead to cases where a set of knowledge items can be considered separately, but still in relation to the context of a particular collaboration); (iii) generalization/specialization (i.e. the ability to create semantically coarse or more detailed knowledge items in order to help users manage information pollution of the collaboration space); (iv) patterns (i.e. the ability to specify instances of interconnections between knowledge items of the same or a different type, and accordingly define "collaboration templates"). The mind-map view aims at supporting sense-making during cognitive complex tasks.

Formal view: The formal view of the collaboration space will permit only a limited set of discourse moves for a limited set of message types whose semantics is fixed. In particular, the formal view will enable the posting of messages which can be of type "issue", "alternative" or "position". The type "issue" will be used to indicate decisions to be made, while "alternative" will be used to represent potential solutions to the issue being discussed. The type "position" will be used for messages that comment either on alternatives or on other positions. Positions will either support or be against alternatives and positions and their relationship will be explicitly specified when users post them to the collaboration space. Files can be attached to positions to further support their validity. The formal view will also support the notion of preferences. A preference is used to weigh the importance of two positions and reflect the importance of one position over another. Specialized decision making support algorithms (e.g. a voting algorithm with equal weights to all or a multiple criteria decision making algorithm), which are associated with the collaboration space, will take into consideration the relationships of positions as well as existing preferences and calculate which alternative is currently prevailing or which position has been defeated. The aim of the formal view is to make the collaboration space machine-understandable and further support decision making.

THE DICODE WORKBENCH

A particular goal of the Dicode project is to support researchers in collaboratively solving problems and making decisions on complex scenarios with very large, potentially conflicting and incomplete amounts of information. Dicode aims to achieve this by providing an integrated environment, where machine and human reasoning capabilities coexist, interplay and augment each other. The Dicode Workbench, an innovative web-based tool that is described in this section, offers such an integrated environment. Through a typical scenario of collaboration among researchers engaged in addressing a clinico-genomic research issue, we present how the Dicode Workbench may address the issues under consideration.

Figure 1 presents the main user interface of the Dicode Workbench, which integrates the computational and collaborative services needed to address the issue under consideration. A widget-based approach has been adopted in this tool. Each widget provides different functionalities. In particular, the section 'Sources' (top left side of Figure 1) lists all resources that researchers require. The section 'Services' lists computational tools that can be invoked to process data sets and produce results. These services can be references to Web services or other applications. Selecting an item in the 'Services' section results in executing the selected service or application. The section 'Processing results' displays the outcomes of the executed services and may include scatter plots, heatmap plots and descriptive statistics of the analyzed data.

The section 'Search' (top right side of Figure 1) provides a single interface enabling the team to search a list of public biomedical repositories. The results of such search activities can be added to the 'Sources' section of the workbench by simply dragging them from the search results. The section 'Experiment Info' displays metadata about the workbench. At the center of the Dicode Workbench is the 'Dicode Logbook', which enables the argumentative collaboration between researchers.

Figure 1 The Dicode workbench (forum view).

To get more specific, imagine two researchers, Jim and Alice, who aim to investigate which genes or groups of genes are associated with breast cancer disease. Initially, they create a new collaboration session (logbook), where they exchange ideas related to which data sources to use, based on their own data analysis experience and literature knowledge. They search relevant literature using the appropriate search services. Jim has conducted an initial analysis with some in-house gene-expression datasets but his findings were not very encouraging, which was attributed to the small sample size (i.e. number of patients) available. He informs Alice about it and suggests potential solutions. The discussion proceeds and finally, in order to overcome the limited sample size problem, they decide to augment their samples with publicly available gene-expression data. After deciding what data to use, they start a new collaboration in order to discuss how the data will be processed. Both researchers suggest solutions and comment on each suggestion and finally decide to use the normalized data for each platform and the UniGene annotation database (http://www.ncbi.nlm.nih.gov/unigene) to uniformly map all genes. Jim and Alice aim at identifying novel or already reported groups of genes associated with breast cancer disease and at comparing the findings of the chosen methodologies to those of the simple analysis conducted by Jim. Both researchers can execute the available services and retrieve the results of the invoked tool. Once the results are available, they engage into interpreting the results in terms of the initial research question.

