



Mastering Data-Intensive Collaboration and Decision Making

FP7 - Information and Communication Technologies

Grant Agreement no: 257184

Collaborative Project

Project start: 1 September 2010, Duration: 36 months

D7.2.1 - Dissemination and Exploitation Activities Report (initial version)

Due date of deliverable: 29 February 2012
Actual submission date: 29 February 2012
Responsible Partner: NEO
Contributing Partners: CTI, UOL, FHG, UPM, IMA, BRF, PUB

Nature: Report Prototype Demonstrator Other

Dissemination Level:

- PU : Public
- PP : Restricted to other programme participants (including the Commission Services)
- RE : Restricted to a group specified by the consortium (including the Commission Services)
- CO : Confidential, only for members of the consortium (including the Commission Services)

Keyword List: Dissemination, Exploitation, Open source, Software license, IPR



The Dicode project (dicode-project.eu) is funded by the European Commission, Information Society and Media Directorate General, under the FP7 Cooperation programme (ICT/SO 4.3: Intelligent Information Management).

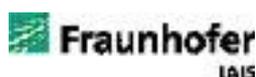
The Dicode Consortium



Computer Technology Institute & Press "Diophantus"
(CTI) (coordinator), Greece



University of Leeds (UOL), UK



Fraunhofer-Gesellschaft zur Foerderung der angewandten
Forschung e.V. (FHG), Germany



Universidad Politécnica de Madrid (UPM), Spain



Neofonie GmbH (NEO), Germany



Image Analysis Limited (IMA), UK



Biomedical Research Foundation, Academy of Athens
(BRF), Greece



Publicis Frankfurt Zweigniederlassung der PWW GmbH
(PUB), Germany

Document history			
Version	Date	Status	Modifications made by
1	27-01-2012	First draft	Doris Maassen, NEO
2	27-01-2012	Second draft	Max Jakob, NEO
3	07-02-2012	Input from UOL, CTI, UPM, BRF incorporated - sent to internal reviewers	Doris Maassen, NEO
4	13-02-2012	Reviewers' comments incorporated (UOL)	Lydia Lau, UOL Doris Maassen, NEO
5	22-02-2012	Reviewers' comments incorporated (CTI)	Nikos Karacapilidis, CTI Doris Maassen, NEO
6	23-02-2012	Sent to Steering Committee	Doris Maassen, NEO
7	29-02-2012	Final version (approved by SC, sent to the Project Officer)	Doris Maassen, NEO

Deliverable manager

- Doris Maassen, NEO

List of Contributors

- Doris Maassen, NEO
- Max Jakob, NEO
- Nena Karagianni, CTI
- Nikos Karacapilidis, CTI
- Guillermo de la Calle Velasco, UPM
- Georgia Tsiliki, BRF
- Lydia Lau, UOL
- Mark Hinton, IMA
- Stefan Rüping, FHG
- Ralf Löffler, PUB

List of Evaluators

- Lydia Lau, UOL
- Nikos Karacapilidis, CTI

Summary

This document is the first version of Dicode's Dissemination and Exploitation Activities Report. Initially, an overview of the dissemination strategy is given. Dicode targets at scientific communities, industry, public sector and open source communities. Dissemination tools and channels ranging from the project web site to social media tools are described in the second section. Subsequently, Dicode's dissemination results of the first half of the project are presented. A list of publications, talks and presentations, prototypes and source code is provided. Selected events and competitions are also listed, followed by details about the Dicode workshops and the scientific workshops organized by the project consortium. The document concludes with Dicode's intellectual property rights strategy and an outlook to future work.

Table of Contents

1	Introduction.....	5
2	Dissemination Strategy.....	6
2.1	Overview.....	7
2.1.1	Typical dissemination and exploitation activities.....	9
2.1.2	Additional dissemination and exploitation activities.....	9
2.1.3	The Dicode workshops.....	9
2.2	Dissemination target groups.....	9
2.2.1	Scientific communities.....	10
2.2.2	Industry and public sector.....	11
2.2.3	Open Source communities.....	13
2.2.4	Dissemination activities matrix.....	13
3	Dissemination Tools and Channels.....	14
3.1	Project web page.....	14
3.2	Leaflets.....	16
3.3	Social Media.....	17
3.3.1	Twitter.....	17
3.3.2	Facebook.....	18
3.3.3	Other tools and platforms.....	19
4	Dissemination activities report.....	20
4.1	Publications Strategy.....	20
4.2	List of publications.....	21
4.3	Talks and presentations.....	23
4.4	Prototypes and Source Code.....	24
4.4.1	Repository at Google Code.....	24
4.4.2	Social Coding: GitHub.....	24
4.4.3	Ontology resources.....	26
4.4.4	Contributions to free software projects.....	26
4.5	Selected Events and Competitions.....	26
4.5.1	Semantic/NLP Hackathon.....	26
4.5.2	Hadoop Get Together.....	26
4.5.3	Discovery Challenge.....	27
4.6	The Dicode workshops.....	27
4.7	Collaboration with relevant EU projects and initiatives.....	29
5	Intellectual property rights strategy.....	31
5.1	General objectives and principles.....	31
5.2	Licensing strategy.....	31
5.2.1	Open Source Licensing for base components.....	31
5.2.2	Proprietary licenses for additional modules.....	32
6	Outlook.....	34

1 Introduction

The main objective of all dissemination and exploitation activities in Dicode is to maximize the visibility and sustainability of the project outcomes. This is achieved by a coordinated approach in diverse dissemination and exploitation activities.

This deliverable is the first version of Dicode's "Dissemination and Exploitation Activities Report". It contains both Dicode's dissemination and exploitation plan and a report on Dicode's activities in this field so far.

Activities performed include the development of a corporate identity of the project, the set-up of a web portal, and initial public relations efforts. Moreover, a number of publications appeared in international scientific journals and proceedings of international peer-reviewed scientific conferences and workshops. Dicode members also organized the 1st Dicode workshop in the context of the world leading conference on computer-supported collaborative work (CSCW 2012) and two other international project-related events. Finally, various activities with respect to the availability of software prototypes for public use, the continuous support to active and interested workgroups that use the Dicode services, and the maintenance of close contacts to the industry took place.

2 Dissemination Strategy

Dicode's project proposal contained a very detailed description of strategies concerning dissemination and exploitation. Therefore, Dicode refrained from producing a "Dissemination and Exploitation plan" as a separate Deliverable in the beginning of the project. Dicode instead focussed on living documents as the Dicode web portal (<http://dicode-project.eu>) which contains both a blog section with news about the project and an up-to-date list of all dissemination activities ranging from publications to contributions to Free Software Projects. Details about the web portal can be found in Deliverable 7.1 <http://dicode-project.eu/sites/default/files/D7.1-WebPortal.pdf> which is publicly available at the Dicode web portal.

Exploitation and dissemination activities will follow the accomplishment of the results expected by the Dicode project. These are described in Table 3.3 of Dicode's Description of Work (DoW)¹, which also refers to their maturity level at the beginning and end of the project.

Exploitation and dissemination activities are among the basic actions to ensure the success of the project. Such actions have been also reported in DoW (Table 3.1) and include:

- Early and continuous involvement of end users in the development and evaluation of the Dicode platform towards achieving a high level of the Dicode framework's acceptance by users involved in Dicode's use cases';
- Exploitation of Dicode partners' liaisons with scientific and business stakeholders, organization of Dicode workshops and diverse dissemination activities, as well as relevant market watch towards achieving acceptance of the Dicode framework by users outside the Dicode consortium and, consequently, recognition of the Dicode platform's value from relevant groups and communities;
- Generic and flexible development approach, show cases at relevant scientific and business events and stakeholders, organization of Dicode workshops and other scientific dissemination activities, as well as market analysis, towards achieving adaptability and proven portability of the Dicode framework in a wide range of application domains and, consequently, acceptance of the Dicode framework by industry and academia
- Dissemination activities related to dedicated workgroups and communities towards proving the added value of the Dicode platform;
- Development of a detailed and coherent overall marketing strategy and perspectives (see DoW, Section 3.2.1) to hit the optimal or create a new market.

In this section the dissemination strategy is discussed in detail. Dicode's dissemination strategy is strongly related to the agile approach of the project which is extensively discussed in Deliverable 2.2: The Dicode approach (http://dicode-project.eu/sites/default/files/D2.2-TheDicodeApproach-v6-EC_0.pdf).

After giving some details about the basic concept, this section presents detailed information about Dicode's dissemination and exploitation strategy. Which are the target organizations, communities and markets Dicode is about to tackle? Later in this report the achievements in all areas are discussed.

¹ See Dicode Grant Agreement, Annex 1 Description of Work, Part B, page 93.

