

Mastering Data-Intensive Collaboration and Decision Making: The Dicode Project

Nikos Karacapilidis^(✉)

Computer Technology Institute and Press “Diophantus”,
University of Patras, 26504 Rio Patras, Greece
nikos@mech.upatras.gr

Abstract. Many collaboration and decision making settings are nowadays associated with huge, ever-increasing amounts of multiple types of data, which often have a low signal-to-noise ratio for addressing the problem at hand. The Dicode project aimed at facilitating and augmenting collaboration and decision making in such data-intensive and cognitively-complex settings. To do so, whenever appropriate, it built on prominent high-performance computing paradigms and proper data processing technologies to meaningfully search, analyze and aggregate data existing in diverse, extremely large, and rapidly evolving sources. At the same time, particular emphasis was given to the deepening of our insights about the proper exploitation of big data, as well as to collaboration and sense making support issues. This chapter reports on the overall context of the Dicode project, its scientific and technical objectives, the exploitation of its results and its potential impact.

1 Introduction

Individuals, communities and organizations are currently confronted with the rapidly growing problem of information overload [1]. An enormous amount of content already exists in the digital universe (i.e. information that is created, captured, or replicated in digital form), which is characterized by high rates of new information that is being distributed and demands attention. This enables us to have instant access to more information (that is of interest) than we can ever possibly consume. As pointed out in a recent IDC’s White Paper [2], the amount of information created, captured, or replicated exceeded available storage for the first time in 2007, while the digital universe is expanding by a factor of 10 every five years.

People have to cope with such a diverse and exploding digital universe when working together; they need to efficiently and effectively collaborate and make decisions by appropriately assembling and analyzing enormous volumes of complex multi-faceted data residing in different sources [3–5]. For instance, imagine:

- A community of clinical researchers and bio-scientists, supported in their scientific collaboration by a system that allows them to easily examine and reuse heterogeneous clinico-genomic data and information sources for the production of new insightful conclusions or the formation of reliable biomedical knowledge, without having to worry about the method of locating and assembling these huge quantities of data (clinical and genomic data, molecular pathways, DNA sequence data, etc.).

- Or a community of clinicians, radiologists, radiographers, patients and pharmaceutical researchers being able to contribute more effectively to clinical decisions and drug testing by combining heterogeneous, collaboratively annotated datasets from patient results (e.g. blood tests, physical examinations, free text journals from patients on their experience from treatment) and different scan modalities (e.g. X-Ray, Static and Dynamic MRI), without having to be anxious about tracking the data and their provenance through the complex decision making process, and the handling of the associated multimedia material.
- Or even, a marketing and consultancy company being able to effortlessly forage the Web (blogs, forums, wikis, etc.) for high-level knowledge, such as public opinions about its products and services; it is thus able to capture tractable, commercially vital information that can be used to quickly monitor public response to a new marketing launch; having the means to meaningfully filter, collate and analyse the associated findings; and use the information to inform new strategy.

The goal of the Dicode project [6] was to turn this vision into reality. The project was funded by the European Commission under the FP7 Work Programme (contract number: FP7-ICT-257184 - <http://dicode-project.eu>). It started on September 1st, 2010 and its duration was 36 months. The partners of the Dicode consortium were: Computer Technology Institute & Press “Diophantus” (project coordinator, Greece), University of Leeds (United Kingdom), Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V. (Germany), Universidad Politécnica de Madrid (Spain), Neofonie GmbH (Germany), Image Analysis Ltd (United Kingdom), Biomedical Research Foundation - Academy of Athens (Greece), and Publicis Frankfurt Zweigniederlassung der PWW GmbH (Germany).

This chapter describes the overall context of the Dicode project (Sect. 2), its scientific and technical objectives (Sect. 3), as well as the exploitation of its results and its potential impact (Sect. 4).¹

2 Overall Project Context

Many collaboration and decision making settings are nowadays associated with huge, ever-increasing amounts of multiple types of data, obtained from diverse sources, which often have a low signal-to-noise ratio for addressing the problem at hand. These data may also vary in terms of subjectivity and importance, ranging from individual opinions and estimations to broadly accepted practices and trustable measurements and scientific results. Their types can be of diverse level as far as human understanding and machine interpretation are concerned. At the same time, the associated data are in most cases interconnected, in a vague or explicit manner.

Additional problems start when we want to consider and exploit accumulated volumes of data, which may have been collected over a few weeks or months, and meaningfully analyze them towards making a decision. Admittedly, when things get complex, we need to identify, understand and exploit data patterns; we need to

¹ A shorter version of this chapter appears in [6].

aggregate appropriate volumes of data from multiple sources, and then mine them 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. The way that data will be structured for query and analysis, as well as the way that tools will be designed to handle them efficiently are of great importance and certainly set a big research challenge.