As the initial goal of Jim and Alice was to accumulate a critical mass of relevant resources, they first create a new collaboration workspace (logbook) and start using it in the *forum view* (Figure 1). The forum view primarily aims to effortlessly collect and share the available resources. During this collaboration phase, Jim and Alice upload available resources and assess them informally, by briefly commenting on them. When many resources start appearing in the forum view (Jim was away for a number of days), Alice decides to switch to a *mind-map view* (where she can better manage the numerous resources). When Jim comes back, being at a loss to realize the evolution of the collaboration, he uses the provided awareness mechanisms (the 'Replay' tool to check step-by-step the progress of the collaboration, the 'Head Up Display' to browse through the actions performed by Alice and the 'Chat' tool to talk to Alice). In the mind-map view, Jim and Alice may organize the available items in more advanced ways and exploit dedicated item types such as ideas, notes and comments (Figure 2). The provided 'mini-map' may help them easily explore the mind-map area which, as collaboration evolves, becomes larger and larger.

In the mind-map view, *ideas* stand for items that deserve further exploitation; they may correspond to an alternative solution to the issue under consideration and they usually trigger the evolution of the collaboration. *Notes* are generally considered as items expressing one's knowledge about the overall issue, an already asserted idea or note. Finally, *comments* are items that usually express less strong statements and are uploaded to express some explanatory text or point to some potentially useful information. Adding metadata (tags) on an item and voting for an item are also supported to enhance searching and filtering functionalities.

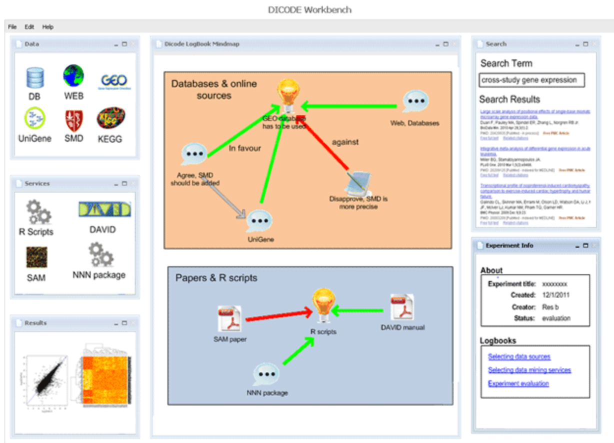

Figure 2 The Dicode workbench (mind-map view).

All the above items can be interrelated by trouble-free actions (as in the case of their creation and uploading, such actions are performed through the mouse). When interrelating items, Jim and Alice may select the colour of the connecting arrow and provide (if they wish) a legend describing the interrelationship they conceive. These legends are intentionally arbitrary. The visual cues of the arrows bear well-defined semantics: for instance, green arrows declare support, whereas red ones declare opposition. Furthermore, the thickness of the arrow may express how strong a resource/idea may object/support an item. Another visual cue that appears in Figure 2 concerns the coloured rectangles that have been created by Jim and Alice to group/cluster related items. Items on a rectangle, along with the interrelations amongst them, may be grouped together to form a new “compound” item.

Although at this collaboration instance these rectangles are simply visual conveniences, they may play an important role during the switch to a more formal projection, enabling the implementation of appropriate abstraction mechanisms. By using the mind-map view, Jim and Alice can transform the resources from a mere collection of items into coherent knowledge structures that facilitating sense making on the available resources. By using the search facilities of the workbench, they are also able to search for relevant literature or data sets, which can be also uploaded on the collaboration logbook (by drag-and-drop). Searching is also possible in the context of the collaboration logbook, while tags added on the items of the mind-map view may be also used to refine the search results.

Jim and Alice may need to further elaborate the knowledge items considered so far, and exploit additional functionalities to advance their argumentative collaboration towards reaching a decision. Such functionalities can be provided by the formal view that enables the semantic annotation of knowledge items, the formal exploitation of collaboration items patterns, and the deployment of appropriate formal argumentation and reasoning mechanisms (Figure 3). While a mind map view aids the exploitation of information by Jim and Alice (i.e. makes the logbook mainly human-interpretable), a *formal view* aims mainly at the exploitation of information by the machine (i.e. makes the logbook mainly machine-interpretable). A formal view provides a fixed set of discourse element and relationship types, with predetermined, system interpretable semantics. In particular, *issues* correspond to problems to be solved, decisions to be made, or goals to be achieved. For each issue, both users may propose *alternatives* (i.e. solutions to the problem under consideration) that correspond to potential choices. *Positions* are asserted in order to support the selection of a specific course of action (alternative), or avert the users’ interest from it by expressing some objection. A position may also refer to another (previously asserted) position, thus arguing in favor or against it.