2.1 Overview

Core concepts of Dicode include the use of an agile methodology both in requirements engineering and in software development. Crucial rules are the following:

- Release early, release fast.
- Get feedback from the software and the user.
- Refine requirements and overall approach.

An agile approach requires both early dissemination and early exploitation. The dissemination strategy strongly depends on the type of users. We have identified three main groups of users: The end user of platforms and services “powered by Dicode”, software developers integrating publicly available Dicode services and software developers contributing new services to Dicode.

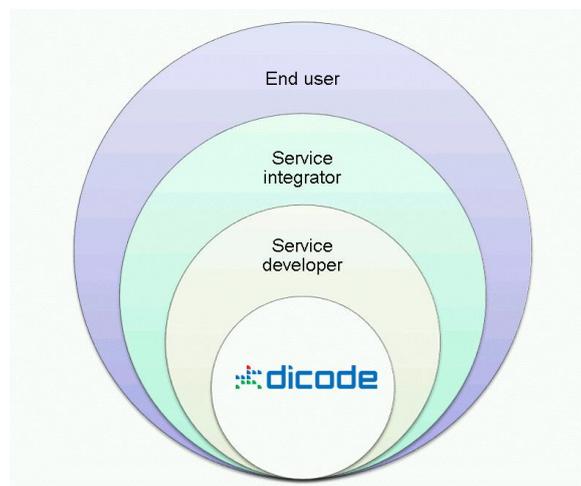


Figure 1: Dicode's users

All three roles shown in the diagram can be undertaken both by members of the Dicode consortium and by external users. At the beginning, only Dicode members act as service developers and service integrators. Later in the project, external developers and IT companies can take over these roles. Basic services developed by Dicode have to be published as open source.² As a consequence, Dicode's dissemination strategy emphasizes the dissemination to open source communities.

Dissemination and exploitation activities take place in Dicode's work package 7 (WP7) which is a joint effort of all partners in Dicode as shown in Figure 2.

² Details on Dicode's Intellectual Property Rights Strategy can be found in Section 5.

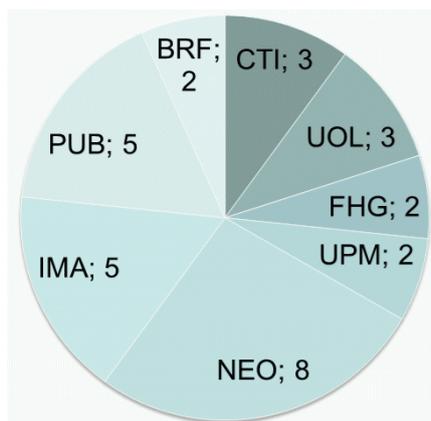


Figure 2: Distribution of person month in WP7

NEO as WP lead has a significantly bigger budget than all other partners. IMA and PUB also have a bigger share of the budget, because they play an important role in helping to build up strong ties with the industry and the end users, right from the beginning of the project.

WP7 consists of three tasks and the corresponding deliverables which are briefly described in the following sections.

Task	Description
T7.1	Typical dissemination and exploitation activities Website, Publications, Press releases
T7.2	Additional dissemination and exploitation activities Prototypes, developer/user community, industry contacts
T7.3	Dicode workshops

Table 1: Tasks in WP7

Deliverable	Description
D7.1	Web portal
D7.2.1	Dissemination and Exploitation Activities Report (initial version)
D7.2.2	Dissemination and Exploitation Activities Report (enhanced version)
D7.3.1	First Dicode Workshop Proceedings
D7.3.2	Second Dicode Workshop Proceedings

Table 2: Deliverables in WP7

2.1.1 Typical dissemination and exploitation activities

T7.1 lays the ground for all dissemination and exploitation activities. This included the development of a corporate identity of the project, the setup of a web portal and initial public relations efforts:

- Production of deliverable *D7.1: Web Portal*.
- Creation of logos to be used in all Dicode communication.
- Setup and maintenance of the project's web portal: work on both content and structure.
- Setup of Dicode pages (profiles) at popular social networking sites including Facebook, LinkedIn and Twitter.
- Production of press release on the launch of the Dicode project (preparation of English, German, Spanish and Greek versions) and dissemination of it through various channels (e-mailing lists, press, etc.).
- Creation and dissemination of Dicode brochure (project flyer).

The second focus point of this task was the dissemination of research results through scientific publications and presentations on conferences concerning both the academic field and the open source communities. A significant number of publications resulted out of joint work among consortium members. These publications appeared in international scientific journals and proceedings of international scientific conferences and workshops. For details on publications and events see Section 4.

2.1.2 Additional dissemination and exploitation activities

This task focuses on the availability of software prototypes for public use, the continuous support to active and interested workgroups that use the Dicode services and the maintenance of close contacts with the industry. There were numerous activities presenting the idea and first prototypes to an interested audience both in developer communities and the industry. Details on the achievements can be found in Section 4.

2.1.3 The Dicode workshops

Dicode organizes two workshops – one in the second and one in the third year. The objective is to set up and conduct high-quality scientific workshops that will bring together industry and academia. The outcome of the workshops will be published in the Dicode workshop proceedings (Deliverables D7.3.1 and D7.3.2). In February 2012, i.e. during the preparation of this deliverable, the first Dicode workshop took place in Seattle. Details about the concept, participant and outcome can be found in Section 4.6.

During the reporting period (first 18 months of the project), Dicode members were also involved in the organization of the *International Workshop on Adaptive Support for Team Collaboration* (<http://astc2011.ascolla.org/home/>), held in conjunction with UMAP 2011 (Girona, Spain, July 15, 2011).

2.2 Dissemination target groups

Dicode focuses on advances in different areas, ranging from high-performance and scalable data mining to collaboration and decision making support (Figure 3). In Dicode's dissemination strategy, scientific communities and the industry are equally important. A key aspect is the dissemination to open source communities.

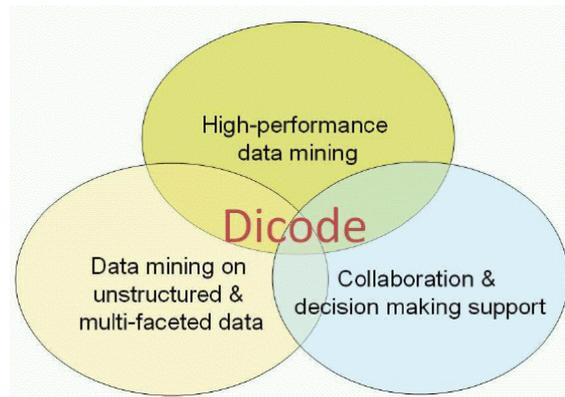


Figure 3: Dicode's main research areas.

Dicode's dissemination strategy is described below. The section is divided into three parts concerning dissemination to scientific communities, to the industry and public sector, and - last but not least - to open source communities. The division is somehow artificial: Dicode strives for applied research based on open source development - an area which links all three domains.

2.2.1 Scientific communities

The strategy concerning the dissemination of scientific results has been clearly laid out in the DoW.³ According to the strategy, Dicode's scientific results are published in conferences and journals dedicated to fields such as Computer-Supported Collaborative Work, Data Mining, Machine Learning and Knowledge Discovery, Decision Making and Decision Making Support Systems, Knowledge Management, and Future Generation Computing Systems, as well as to interdisciplinary and applied research related to both ICT and Social Sciences.

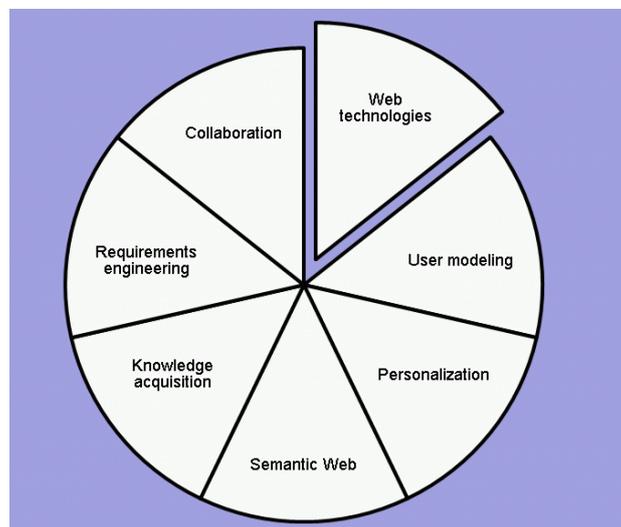


Figure 4: Fields of research in Dicode

The extensive publication list (Section 4.1) shows that many of the fields have already been tackled in the first half of the project. The main dissemination activities were via international workshops and conferences in relevant research communities to report our early efforts. Those events are usually in conjunction with a main conference with a total

³ For a list of target journals and conferences see DoW, p. 104.

attendance of 200-300 people and truly international. In 2012 and 2013, Dicode plans to disseminate the scientific outcome of the project to targets such as:

- WWW - World Wide Web Conference,
- UMAP - International Conference on User Modeling, Adaptation, and Personalization ,
- Web Intelligence (WI),
- International or European Semantic Web Conference (ESWC/ISWC),
- International Conference of Intellifent User Interfaces (IUI),
- K-CAP/ EKAW - European Knowledge acquisition workshop),
- CSCW/ECSCW - Computer Supported Cooperative Work (International or European),
- International Requirements Engineering Conference,
- LREC - International conference on Language Resources and Evaluation,
- SIGCHI Conference on Human Factors in Computing Systems,
- Journal of UMUI (User modelling and user adaptive interaction),
- Journal of Web Semantics,
- EURO Journal on Decision Processes,
- Journal of Data and Knowledge Engineering.