In the settings under consideration, “big data” analytics technology currently receives much criticism, in that it does not provide proper insight into what the data means. To make sense of big data and come with discoveries that help improve decision making in practical contexts, human intelligence should be also exploited. We need to provide the appropriate ways to nurture and capture this human intelligence in order to extract the necessary insights and improve the way machines deal with complex situations.



Fig. 1. The Dicode services exploit the cloud computing paradigm and build on the synergy of machine and human reasoning.

Taking the above issues into account, the Dicode project aimed at facilitating and augmenting collaboration and decision making in data-intensive and cognitively-complex settings. To do so, whenever appropriate, it built on prominent high-performance computing paradigms and proper data processing technologies to meaningfully search, analyze and aggregate data existing in diverse, extremely large, and rapidly evolving sources. At the same time, particular emphasis was given to the deepening of our insights about the proper exploitation of big data, as well as to collaboration and sense making support issues. Building on current advancements, the solution proposed by the Dicode project brings together the reasoning capabilities of both the machine and the humans (Fig. 1).

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 effective in their work practices. Services that were developed and integrated in the context of the Dicode project are released under an open source license.

The achievements of the Dicode project were validated through three use cases:

- ***Clinico-Genomic Research Assimilator.*** The need to collaboratively explore, evaluate, disseminate and diffuse relative scientific findings and results is more than profound today. Towards this objective, Dicode elaborated an integrated clinico-genomic (tacit) knowledge discovery and decision making use case that targets the identification and validation of predictive clinico-genomic models and biomarkers [7].
- ***Trial of Clinical Treatment Effects.*** The goal of this case (which has been expanded in the second year of the project to cover broader clinical trials, not just for Rheumatoid Arthritis) was to facilitate the process of making clinical decisions in drug trials by combining datasets from patient results (blood tests, physical examinations) and the different scan modalities (X-Ray, Static and Dynamic MRI scan images) to reveal the effectiveness of a drug within a trial.
- ***Opinion Mining from unstructured Web 2.0 data.*** Through this case, we validated the Dicode services for the automatic analyses of the voluminous amount of unstructured information existing on the Web, especially in the highly dynamic social media space. Data for this case were primarily obtained from spidering the most popular social Web sites making use of APIs from various Web 2.0 platforms [8].

3 Scientific and Technical Objectives

The project’s objectives have been fully accomplished through an evolutionary approach characterised by:

- the active engagement of all stakeholders (technical partners and use case representatives) in the specification, design and evaluation of the foreseen technological solutions throughout the project;
- the adoption of an incremental development approach, which ensured that end users can experiment with the Dicode services from the early stages of the project (operational prototype versions of the Dicode services were available at the end of the first year of the project, enhanced versions were delivered in month 24, final versions were ready in month 33);
- the continuous refinement of user requirements through testing (involving users from all three use cases), and
- the early availability of an operational integrated suite of services, which facilitated trials and proof-of-concept purposes, enabled proper exploitation and dissemination activities, and ensured project sustainability.

The association between the project’s objectives and the project’s milestones is illustrated in Fig. 2. As shown, “Laying Foundations”, “Integration, Validation and

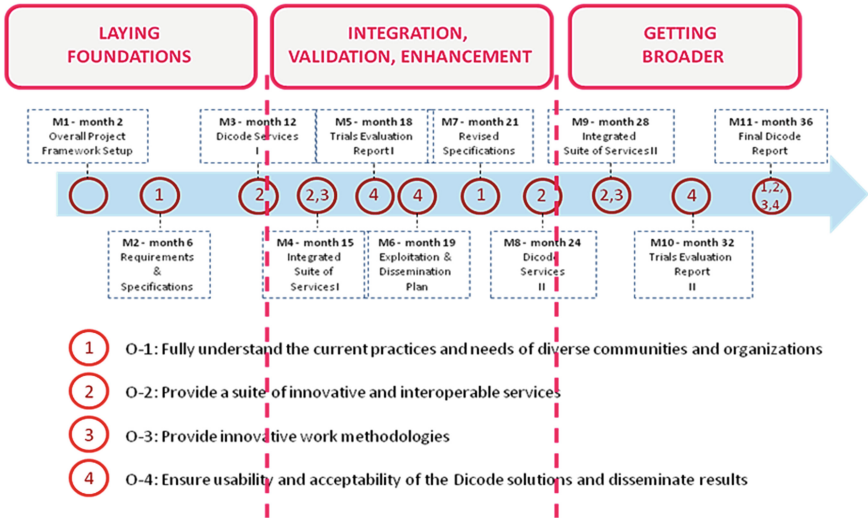


Fig. 2. S&T objectives, project’s milestones and goals set for each year of the Dicode project.

Enhancement” and “Getting Broader” was the overall goal for each year of the project, respectively. As justified in the following, the Dicode project successfully reached these goals.