Figure 3 The Dicode workbench (formal view).

By switching from the mind map to the formal view, existing item types are transformed, filtered out, or kept “as-is” based on a specific set of rules. These rules also take into consideration the item’s visual cues. In addition, the formal view integrates a reasoning mechanism that determines the status of each discourse entry, the ultimate aim being to keep Jim and Alice aware of the discourse outcome. Jim and Alice may continue their collaboration in this formal view; each time an element is added to the discussion, this triggers the underlying reasoning mechanism which informs the team about the most prominent (at this collaboration instance) solution. Using the formal view, Jim and Alice receive active support from the system to make a decision concerning the most appropriate resources for their research. Switching from formal to informal views (i.e. forum or mind-map views) is also possible.

CONCLUSION

Dealing with data-intensive and cognitively complex settings is not a technical problem alone. To deal with the related issues in collaboration and decision making settings, our approach brings together the reasoning capabilities of the machine and the humans. It can be viewed as an innovative workbench incorporating and orchestrating a set of interoperable services that reduce the data-intensiveness and complexity overload at critical decision points to a manageable level, thus permitting stakeholders to be more productive and concentrate on creative activities. In addition, our approach builds on a conceptual framework where formality and the level of knowledge structuring during argumentative collaboration is not considered as a predefined and rigid property, but rather as an adaptable aspect that can be modified to meet the needs of the tasks in hand.

ACKNOWLEDGEMENT

This publication has been produced in the context of the EU Collaborative Project "DICODE - Mastering Data-Intensive Collaboration and Decision" which is co-funded by the European Commission under the contract FP7-ICT-257184. This publication reflects only the author's views and the Community is not liable for any use that may be made of the information contained therein.

REFERENCES

1. Economist. A special report on managing information: Data, data everywhere. Economist.(2010)
2. Schick, A. Information overload: A temporal approach, Accounting, Organizations and Society, vol. 15, no. 3, (1990), pp.199-220.
3. Kirsh, D.: A Few Thoughts on Cognitive Overload, intellectica, (2000), vol. 1, no. 30, pp.19-51.
4. Batra, D.: Cognitive complexity in data modeling: causes and recommendations, Requir. Eng., vol. 12, (2007), pp.231-244.
5. Beardsley, S. C., Johnson, B. C. and Manyika, J. M. Competitive advantage from better interactions McKinsey Quarterly. (2006).
6. Eppler, M.J. and Mengis, J. The Concept of Information Overload: A Review of Literature from Organization Science, Accounting, Marketing, MIS, and Related Disciplines. The Information Society, Vol. 20, (2004), pp. 325–344.
7. Rowe, G. W. A., Mancagno, F., Reed, C. A. and Walton, D. "Araucaria as a tool for diagramming arguments in teaching and studying philosophy", Teaching Philosophy, 29:2, (2006), pp.111-124. ISSN: 0145 5788
8. Eppler, Martin & Mengis, Jeanne. "Side-effects of the E-Society: The Causes of Information Overload and Possible Countermeasures", in Proceedings of the IADIS International Conference on e-Society, (2004), pp. 1119-1124.
9. Shim JP, Warkentin M, Courtney JF, Power DJ, Sharda R, Carlsson C Past, Present and Future of Decision Support Technology. Decision Support Systems vol. 33, (2002), pp. 111-126.
10. Bolloju N, Khalifa M, Turban E. Integrating Knowledge Management into Enterprise Environments for the Next Generation Decision Support. Decision Support Systems (2002), vol. 33, pp. 163-176,
11. Marshall, C., Shipman, F. Spatial Hypertext and the Practice of Information Triage. In Proceedings of the 8th ACM Conference on Hypertext, Southampton UK, (1997), pp. 124–133, ACM Press, New York