Products and proof of concept prototypes play a central role in Dicode's dissemination activities in years 2 and 3. One of the most relevant and visible products is the Dicode Workbench. The objective of the Dicode Workbench is to make the daily work of researchers easier by providing uniform and integrated access to different resources in a common web application. Researchers coming from the three Dicode use cases will use this environment although they have completely different interests and expertise. The potential target audience for the Dicode Workbench will be very heterogeneous, ranging from biologists or physicians to market analysts or publicists. We expect to publish the scientific results of the project in different top-ranking scientific journals and conferences in the biomedical and informatics fields.

A detailed plan for the use and dissemination of foreground per partner can be found in Section B3.2 of the DoW.

2.2.2 Industry and public sector

The Dicode consortium has strong ties to large media corporations and leading business enterprises and to the public sector. While partners like FHG (Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V.) and NEO (Neofonie GmbH) aim at offering the Dicode technologies to their customers, IMA (Image Analyses Ltd) aims at a direct exploitation in their MRI analysis software "Dynamika".

The following diagram gives an overview over Dicode's dissemination target groups. In the public sector, the focus lies on 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.

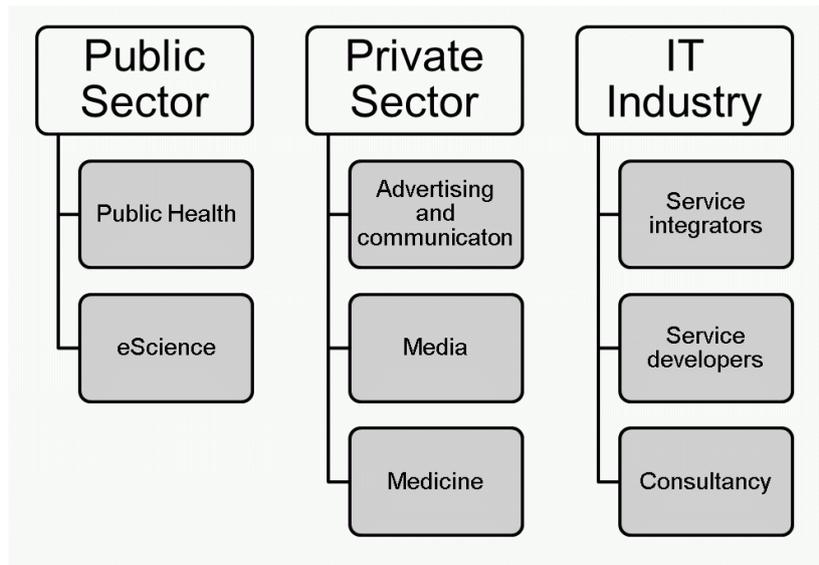


Figure 5: Dicode targets a wide audience

During the first year of the project, the dissemination strategy outlined in the DoW was concretized. In the industry, Dicode's dissemination strategy mainly focussed on the advertising and communication sector, media and the medical field.

Advertising and Communication

Dicode's strategy is two-folded: on the one hand, we focus on marketing and market research conferences, on the other hand on Social Media events. Especially PUB will increase the activities in this area in the second half of the project. The main goal of all activities is to gain a broader understanding of the consumers and of their active engagement in social media.

Dicode's mission is to enhance the relationship between companies and consumers. Social media allows for a new quality of this relationship. To make use of the big potential that social media holds in this respect, there is a fundamental need for monitoring data. We created some awareness for the Dicode approach in the marketing community. Dicode developed a couple of data visualization widgets which help to present early results of social media analysis to marketing analysts and to customers.⁴

Media

Most publishers added Web 2.0 features such as blogs, forums and twitter to their platforms recently, which allows for direct feedback by the readers. The Dicode platform gives publishing companies the basis technology to easily measure and analyze feedback from their customers on the basis of blog comments, forum entries and tweets. Additionally Dicode offers base technologies for content enrichment (e.g. automatic metadata generation) which can be used for the development of innovative search and navigation technologies on media platforms. NEO aims at integrating the Dicode services into prototypes for media customers. NEO's new news platform <https://beta.nachrichten.de/> can serve as a showcase for Dicode's text mining services and semantic technologies.

⁴ See Section 4.3 for a list of events.

Medical industry

Direct exploitation of Dicode results is at heart of IMA's roadmap for the enterprise version of its software package Dynamika. Currently Dynamika is a single user, PC based solution. It is apparent that the future of clinical trials and research in MRI (Magnetic Resonance Imaging) techniques to be used for all diseases requires the use of data sharing, collaboration and the ability to mine and make sense of large numbers of unstructured patient records, image sets, clinician notes and calculated parameters. During the first year of Dicode, a roadmap for an enterprise version of Dynamika was developed that will bring these capabilities to customers. This roadmap is designed to enable the ready use of Dicode services. IMA has been awarded a grant from the Technology Strategy Board in the UK to help develop this roadmap. It is aligned with, and has been influenced by, the Dicode vision.

2.2.3 Open Source communities

As special attention has been given to the use of open source software and standards, partial success of the project's dissemination and exploitation activities will depend on the acceptance of the project within the open source community. To improve the chances of acceptance by this community, the project was positioned from the beginning at various conferences held by the open source community (e.g. ApacheCon, Berlin Buzzwords, FOSDEM, etc.) and/or such conferences where the primary and secondary target audience is members from the open source community.

2.2.4 Dissemination activities matrix

Table 3 provides an overview of the dissemination activities for each target group.

Target	Activities / Medium	Objectives
Scientific communities (Professors, Post-doc students, Research Assistants, PhD students, other related research projects)	Refereed Journal Publications; Oral and Poster Presentations at Conferences; Web site; Flyer; Dicode workshops.	Conduct specific research discussions; raise awareness on the project advances; organization of clusters with interested research teams.
Industry and Public Sector (Directors, Managers, Team Leaders, Public Sector Bodies, Policy Makers)	Dedicated meetings and presentations; Press Releases; partners' networks; dedicated workshops.	Raise awareness about specific Dicode advances that are of interest to their business; broaden the list of Dicode use cases; investigate future exploitation opportunities.
Open Source Communities	Dedicated meetings and presentations at big open source community events; organization of dedicated workshops.	Exploitation of open source software and standards; acceptance of the Dicode project by the open source community.
General Public	Web site, Press Releases, Web 2.0 applications	Raise awareness on latest science and technology achievements.

Table 3: Dissemination activities matrix

3 Dissemination Tools and Channels

Dicode has produced a variety of communication and marketing material to promote the project. This section gives a brief description of all dissemination tools and channels used within Dicode.

3.1 Project web page

The Dicode web page serves as a central hub which connects all dissemination tools and channels used by the project. Dicode constantly updates the website with news about the project.