In particular, the project’s scientific and technical objectives were:

- *O-1: To fully understand the current practices and needs of diverse communities and organizations as far as data-intensive and cognitively-complex collaboration and decision making is concerned.* Three representative use cases were continuously elaborated throughout the project. Related settings were also considered, aiming to reveal practices and needs associated with both large data sets and real-time data [9]. The accomplishment of this objective was critical for the applicability of the Dicode approach in a wide variety of settings.

This objective was of high importance throughout the project. Thoroughly considering the feedback from the two evaluation rounds of Dicode services across the project’s use cases, an analysis of the lessons learned was documented and services’ specifications were revised to inform the final iteration of development. A much deeper understanding of the use cases’ differences and similarities, as well as of their potential to explore the full range of Dicode services, was achieved through close collaboration between technical partners and end users.

- *O-2: To provide a suite of innovative, adaptive and interoperable services (both at a conceptual and a technical level) that satisfies the full range of the associated requirements.* The development of Dicode services facilitated and augmented collaboration, sense-making and decision-making in data-intensive and cognitively-complex settings, while also serving the underlying requirements of capturing, delivering and analyzing pertinent information (Fig. 3). Dicode services are running on the Web. Throughout the project, much attention was given to the adaptability of

Dicode services with respect to changes in user requirements and operating conditions. Moreover, especially during the third year of the project, development efforts paid much attention to usability issues. Particular sub-objectives concern the development and seamless integration of:

- *O-2.1: Data Acquisition Services*, which enable the purposeful capturing of tractable information that exists in diverse data sources and formats. Particular attention was paid to web resources and the integration of social media APIs and high quality third-party feeds.
- *O-2.2: Data Pre-processing Services*, which efficiently manipulate raw data before their storage to the foreseen solution. Transformation of different kinds of documents into a canonical form, structuring of documents from layout information (e.g. detection of navigation, comments, abstracts), data cleansing (e.g. removing noise from web pages, discarding useless database records), as well as language detection and linguistic annotations are some of the functionalities falling in this category of services.
- *O-2.3: Data Mining Services*, which in many cases exploit and are built on top of a cloud infrastructure and other most prominent large data processing technologies to offer functionalities such as high performance full text search, data indexing, classification and clustering, directed data filtering and fusion, and meaningful data aggregation. Advanced text mining techniques, such as named entity recognition, relation extraction and opinion mining, help to extract valuable semantic information from unstructured texts. Intelligent data mining techniques elaborated include local pattern mining and similarity learning.

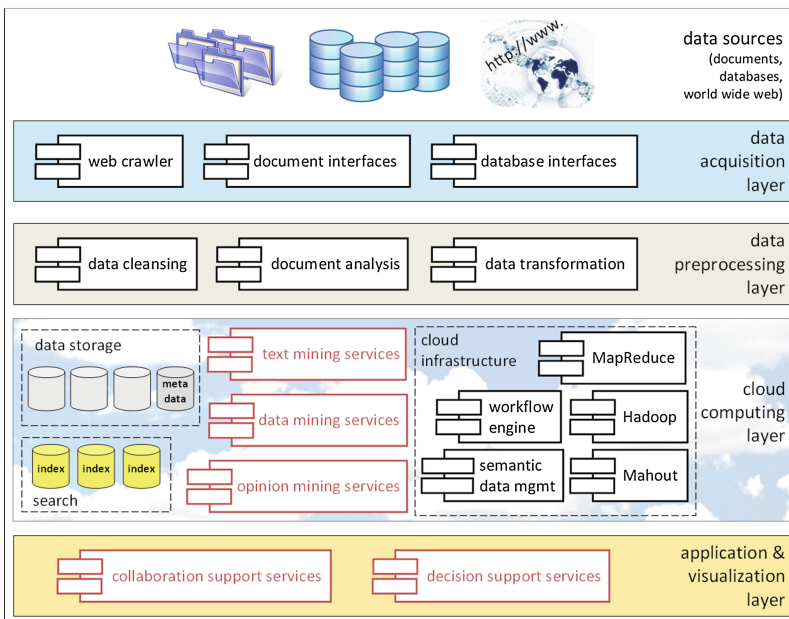


Fig. 3. The Dicode architecture and suite of services.

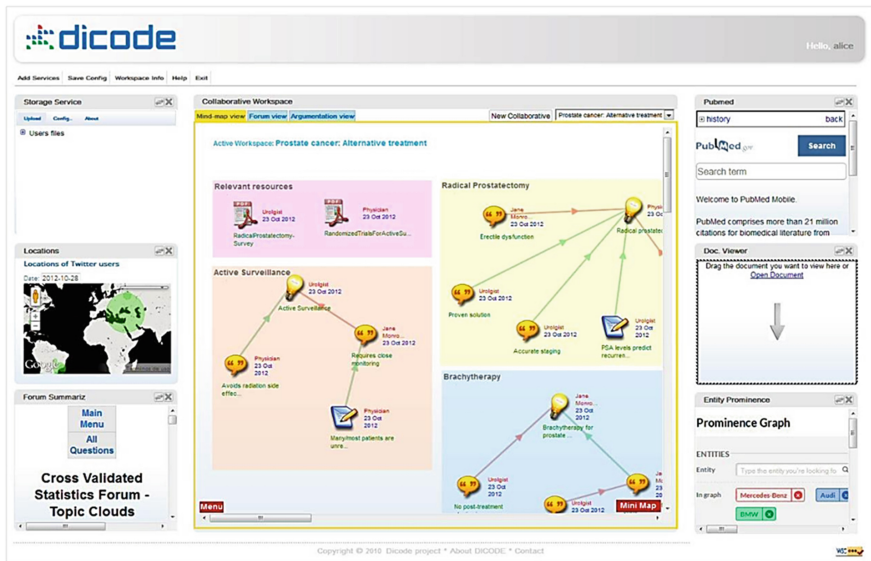


Fig. 4. Integrating Dicode services: An instance of the Dicode Workbench.

- *O-2.4: Collaboration Support Services*, which facilitate the synchronous and asynchronous collaboration of stakeholders through adaptive workspaces, efficiently handle the representation and visualization of the outcomes of the data mining services (through alternative and dedicated data visualization schemas), and accommodate a series of actions for the appropriate handling of data in each use case.
- *O-2.5: Decision Making Support Services*, which augment (both individual and group) sense-making and decision-making by supporting stakeholders in locating, retrieving and arguing about relevant information and knowledge, as well as by providing them with appropriate notifications and recommendations.

This objective was of paramount importance for the success of the project. Taking into account feedback from the two evaluation rounds of the project, as well as recommendations of the Project officer and Project Reviewers, the final operational versions of the Dicode Workbench and integrated Dicode services were developed and tested across the project's use cases (Fig. 4). Much attention was given to the openness of the Dicode solution, in order to augment exploitation purposes. An appropriate infrastructure of in-house computer clusters for running large scale data mining experiments and testing prototype implementations, as well as data collections for benchmarking based on textual and structured data, were set and maintained. Standards and guidelines for the development of Dicode services - aiming at ensuring interoperability between the services to be developed and reusability of them through diverse scenarios of use – were defined and revised upon the evolution of the project. Issues around both the conceptual and technical integration of the full range of Dicode services were thoroughly elaborated to upgrade user experience. According to the workplan, the

final versions of the Dicode Data Mining Services [10], the Dicode Collaboration Support Services [11], the Dicode Decision Making Support Services [11], and the Dicode Workbench [12] were produced. In addition, a set of practical lessons learned while developing the Dicode's services and using them in data-intensive and cognitively-complex settings were reported. These lessons concern experiences, concrete recommendations and best practices from the development of the project's services, and they have been presented in a way that could aid people who engage in various phases of developing similar kind of systems [13, 14].

- *O-3: To provide innovative work methodologies that exploit the abovementioned suite of services and advance the current practices in terms of efficiency, creativity, as well as time and cost effectiveness.* These methodologies take into account the nature and needs of contemporary organisations and communities operating in a knowledge-driven economy.

This objective was highly important throughout the project. The established consensus on the role of the envisioned suite of Dicode services was significantly augmented through the two rounds of validation of the Dicode services, which provided valuable insights for the shaping of novel methodologies to be followed in stakeholders' daily work practices. During these evaluation rounds, a long and diverse set of end users tested the Dicode solution (services and Workbench) and provided valuable feedback by pointing out both strengths and weaknesses. These were considered through various real-world scenarios, which actually constituted the base for the definition of Dicode's innovative work methodologies [7, 8]. The proposed methodologies reflect our experiences gained from the overall validation of the project's results and provide useful suggestions and insights to relevant communities and organizations.

- *O-4: To ensure usability and acceptability of the above services and work methodologies through their validation in real use cases, and disseminate the project's results by dedicated actions.*

This objective was also of paramount importance for the success of the project. Two rounds of evaluation of the Dicode Workbench and integrated services through the project's use cases were performed. Properly formulated metrics and questionnaires were employed to analyse the feedback received. Appropriate video-casts – based on everyday user stories from user communities, developers and early adopters – were prepared for each use case. The parameters assessed for each service concerned their acceptability, ease of use, usability, and overall quality [7, 8].