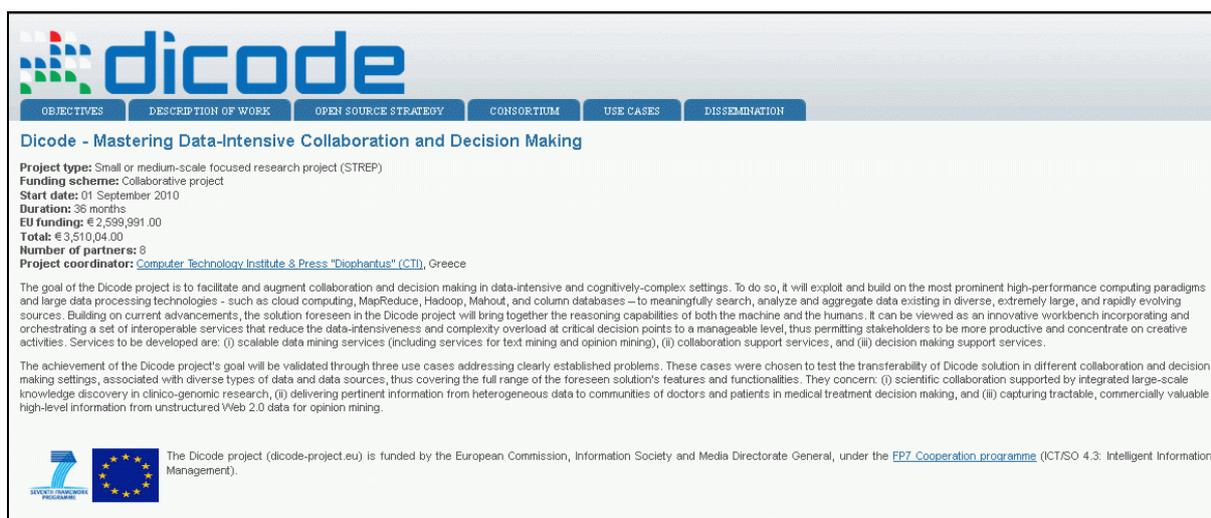


Figure 6: The Dicode web page

On the right side of the page, the user finds links to recent blog posts and to Social Media platforms like Twitter, Facebook, CiteULike, and LinkedIn. Additionally, links to Dicode's open source repositories are supplied.



Figure 7: Connect with Dicode

News about the project can be found in Dicode's blog. The top entries also appear on the initial page. A click on the author of a blog entry leads to the author's blog. The following screenshot shows the top blog posts as of January 2012.

LATEST BLOG POSTS

Dicode at Apache Hadoop Get Together Berlin
 Tuesday, December 13, 2011 - 16:46 by doris
 Max Jakob from Neofonie's Dicode team gave a hands on presentation on Apache Pig for Natural Language Processing at the Hadoop Get Together in Berlin. <http://blog.isabel-drost.de/index.php/archives/343/apache-hadoop-get-together>

 [doris's blog](#) | [Login to post comments](#)

dicoSyn 2012 - A Dicode workshop at CSCW 2012 / submissions due: December 2, 2011
 Monday, November 21, 2011 - 17:47 by nikoskaracapilidis
 Contemporary collaboration settings are often associated with huge, ever-increasing amounts of multiple types of data, which may vary in terms of relevance, subjectivity and importance, ranging from individual opinions to broadly accepted practices. In such settings, collective sense making is crucial for well-informed decision making. This sense making process may both utilize and provide input to intelligent information analysis tools. Through position papers and interactive discussions, this workshop aims to bring together researchers and practitioners from different scientific fields and research communities to further explore (i) the synergy between human and machine intelligence, and (ii) larger issues surrounding analytical practices and data sharing practices in the above settings.

Submissions due: December 2, 2011
 For details, see: <http://dicode.cti.gr/dicosyn12/>

 [nikoskaracapilidis's blog](#) | [Login to post comments](#)

Dicode @ the 10th IEEE International Workshop on Biomedical Engineering
 Friday, October 14, 2011 - 16:14 by nikoskaracapilidis
 A paper entitled "Facilitating Scientific Collaboration in Data-Intensive Biomedical Settings", authored by Nikos Karacapilidis, Manolis Tzagarakis, Spyros Christodoulou and Georgia Tsilki, was presented at the 10th IEEE International Workshop on Biomedical Engineering (10BioEng), which took place in Kos Island, Greece (5-7 October 2011). Through a particular collaboration scenario, this paper explores various possibilities and challenges of managing biomedical collaboration with the use of the foreseen Dicode services.

 [nikoskaracapilidis's blog](#) | [Login to post comments](#)

Best Research Paper Award!
 Monday, October 10, 2011 - 15:52 by nenakaragianni
 The paper entitled "Building on the synergy of machine and human reasoning to tackle data-intensive collaboration and decision making" (authored by Nikos Karacapilidis, Stefan Rüping, Manolis Tzagarakis, Axel Poigné and Spyros Christodoulou), which reports on the Dicode approach, was judged worthy of the "Best Research Paper Award" at the KES 3rd International Conference on Intelligent Decision Technologies (July 2011, Piraeus, Greece).

 [nenakaragianni's blog](#) | [Login to post comments](#)

Figure 8: Blog posts

The most important section of the web site is the dissemination page which lists all publications and presentations of Dicode. All press releases, brochures and public deliverables can be found there as well. Additionally, links to prototypes and code repositories and a list of contributions to free software projects are given.

Publications

- N. Karacapilidis, S. Rüping and I. Drost: [Mastering Data-Intensive Collaboration and Decision Making through a Cloud Infrastructure](#). ERCIM News, No 83, October 2010, pp. 49-50.
- M. Garcia-Renessal, A. Cuevas, D. Pérez-Rey, L. Martín, A. Anguita, D. de la Iglesia, G. de la Calle, J. Crespo and V. Majo: [PubDNA Finder: a web database linking full-text articles to sequences of nucleic acids](#). Bioinformatics. November 2010, 26(21):2801-2802. Available from: <http://dx.doi.org/10.1093/bioinformatics/btq520>
- V. Majo, J. Crespo, M. Garcia-Renessal, D. de la Iglesia, D. Perez-Rey and C. Kulkowski: [Biomedical Ontologies: Toward Scientific Debate](#). Methods of information in medicine. May 2011, 50(3):203-216. Available from: <http://dx.doi.org/10.3414/ME10-05-0004>
- A. Paramythi, L. Lau, S. Demetriadis, M. Tzagarakis and S. Kleanthous (eds.): [Proceedings of the International Workshop on Adaptive Support for Team Collaboration 2011 \(ASTC 2011\)](#) - held in conjunction with the International Conference on User Modeling, Adaptation, and Personalization (UMAP 2011), Girona, Spain, July 15, 2011, CEUR Workshop Proceedings, Vol. 743.
- A. Amnari, V. Dimitrova, L. Lau, M. Tzagarakis, and N. Karacapilidis: [Augmented Collaborative Spaces for Collective Sense Making: The Dicode Approach](#). In: A. Paramythi, L. Lau, S. Demetriadis, M. Tzagarakis, & S. Kleanthous (Eds.), Proceedings of the International Workshop on Adaptive Support for Team Collaboration 2011 (ASTC 2011), Girona, Spain, July 15, 2011, CEUR Workshop Proceedings, Vol. 743, pp. 3-13.
- N. Karacapilidis: [Mastering Data-Intensive Collaboration and Decision Making through a Cloud Infrastructure: The Dicode EU project](#). EMBnet Journal, Vol. 17, No 1, July 2011, p. 3
- N. Karacapilidis, S. Rüping, M. Tzagarakis, A. Poigné and S. Christodoulou: [Building on the synergy of machine and human reasoning to tackle data-intensive collaboration and decision making](#). In: J. Watada, G. Philips-Wren, L.C. Jain and R.J. Howlett (eds.), Intelligent Decision Technologies - Proceedings of the 3rd International Conference on Intelligent Decision Technologies (IDT 2011), Piraeus, Greece, July 20-22, 2011, Springer-Verlag, Berlin, Smart Innovation, Systems and Technologies, Vol. 10, pp. 113-122. **Best Research Paper Award**
- F. Yang-Turner and L. Lau: [A Pragmatic Strategy for Creative Requirements Elicitation](#). In Proceedings of the 1st International Workshop on Requirements Engineering for Systems and Systems-of-Systems. Held in conjunction with the 19th IEEE International Requirements Engineering Conference, 29 Aug - 2 Sept, 2011, Trento, Italy.
- F. Yang-Turner and L. Lau: [Extending Use Case Diagrams to Support Requirements Discovery](#). In Proceedings of the 1st International Workshop on Requirements Engineering for Systems and Systems-of-Systems. Held in conjunction with the 19th IEEE International Requirements Engineering Conference, 29 Aug - 2 Sept, 2011, Trento, Italy.
- Dhavalakumar Thakker, Vania Dimitrova, Lydia Lau, Ronald Densau, Stan Karanasiou and Fan Yang-Turner: [A Priori Ontology Modularisation in Ill-defined Domains](#). In Proceedings of the 7th International Conference on Semantic Systems (I-Semantics 2011), 7-9 September 2011, Graz, Austria.
- N. Karacapilidis, M. Tzagarakis, S. Christodoulou and G. Tsilki: [Facilitating Scientific Collaboration in Data-Intensive Biomedical Settings](#). In Proceedings of the 10th IEEE International Workshop on Biomedical Engineering, Kos Island, Greece, October 5-7, 2011, CD-ROM version (paper #369).
- D. Thakker, F. Yang-Turner, L. Lau and V. Dimitrova: [Socio-technical Ontology Development for Modelling Sensemaking in Heterogeneous Domains](#). In Proceedings of the Workshop on Ontologies Come of Age in the Semantic Web (OCAS2011) at the 10th International Semantic Web Conference (ISWC), Bonn, Germany, October 24, 2011.
- G. Tsilki and S. Kossida: [Fusion methodologies for biomedical data](#). Journal of Proteomics (in print).
- N. Karacapilidis, M. Tzagarakis, N. Karacousos and S. Christodoulou: [Mitigating the cognitive overload of contemporary argumentation-based collaboration settings](#). In Proceedings of the 7th International Conference on Collaborative Computing (CollaborateCom 2011), Orlando, Florida, USA, October 15-18, 2011.
- N. Karacapilidis, M. Tzagarakis, S. Christodoulou and G. Tsilki: [Facilitating and Augmenting Collaboration in the Biomedical Domain](#). International Journal of Systems Biology and Biomedical Technologies (in print).