In addition, a comprehensive exploitation and dissemination plan has been produced, ensuring the impact and sustainability of the Dicode outcomes. Initial dissemination and exploitation activities included the development of a corporate identity of the project, the set-up of a web portal, and initial public relations efforts. A significant number of publications have resulted out of joint work among consortium members. These publications appear in international scientific journals and proceedings of international peer-reviewed scientific conferences and workshops (a detailed list of Dicode's dissemination activities appears at <http://dicode-project.eu/index.php?q=news>). Presentations of project-related work were also given in some of the top technology and marketing conferences. Moreover, Dicode organized four scientific workshops, one in

the context of the world leading conference on collaboration support (CSCW 2012), another in the context of the best European conference on machine learning and knowledge discovery (ECML-PKDD 2012), a third one at the leading international conference on knowledge engineering and knowledge management (EKAW 2012), and a fourth one at world leading conference on hypertext and social media (Hypertext 2013). A series of exploitation activities has been also carried out, especially during the last two years of the project. Each Dicode partner put much effort in developing a concrete and realistic exploitation strategy (see next section). Several success stories concerning exploitation of Dicode results, development of strategic partnerships with industry and co-operation with other EU projects have been already reported.

4 Exploitation of Results and Potential Impact

The combination of academic and industrial partners within the Dicode consortium was perfectly suited for working with existing customers and collaborators in a variety of industry and academic segments to develop the Dicode platform for market use. Suitable targets were defined in the early stages of the project. As each target has specific needs which can be met through the technology developed in the Dicode project, partners in the project consortium were involved in cultivating and extending ties to their existing customer base to keep these key assets informed of project developments. The project partners also organized dedicated demonstrations of running prototypes and scenarios of use for key persons in the target organisations.

Figure 5 gives an overview of Dicode's target groups for the exploitation of the project's foreground. In the public sector, the focus lies on public services, public health and e-Science. In the private sector, advertising and communication, media and medicine are the main target areas. In the IT industry, Dicode caters to service integrators, service developers and consultancy companies.

To ensure the sustainability of the Dicode project, each consortium partner formulated a detailed exploitation plan and carried out a set of associated activities, based on modern marketing and communication best practices. Market entry strategies followed in the context of the Dicode project included:

- The definition of appropriate targets, both public and private entities, and partners in the network of the consortium partners who have an interest in the outcomes of the Dicode project, and are also suitable for obtaining first experiences and willing to be used as success stories;
- Building out additional use-cases to fit the needs of the defined targets;
- Strategic partnerships with established players in the market.

The key success indicators of the Dicode project, together with their high effects and actions taken towards ensuring their accomplishment are summarized in Table 1. The final results of the Dicode project advance the state-of-the-art in approaches on (i) the proper exploitation of big data (dealing with the “big data fallacy” issue) and the integrated consideration of data mining and sense-making issues, (ii) recommender systems, with respect to recommendations in heterogeneous, multi-faceted data and the identification of hidden links in complex data types, (iii) understanding text to

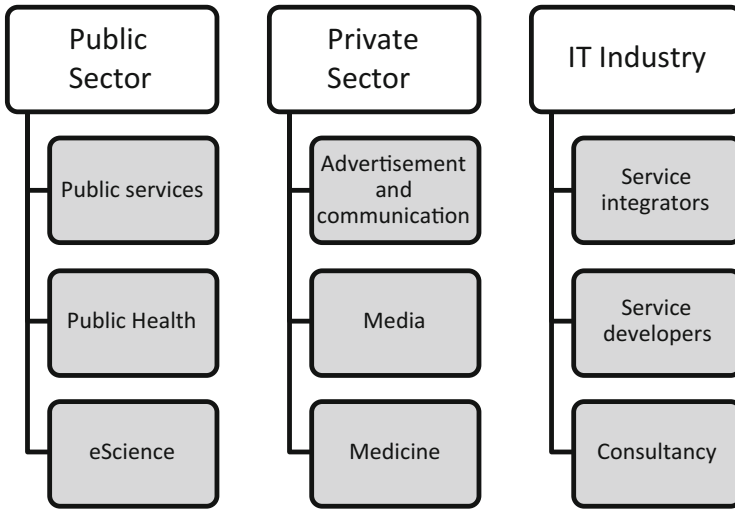


Fig. 5. Dicode targets a wide audience.

drastically reduce the annotation effort for extracting relations, (iv) opinion mining by considering opinion statements as n -ary relations and apply the highly scalable methodology implemented for their recognition, (v) Web 2.0 collaboration support tools in terms of interoperability with third party tools and integration of appropriate reasoning and data mining services, and (vi) decision making support applications, by integrating knowledge management and decision making features as well as by building on the synergy of human and machine argumentation-based reasoning.

Such advancements have shaped innovative work methodologies for dealing with the problems of information overload and cognitive complexity in diverse collaboration and decision making contexts. Adopting the proposed solution, both individual and collaborative sense making are augmented through the meaningful exploitation of prominent data processing and data analysis technologies. The Dicode solution is user-friendly and built on the synergy of human and machine intelligence. Adopting open standards, and in accordance with EU's recent initiatives on Open Systems and Data, the Dicode project has the potential of forming a rich ecology of domain specific and non-specific extensions. The Dicode platform allows for external data service providers to supply information, as well as for external developers to supply additional modules and applications, which are tailored to evolving market conditions. Finally, it enables diverse public and private entities to aggregate, structure, semantically enrich and analyse vast amounts of information. This turns the problem of information overload into a benefit of structured data, which can be used as the basis for decisions of better quality. Simply put, the Dicode solution is able to turn information growth into economic growth.

In particular, the potential impact of the Dicode project (including the socio-economic impact and the wider societal implications so far) concerns:

Table 1. Success indicators and actions to ensure them.

Key success indicators	Actions
Deployment of the Dicode framework is not too costly → high rate of Dicode technology adoption	Adoption of standards, exploitation of existing prototypes and background technology, open-source policy
High level of the Dicode framework's acceptance by users involved in Dicode's use cases → increased users' productivity and creativity	Early and continuous involvement of end users in the development and evaluation of the Dicode platform
Acceptance of the Dicode framework by users outside the Dicode consortium → recognition of the Dicode platform's value from relevant groups and communities	Exploitation of Dicode partners' liaisons with scientific and business stakeholders; Dicode workshops and diverse dissemination activities; relevant market watch
Adaptability and proven portability of the Dicode framework in a wide range of application domains → acceptance of the Dicode framework by industry and academia	Generic and flexible development approach; show cases at relevant scientific and business events and stakeholders; Dicode scientific workshops; scientific dissemination activities; market analysis
Foundations for high-performance scalable data mining in the cloud computing initiative and related open source community → proven added value of the Dicode framework	Adoption of open source principles and concepts; advancement of cloud computing paradigm; dissemination activities related to dedicated workgroups and communities
Hit the optimal market → strengthened EU leadership in the domain of intelligent information management	Development of a detailed and coherent exploitation strategy and thorough consideration of associated perspectives

Better leveraging of human skills, improved quality and quantity of output and reduced time and cost allowing users to concentrate on more creative and innovative activities.

- The Dicode integrated suite of services and corresponding work methodologies facilitate and enhance the integration and aggregation of different stakeholders' perspectives across different collaboration and decision making activities, by explicitly addressing their knowledge and social dynamics. The Dicode platform is able to augment the creativity of stakeholders (stakeholders save time by skipping unnecessary tasks, accomplishing trivial tasks faster, while the platform provides a remedy to the information and cognitive overload). Stakeholders may easily customize the Dicode platform and concentrate on more creative and innovative activities.
- The Dicode platform enables new working practices for stakeholders involved in data-intensive and/or cognitively-complex settings. It has followed a component-based approach, based on open standards. This allows for further development by using and adapting existing modules, or developing new ones to cover the needs of related contexts.

Increased ability to identify and respond appropriately to evolving conditions (e.g. in finance, epidemiology, environmental crises ...) faster and more effectively. Reinforced

ability to collaboratively evolve large-scale, multi-dimensional models from the integration of independently developed datasets.

- In Dicode, machine-tractable knowledge concerning the full lifecycle of collaboration and decision making is accumulated and maintained. Consequently, the Dicode platform augments the productivity of stakeholders, e.g. by enabling them to easily locate and meaningfully reuse existing content. This affects both individuals and the workgroups they belong to.
- The Dicode platform improves the quality and quantity of the collaboration process. Since needs and user types evolve over time, the platform can be easily customized and adapted to address diverse needs and user types.
- The Dicode platform enhances collaboration between individual stakeholders through the meaningful integration and aggregation of independently developed applications (and associated datasets), which allows for a quicker consensus in the decision making process.
- The Dicode platform allows for external data service providers to supply information, as well as for external developers to supply additional modules and applications, which are tailored to evolving market conditions.

Higher levels of information portability and reuse by creating an ecology of systems and services that are dynamic, interoperable, trustworthy and accountable by design.

- The Dicode platform advances the state-of-the-art in information portability and reuse by considering interoperability issues, while also fostering standards-based integration and exploitation of information resources across organisational boundaries.
- The Dicode platform has been developed using existing standards and exploiting existing open source software.
- The Dicode project has developed a large-scale data processing platform. This platform allows for diverse data processing modules to be integrated through appropriate interfaces.
- The Dicode platform exploits, whenever appropriate, a cloud computing environment, which allows for improved information portability and reuse.
- The Dicode platform is web-based. This allows system independence for the end user.
- During the development of the Dicode platform, strong cooperation with committees and organizations which set standards in the fields of cloud computing was held. A series of contributions to free software projects concerning large-scale data processing in a cloud environment have been performed.
- Being designed with “openness” in mind, the Dicode platform is able to create a rich ecology of domain specific and non-specific extensions.

Increased EU competitiveness in the global knowledge economy by fostering standards-based integration and exploitation of information resources and services across domains and organisational boundaries.