Talks and Presentations

- [SemanticALP Hackathon at Berlin Buzzwords Conference, June 2011. Report by Olivier Grisel](#)
- [Scaling Data Analysis with Apache Mahout](#) - presentation by Isabel Drost at O'Reilly Strata, Santa Clara, February 2011.
- [Intelligent data analysis - Apache Mahout](#) - presentation by Isabel Drost at Devovx, Antwerp, November 2010. [\[slides\]](#)[\[video\]](#)
- [Scaling Intelligent Data Analysis with Apache Mahout](#) - presentation by Isabel Drost at codebits, Lisbon, November 2010 (travel and hotel kindly sponsored by [Sapo.pt](#)). [\[slides\]](#)[\[video\]](#)
- [Apache Mahout - making data analysis easy](#) - presentation by Isabel Drost at Apache Con US, Atlanta, November 2010. [\[slides\]](#)
- Vania Dimitrova, Keynote speech at the Third International Conference on Software, Services & Semantic Technologies, September 1-3, 2011 Bourgas, Bulgaria; presenting the work on ROO, DON, and Augmentor

Prototypes and Source Code

- Dicode source code repository (WP3NEO) <http://code.google.com/p/dicode/>
- DON (Dicode Ontology): <https://sites.google.com/site/ontomatic/don>
- Dicode Community Modelling Resource: <https://www.dicode-project.eu/index.php?node#8>

Contributions to Free Software Projects

- [MAHOUT-S60](#) improved extensibility of Apache Mahout with respect to text processing for co-location analysis and LDA.
- [MAHOUT-S95](#) improved deployment of Mahout artifacts (now also include reference to source and JavaDoc).
- [MAHOUT-S90](#) adds TVS (Tab Separate Value) input file support to SequenceFilesFromDirectory.

Figure 9: List of publications, presentations and software contributions

3.2 Leaflets

In the beginning of the project, a brochure was published. A second version of the brochure is currently being prepared.

GOALS

The goal of the Dicode project is to facilitate and augment collaboration and decision making in data-intensive and cognitively-complex settings.

To do so, it will exploit and build on the most prominent high-performance computing paradigms and large data processing technologies to meaningfully search, analyze and aggregate data existing in diverse, extremely large, and rapidly evolving sources.

More specifically, the project's S&T objectives are to:

- fully understand the practices and needs of diverse communities and organizations as far as data-intensive and cognitively-complex collaboration and decision making is concerned.
- provide a suite of innovative, adaptive and interoperable services that satisfies the full range of the associated requirements
- provide innovative work methodologies that exploit the abovementioned suite of services and advance the current practices in terms of efficacy, creativity, as well as time and cost effectiveness.
- ensure usability and acceptability of the above services and work methodologies through their validation in real use cases.

CONTACT

Prof. Nikos Karacapiiidis
Research Academic Computer Technology Institute
26504 Rio Patras, Greece

Tel: +30 2610 960305
E-mail: info@dicode-project.eu
<http://dicode-project.eu>

USE CASES

CLINICO-GENOMIC RESEARCH ASSIMILATOR

This use case focuses on scientific collaboration in the clinico-genomic research community managing translational research processes so that relevant findings and results are timely delivered. It helps to explore, evaluate, disseminate and diffuse relative scientific results. The Clinico-Genomic Research Assimilator links and mines disparate clinical and post-genomic data sources.

TRIAL OF RHEUMATOID ARTHRITIS TREATMENT

Focusing on medical decision making, this use case helps doctors and patients in the domain of Rheumatoid Arthritis treatment. Dicode services will enable more effective collaborative decision making to speed up the introduction of life saving treatments.

OPINION MINING FROM UNSTRUCTURED WEB 2.0 DATA

This use case concerns capturing tractable, commercially valuable information to support marketing decisions and strategies. Through this use case, we aim to validate the Dicode suite of services for the automatic analyses of voluminous amounts of unstructured information from Web 2.0 platforms, blogging platforms (Twitter), and social network.

dicode

**Mastering
Data-Intensive
Collaboration and
Decision Making**

Data mining on unstructured and multi-faceted data

Collaboration and decision making support

High performance Data Mining

<http://dicode-project.eu>

Figure 10: Dicode leaflet, outer side

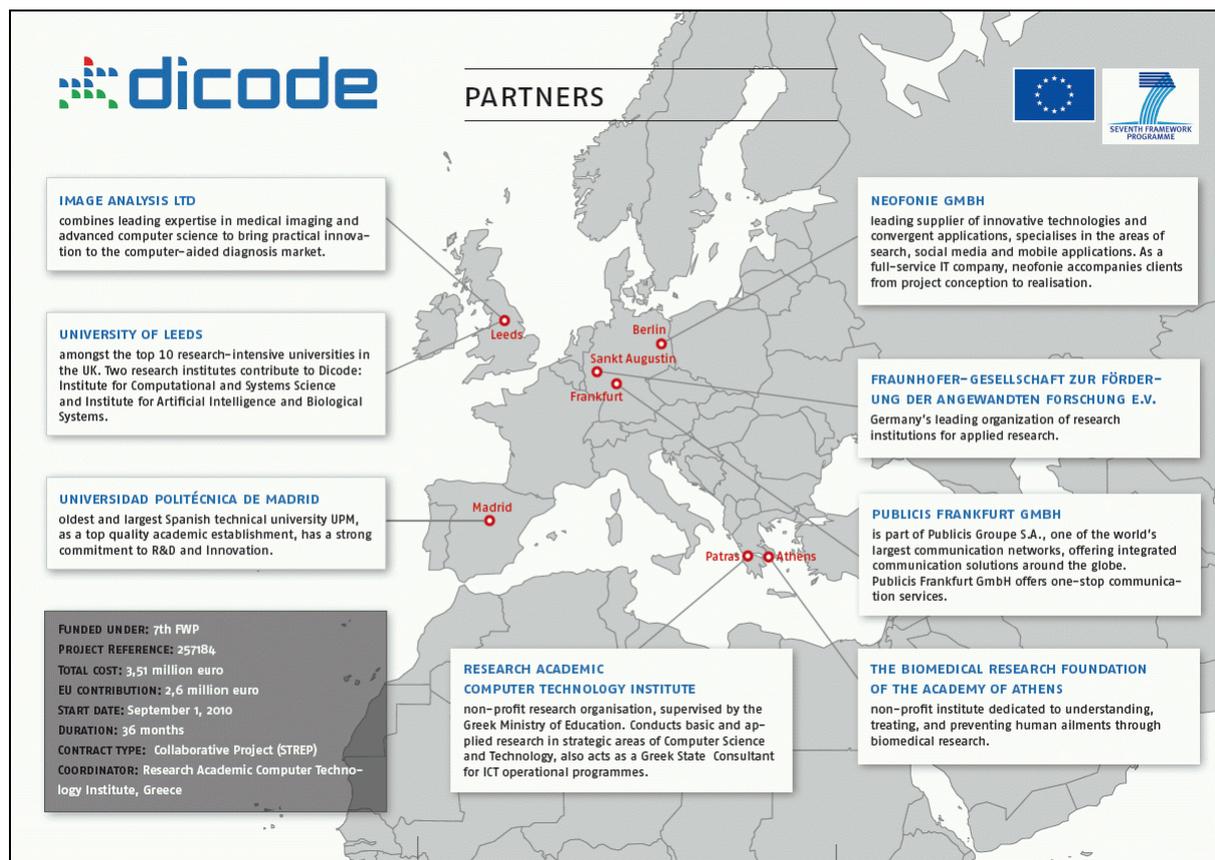


Figure 11: Dicode leaflet, inner side

3.3 Social Media

Social media plays an important role for Dicode, both as an instrument used for dissemination and as an area of research and development in Use Case 3. Besides the project web page Dicode uses the following social media services and platforms for dissemination:

3.3.1 Twitter

Dicode uses Twitter for status updates about current events and for networking within the open source community (see http://twitter.com/DICODE_EU). The list of followers shows that there is a global interest in Dicode's activities. Most followers are involved in open source development.