- The global knowledge economy demands no barriers to entry. Accordingly, the Dicode platform has been developed by adopting open standards. Additionally, the platform allows for easy sharing of data and information. This enables the creation of marketplaces for information and information suppliers. Data rich applications can be implemented more quickly due to easy access of data through a shared environment.
- The web-based development of the Dicode platform, together with the exploitation of the cloud in some of its modules, allows for global access to innovative data processing services. Moreover, the platform can be easily adapted for international use (i.e. no cultural barriers to entry). The above may reduce fixed costs for companies using the Dicode platform, allowing them to invest more resources and money into their core line of business activities, thus providing them a competitive advantage in the international marketplace (i.e. no financial and technological barriers).
- The Dicode platform allows public and private entities to aggregate, structure and analyse vast amounts of information. This turns the problem of information overload into a benefit of structured data which can be used as the basis for better and quicker decisions. The Dicode platform helps stakeholders enrich current information, and turns the problem of information overload into knowledge discovery.

Strengthened EU leadership at every step of the computer-aided information and knowledge management lifecycle, creating the conditions for the rapid deployment of innovative products and applications based on high quality content.

- The European IT landscape is generally comprised of small and medium sized enterprises (SMEs). For many SMEs, it is difficult to develop new, data rich applications from scratch, basically due to the associated high investment costs. Open Source solutions, such as the Dicode platform, will reduce the barriers for SMEs in the development and hosting of data rich applications.
- In Europe, there are many different languages. For developing an application which can handle and process text sources from different EU countries, it is necessary to use different language dependent modules. Data and text processing standards, as supported by the Dicode platform, allow for the simple replacement of compatible modules which switch from one language to another (plug and play integration).
- Based on the existing Dicode infrastructure and services, new applications can be developed in less time. This yields to quicker “time to market” and faster return-on-investment due to decreased development costs.
- The Dicode platform is able to assist European companies in making better decisions quicker, based on the largest data set possible. As much of the data on the Web is text, Dicode solutions for issues such as sentiment analysis, opinion mining, data mining, trend mining etc. will continue to grow in importance for decision makers.

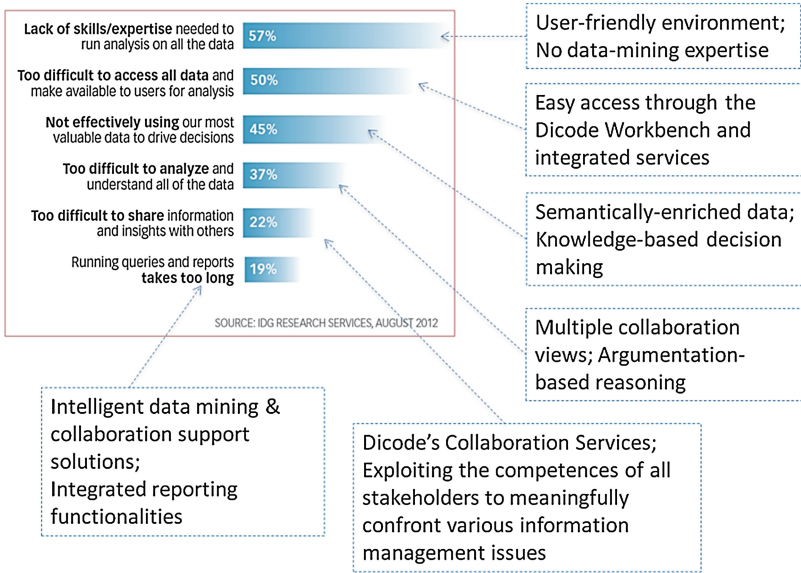


Fig. 6. Big Data challenges and Dicode responses.

5 Discussion and Conclusions

The Dicode platform enables a meaningful aggregation and analysis of Big Data in complex settings. The proposed solution (infrastructure and services) builds on a holistic approach where decision support technologies have been evolved and inter-related in order to efficiently and effectively address the requirements of the knowledge-intensive organization [15]. The Dicode outcomes enable new working practices that turn the problem of information overload and cognitive complexity into the benefit of knowledge discovery. This is achieved through properly structured data that can be used as the basis for more informed decisions. Simply put, the Dicode approach is able to turn information growth into knowledge growth; it improves the quality of collaboration within a Web community, while enabling its users to be more productive and focus on creative activities.

The Dicode approach provides responses to all six Big Data challenges identified in a recent White Paper [16]. Specifically, as shown in Fig. 6, through the user-friendly environment and the integrated approach provided in Dicode, users do not need to possess any particular skills and data mining expertise to run analysis of data; the Dicode Workbench and its integrated services enable easy access to related data, thus making them available for further analysis; the Dicode approach exploits semantically-enriched data and associated knowledge to effectively use the most valuable data and drive decisions; through alternative collaboration views and associated reasoning mechanisms [11], the Dicode Collaboration and Decision Making Support Services facilitate data analysis and understanding; in addition, the abovementioned services exploit the competences of all stakeholders to meaningfully confront diverse

information management issues; finally, a set of intelligent functionalities and solutions offered in Dicode expedite the running of queries and production of reports.