Figure 12: Selection of Dicode’s followers on Twitter

3.3.2 Facebook

Dicode has a Facebook profile and a Facebook group – for more details, see <http://www.facebook.com/#!/profile.php?id=100001390513581>. The following screenshot shows an extract of Dicode’s wall at Facebook.



Figure 13: Dicode’s Wall at Facebook

Dicode's Facebook group automatically displays all Tweets from Dicode's Twitter account.
<http://www.facebook.com/#!/group.php?gid=124336097612281>

3.3.3 Other tools and platforms

Additionally, Dicode uses platforms like CiteULike (http://www.citeulike.org/user/DICODE_EU) and LinkedIn (<http://de.linkedin.com/pub/dicode-eu/24/979/688> (refer to D7.1: "Web portal" for more information on the use of Social Media in the Dicode project)).

4 Dissemination activities report

This section gives a comprehensive account of Dicode's dissemination activities in the first half of the project duration (September 2011 to February 2012). During this period, Dicode:

- reported early efforts at international workshops and conferences in relevant research communities,
- discussed Dicode approach at open source conferences and workshops,
- presented Dicode to Marketing and Social Media professionals, and
- improved Dicode vision together with customers and partners.

A significant number of publications have been published by the end of the first year. Dicode received a best research paper award for one paper. Dicode had also the chance to present itself at some of the top conferences on open source high scalability technologies and semantics.



Figure 14: Dicode at international conferences

4.1 Publications Strategy

Publication of the project's scientific results is of high importance towards augmenting its visibility. A publication strategy has been sketched from the very beginning of the Dicode project. This strategy is based on the following points:

- Key research areas, where Dicode contributes, will be regularly identified and associated to specific journals and international conferences and related events (target list);
- The publication target list will be reviewed by the Steering Committee and further discussed at Dicode consortium meetings;
- Partners are encouraged to collaborate towards preparing joint publications;
- Key parts of Dicode deliverables should be written in a way that allows their easy reuse in associated publications for peer reviewed journals and conferences;
- Publications will be regularly uploaded at the Dicode web site (Dissemination page).

4.2 List of publications

- N. Karacapilidis, S. Rueping and I. Drost: [Mastering Data-Intensive Collaboration and Decision Making through a Cloud Infrastructure](#). ERCIM News, No 83, October 2010, pp. 49-50.
- M. García-Remesal, A. Cuevas, D. Pérez-Rey, L. Martín, A. Anguita, D. de la Iglesia, G. de la Calle, J. Crespo and V. Maojo: [PubDNA Finder: a web database linking full-text articles to sequences of nucleic acids](#). Bioinformatics. November 2010, 26(21):2801-2802. Available from: <http://dx.doi.org/10.1093/bioinformatics/btq520>
- V. Maojo, J. Crespo, M. García-Remesal, D. de la Iglesia, D. Perez-Rey and C. Kulikowski: [Biomedical Ontologies: Toward Scientific Debate](#). Methods of information in medicine. May 2011, 50(3):203-216. Available from: <http://dx.doi.org/10.3414/ME10-05-0004>
- A. Paramythis, L. Lau, S. Demetriadis, M. Tzagarakis and S. Kleanthous (eds.): [Proceedings of the International Workshop on Adaptive Support for Team Collaboration 2011 \(ASTC 2011\)](#) - held in conjunction with the International Conference on User Modeling, Adaptation, and Personalization (UMAP 2011), Girona, Spain, July 15, 2011, CEUR Workshop Proceedings, Vol. 743.
- A. Ammari, V. Dimitrova, L. Lau, M. Tzagarakis, and N. Karacapilidis: [Augmented Collaborative Spaces for Collective Sense Making: The Dicode Approach](#). In: A. Paramythis, L. Lau, S. Demetriadis, M. Tzagarakis, & S. Kleanthous (Eds.), Proceedings of the International Workshop on Adaptive Support for Team Collaboration 2011 (ASTC 2011), Girona, Spain, July 15, 2011, CEUR Workshop Proceedings, Vol. 743, pp. 3-13.
- A. Ammari, L. Lau and V. Dimitrova, Distributed Data Mining for User Sensemaking in Online Collaborative Space. Dicosyn Workshop 2012 in conjunction with the 2012 ACM Conference on Computer Supported Cooperative Work, 11-15 Feb, 2012, Seattle, USA.
- N. Karacapilidis: [Mastering Data-Intensive Collaboration and Decision Making through a Cloud Infrastructure: The Dicode EU project](#). EMBnet.journal, Vol. 17, No 1, July 2011, p. 3
- N. Karacapilidis, S. Rüping, M. Tzagarakis, A. Poigné and S. Christodoulou: [Building on the synergy of machine and human reasoning to tackle data-intensive collaboration and decision making](#). In: J. Watada, G. Philips-Wren, L.C. Jain and R.J. Howlett (eds.), Intelligent Decision Technologies - Proceedings of the 3rd International Conference on Intelligent Decision Technologies (IDT 2011), Piraeus, Greece, July 20-22, 2011, Springer-Verlag, Berlin, Smart Innovation, Systems and Technologies, Vol. 10, pp. 113-122. [Best Research Paper Award](#)
- F. Yang-Turner and L. Lau: A Pragmatic Strategy for Creative Requirements Elicitation. In Proceedings of the 1st International Workshop on Requirements Engineering for Systems and Systems-of-Systems. Held in conjunction with the 19th IEEE International Requirements Engineering Conference, 29 Aug - 2 Sept, 2011, Trento, Italy.
- F. Yang-Turner and L. Lau: Extending Use Case Diagrams to Support Requirements Discovery. In Proceedings of the 1st International Workshop on Requirements Engineering for Systems and Systems-of-Systems. Held in conjunction with the 19th

4.3 Talks and presentations

- “Scaling Data Analysis with Apache Mahout” - presentation by Isabel Drost at O'Reilly Strata, Santa Clara, February 2011.
- “Intelligent data analysis - Apache Mahout” - presentation by Isabel Drost at Devovx, Antwerp, November 2010.
- “Scaling Intelligent Data Analysis with Apache Mahout” - presentation by Isabel Drost at codebits, Lisbon, November 2010.
- “Apache Mahout - making data analysis easy” - presentation by Isabel Drost at Apache Con US, Atlanta, November 2010.
- “Unite to Triumph and Divide to Conquer: Intuitive, Iterative, and Modular Ontology Authoring” - keynote talk by Vania Dimitrova at the 3rd International Conference on Software, Services & Semantic Technologies, September 1-3, 2011, Bourgas, Bulgaria (presenting the work on ROO, DON, and Augmentor).
- “Social Media Monitoring as illustrated by Publicis Conversation Reader” - short presentation by Ralf Löffler of the Dicode Approach, Anhalt Universtiy, Bernburg, May 2011.
- Panel discussion at the Marketing Horizonte, Cologne, Germany, September 2011 (Ralf Löffler as participant).
- “I love this brand”, short presentation by Ralf Löffler of the Dicode approach at the Marketing Horizonte, Cologne, Germany, September 2011.
- “Entendre: Interactive Semantic Feedback for Ontology Authoring”, demo by R. Denaux, D. Thakker, V. Dimitrova and A.G. Cohn. 10th International Semantic Web Conference (ISWC 2011), Bonn, Germany, October 2011.



Figure 16: Dicode at Open Source Conferences

4.4 Prototypes and Source Code

This section lists Dicode's open source artefacts and the contributions to exiting free software projects.

4.4.1 Repository at Google Code

The Dicode Analysis project containing modules for data acquisition, data processing and data services is available at Google code.

- Dicode source code repository (WP3): <http://code.google.com/p/dicode/>

Users interested in the project can track changes easily by subscribing to the Atom feed from <http://code.google.com/p/dicode/feeds> or by adding Gadgets (http://www.google.com/ig/add?moduleurl=http://google-code-project-hosting-gadgets.googlecode.com/svn/build/prod/changes/gcChanges.xml&up_projectName=dicode) to their page. An extract of an example Atom feed is shown below:



Figure 17: Atom Feed showing subversion commits of Dicode

4.4.2 Social Coding: GitHub

Since the start of Dicode, there has been a noticeable shift in open source development towards platforms which add networking support for software developers. One of the most popular platforms is GitHub (<https://github.com/>), a project hosting platform for projects using the Git version control system.

In contrast to other open source hosting platforms, GitHub focusses on users and user repositories. GitHub encourages "Forks"⁶. Users interested in working on a specific project can create their own version of it as a fork. After adding new features, the developer can issue a pull request to the creator of the project. If the project owner approves the changes, he can pull the committed changes to his project.

⁶ http://en.wikipedia.org/wiki/Fork_%28software_development%29

Dicode started using GitHub in November 2011. As a first project, pignlproc (<https://github.com/ogrisel/pignlproc>) was forked, a project which provides utilities for in Apache Pig for building training corpora for machine learning and NLP based on publicly available Wikipedia and DBpedia⁷ dumps. Dicode added a set of features, e.g. support for redirect resolution while parsing the German Wikipedia dump. After completing a set of features which could be interesting for the owner of the original project, Dicode can issue a pull request by using the following form:

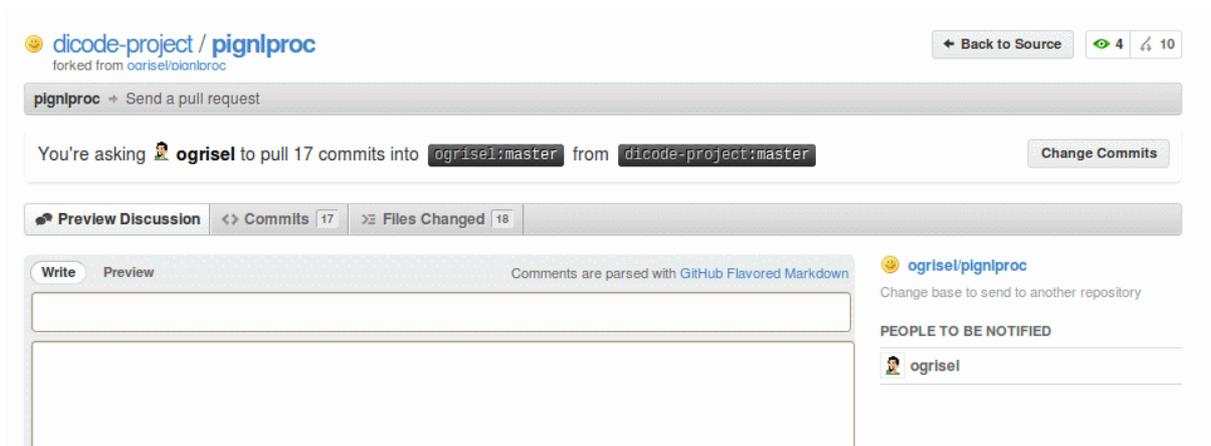


Figure 18: A pull request in GitHub

In huge projects, the graph of forks and merges can become quite big. GitHub provides a timeline which visualises all the network of each project as shown in the following figure.

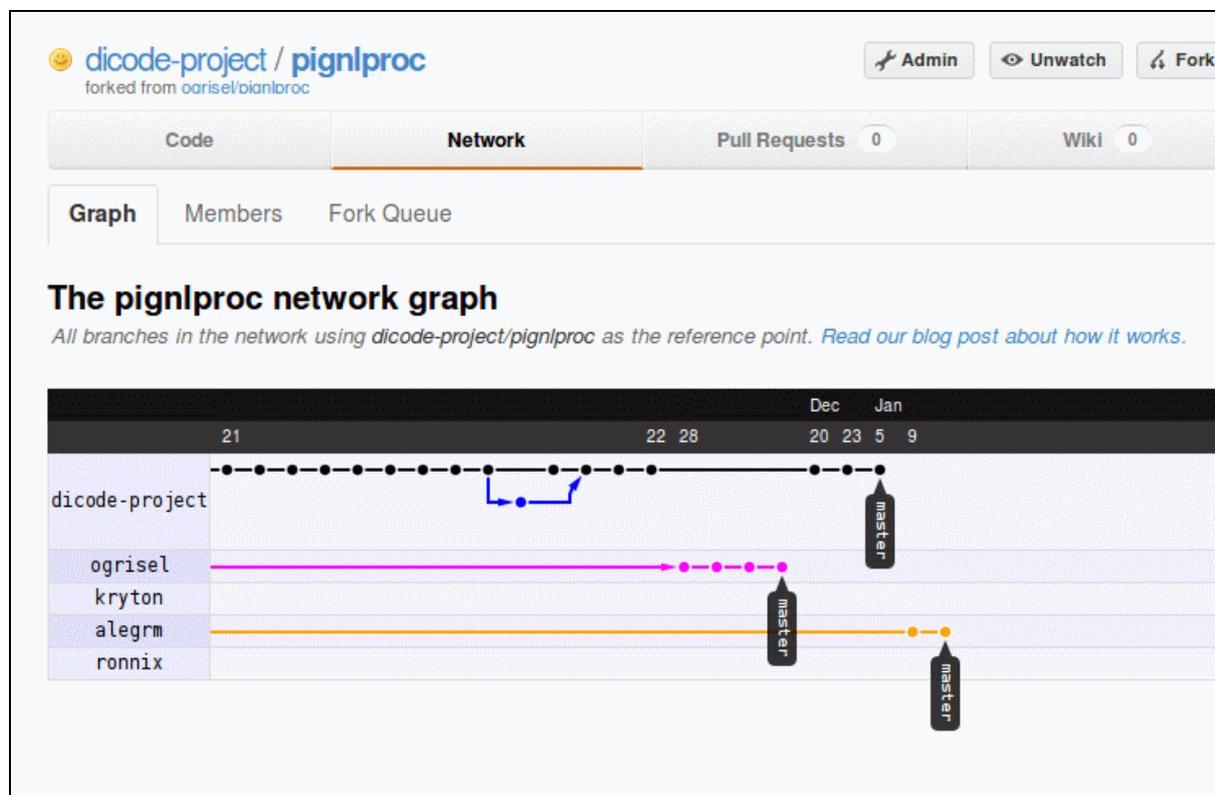


Figure 19: GitHub network graph

⁷ <http://dbpedia.org/downloads>

4.4.3 Ontology resources

Dicode's ontology resources are available publicly:

- DON (Dicode ONtology): <https://sites.google.com/site/ontomatic/don>
- Dicode Community Modeling Resource: <https://www.dicode-project.eu/index.php?q=node/48>

4.4.4 Contributions to free software projects

Dicode contributed several patches to free software projects.

- [MAHOUT-560](#) improved extensibility of Apache Mahout with respect to text processing for co-location analysis and LDA.
- [MAHOUT-595](#) improved deployment of Mahout artifacts (now also includes reference to source and JavaDoc).
- [MAHOUT-590](#) adds TSV (Tab Separate Value) input file support to `SequenceFilesFromDirectory`.

4.5 Selected Events and Competitions

4.5.1 Semantic/NLP Hackathon

Dicode used the Berlin Buzzwords conference, the biggest European conference on high scalability development based on Hadoop, as a platform for networking and collaboration with the open source community.

A hackathon organized and hosted by Neofonie GmbH (NEO) brought together developers interested in Hadoop-based technologies from various countries and backgrounds ("Semantic/NLP Hackathon" at "Berlin Buzzwords - The Conference of High Scalability" - Conduct of Hackathon on Semantics and NLP, June 8-9, 2011). NEO aims at hosting a regular yearly hackathon with a focus on Dicode technologies in conjunction with the Berlin Buzzwords conference.

NEO has also been working towards networking and strengthening of links with the open source community through participation in the Apache HBase Hackathon at Berlin Buzzwords and review of the book "HBase - The Definite Guide" published by O'Reilly.