As a last note, we point out that the overall Dicode approach is fully in line with a set of imperatives concerning challenges and opportunities with Big Data, which are reported in another White Paper authored by 21 prominent researchers [17]. Specifically, the Dicode platform enables stakeholders “to run heterogeneous workloads on a single infrastructure that is sufficiently flexible to handle all these workloads”; it is “designed explicitly to have a human in the loop”, thus enabling “humans to easily detect patterns that computers algorithms have a hard time finding”; it provides “supplementary information that explains how each result was derived, and based upon precisely what inputs”; it offers “a rich palette of visualizations”, which are “important in conveying to the users the results of the queries in a way that is best understood in the particular domain”.

References

1. Eppler, M., Mengis, J.: The concept of information overload: a review of literature from organization science, accounting, marketing, MIS, and related disciplines. *Inf. Soc.* **20**(5), 325–344 (2004)
2. IDC. The Diverse and Exploding Digital Universe. White Paper, March 2008. www.idc.com
3. Economist. A Special Report on managing information: Data, data everywhere. Economist (2010)
4. Hara, N., Solomon, P., Kim, S.L., Sonnenwald, D.H.: An emerging view of scientific collaboration: Scientists’ perspectives on collaboration and factors that impact collaboration. *J. Am. Soc. Inform. Sci. Technol.* **54**, 952–965 (2003)
5. Shim, J.P., Warkentin, M., Courtney, J.F., Power, D.J., Sharda, R., Carlsson, C.: Past, present and future of decision support technology. *Decis. Support Syst.* **33**, 111–126 (2002)
6. Karacapilidis, N. (ed.): *Mastering Data-Intensive Collaboration and Decision Making: Cutting-edge research and practical applications in the Dicode project*, Studies in Big Data Series, vol. 5, Springer (2014)
7. Tsiliki, G., Kossida, G.: Clinico-genomic research assimilator: a dicode use case. In: [6], pp. 165–180 (2014)
8. Löffler, R.: Opinion mining from unstructured web 2.0 data: a dicode use case. In: [6], pp. 181–200 (2014)
9. Lau, L., Yang-Turner, F., Karacapilidis, N.: Requirements for big data analytics supporting decision making: a sensemaking perspective. In: [6], pp. 49–70 (2014)
10. Friesen, N., Jakob, M., Kindermann, J., Maassen, D., Poigné, A., Rüping, S., Trabold, D.: The dicode data mining services. In: [6], pp. 89–118 (2014)
11. Tzagarakis, M., Karacapilidis, N., Christodoulou, S., Yang-Turner, F., Lau, L.: The dicode collaboration and decision making support services. In: [6], pp. 119–139 (2014)
12. de la Calle, G., Alonso-Martínez, E., Rojas-Vera, M., García-Remesal, M.: Integrating dicode services: the dicode workbench. In: [6], pp. 141–164 (2014)
13. Friesen, N., Kindermann, J., Maassen, D., Rüping, S.: Data mining in data-intensive and cognitively-complex settings: lessons learned from the dicode project. In: [6], pp. 201–212 (2014)

14. Christodoulou, S., Tzarakis, M., Karacapilidis, N., Yang-Turner, F., Lau, L., Dimitrova, V.: Collaboration and decision making in data-intensive and cognitively-complex settings: lessons learned from the dicode project. In: [6], pp. 213–226 (2014)
15. Karacapilidis, N.: An Overview of Future Challenges of Decision Support Technologies. In: Gupta, J., Forgionne, G., Mora, M. (eds.) *Intelligent Decision-Making Support Systems: Foundations*, pp. 385–399. Applications and Challenges, Springer-Verlag, London, UK (2006)
16. SAS. Data Visualization: Making Big Data Approachable and Valuable. White Paper (2013). http://www.sas.com/content/dam/SAS/en_us/doc/whitepaper2/sas-data-visualization-market-pulse-106176.pdf
17. Computing Community Consortium - Computing Research Association. Challenges and Opportunities with Big Data: A community white paper developed by leading researchers across the United States. White Paper, February 2012. <http://www.cra.org/ccc/files/docs/init/bigdatawhitepaper.pdf>

Knowledge Discovery, Knowledge Engineering and
Knowledge Management

5th International Joint Conference, IC3K 2013,

Vilamoura, Portugal, September 19-22, 2013. Revised
Selected Papers

Fred, A.; Dietz, J.; Liu, K.; Filipe, J. (Eds.)

2015, XVII, 425 p. 150 illus., Softcover

ISBN: 978-3-662-46548-6