4.5.2 Hadoop Get Together

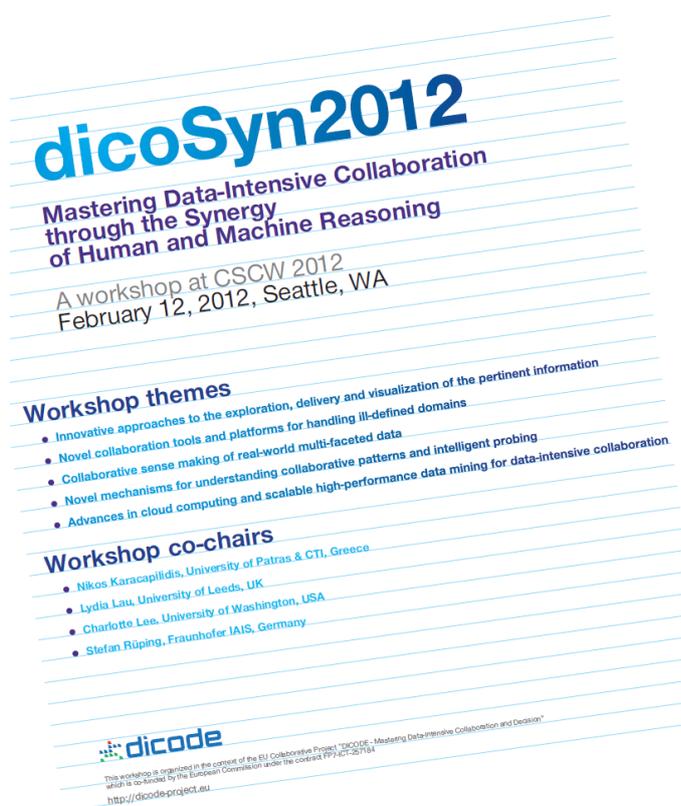
Members of Dicode are regular guests at the "Hadoop Get Together" in Berlin, a quarterly meet-up where developers and scholars involved in Hadoop-related projects discuss their work. In December 2011, Max Jakob from NEO's Dicode team gave a hands-on presentation on Apache Pig for Natural Language Processing at the Hadoop Get Together in Berlin (<http://vimeo.com/33394778>). For a detailed report on the Get Together see:

<http://blog.isabel-drost.de/index.php/archives/343/apache-hadoop-get-together-berlin-december-2011>.⁸

4.5.3 Discovery Challenge

Dicode took part in the Discovery Challenge (<http://www.ecmlpkdd2011.org/challenge.php>) at the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases ECML PKDD 2011. The Similarity Learning Service which was developed as a Dicode Data Mining Service (see WP3 for details) received a medium score in recommending videos in the context of VideoLectures.net while the main challenge was the cold start problem.

4.6 The Dicode workshops



Dicode carries out two scientific workshops, one in the second and one in the third year. The proceedings of both workshops will be published as Deliverables D7.3.1 and D7.3.2. The first workshop “dicoSyn 2012 - Mastering Data-Intensive Collaboration through the Synergy of Human and Machine Reasoning” (<http://dicode.cti.gr/dicosyn12/>) took place in Seattle, in the context of CSCW 2012, during the preparation of this deliverable. A poster was designed for the needs of the workshop, which appeared at a dedicated wall during the conference (see photo on the left).

dicoSyn 2012 brought together researchers and practitioners from different scientific fields and research communities to further explore the synergy between human and machine intelligence, and larger issues surrounding analytical practices and

data sharing practices. Research areas included:

- Innovative approaches to the exploration, delivery and visualization of the pertinent information
- Novel collaboration tools and platforms for handling ill-defined domains
- Collaborative sense making of real-world multi-faceted data
- Novel mechanisms for understanding collaborative patterns and intelligent probing
- Advances in cloud computing and scalable high-performance data mining for data-intensive collaboration

Workshop co-chairs were:

⁸ The schedule of the Hadoop Get Together can be accessed via the Scalable Information Retrieval Group at the business networking platform XING. <http://www.xing.com/net/informationretrieval>

- Nikos Karacapilidis, University of Patras & CTI, Greece
- Lydia Lau, University of Leeds, UK
- Charlotte Lee, University of Washington, USA
- Stefan Rüping, Fraunhofer IAIS, Germany

Eight papers were accepted for presentation. These were:

- Post-Doctoral Researchers' Use of Preexisting Data in Cancer Epidemiology Research, by Betsy Rolland & Charlotte P. Lee
- CoPEXplorer: A Novel Collaboration Platform for Quantitative Proteomics Research in China, by Tun Lu, Xianghua Ding, Xing Huang, Guangmeng Zhai, Zhaocan Chen & Ning Gu
- Two Socio-Technical Gaps of Cyberinfrastructure Development and Implementation for Data-Intensive Collaboration and Computational Simulation in Early e-Science Projects in the U.S., by Kerk F. Kee & Larry D. Browning
- Emergent Properties of Data-Driven Technological Development in Science, by Stephanie Gokhman
- Striving toward ad hococracy in dataland, by Ben Li
- On the identification and integration of countermeasures to cope with data intensiveness in collaboration settings, by Manolis Tzagarakis, Spyros Christodoulou & Nikos Karacapilidis
- Distributed Data Mining for User Sensemaking in Online Collaborative Spaces, by Ahmad Ammari, Lydia Lau & Vania Dimitrova
- Distance Metric Learning for Recommender Systems in Complex Domains, by Natalja Friesen & Stefan Rüping

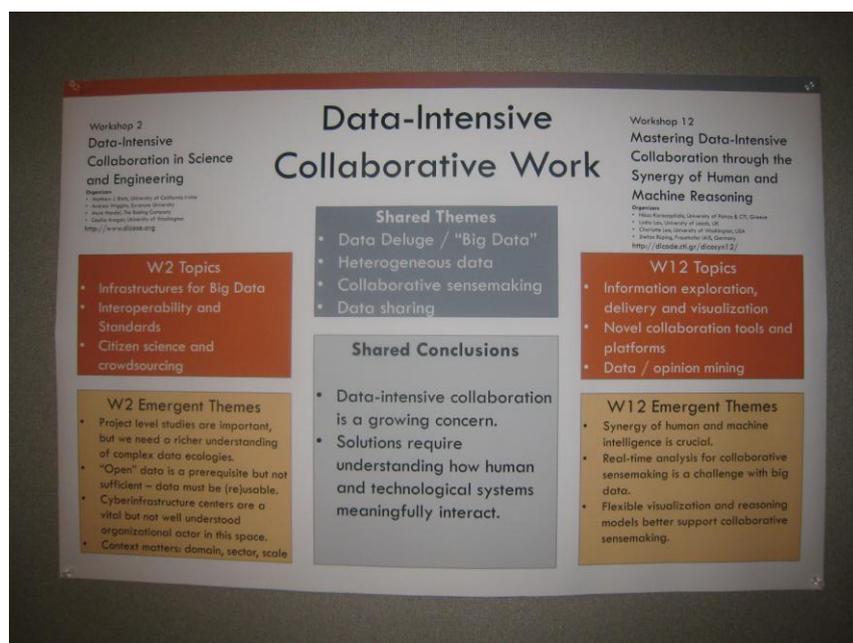
An influential keynote, titled "Enhancing a collaboration space with social media features to support data-intensive collaboration", was also given by Prof. Wolfgang Prinz. In his talk, Prof. Prinz presented the approach and challenges of the EU-funded Granatum project, which aims at the development of a cooperation environment for chemoprevention researchers. This keynote was the first action towards establishing a close collaboration between the Dicode and Granatum projects.



A very productive session, held after the paper and keynote presentations, concerned group discussion on cross cutting themes; this resulted to the collaborative production and elaboration of very useful insights for future work on the workshop's research issues (see photo on the left). The proceedings of the dicoSyn 2012 workshop are in preparation. Plans for producing a special

issue at an international journal – with extended versions of the dicoSyn 2012 papers - are also in progress.

At CSCW 2012, there was a complementary workshop, titled “Data-Intensive Collaboration in Science and Engineering” (<http://www.dicose.org>), which had a broad focus on issues of data-intensive collaboration, and leans more toward social and organizational issues. The two workshops were organized independently, but shared a common theme. The organizers coordinated their planning so that participants could attend just one or the other, but participants should find it valuable to attend both. Participants were encouraged to attend



both events, while organizers attended both. The two workshops shared a list of themes and conclusions and were presented next to each other in the related session of the main CSCW conference (see photo on the left). Through this complementary workshop, further dissemination of the Dicode results and progress was achieved (mainly addressing researchers and academic faculty based on USA).

4.7 Collaboration with relevant EU projects and initiatives

The Dicode consortium has established collaboration with the following European projects and initiatives:

- Co-operation with the FP7 project (IP) “p-medicine” (<http://www.p-medicine.progressima.eu/>): Fraunhofer and UPM are partners of this project, working on workflow patterns for data mining in personalised medicine. Foreseen collaboration concerns architectures and tools for parallel knowledge discovery.
- Co-operation with the project “e-LICO” (<http://www.e-lico.eu/>): It concerns experimentation with RapidAnalytics Data Mining Server.
- Co-operation with the organizers of the CSCW 2012 workshop on Data-Intensive Collaboration in Science and Engineering (<http://www.dicose.org/>).
- Co-operation with the SEQAHEAD project (<http://www.nextgenerationsequencing.org/>): BRF is a partner of this project, working on next generation sequencing data management and analysis.
- Co-operation with the “EDGE” project (http://www.bioacademy.gr/bioinformatics/EDGE/bioinformatics_home.htm). It concerns the development of a Greek national network for genomics research (BRF).

- Co-operation with the GRANATUM project (<http://granatum.org/>) (CTI, see also Section 4.6).
- Co-operation with the MIA project (<http://www.trusted-cloud.de/de/778.php>) (NEO).
- People from UPM, Fraunhofer and CTI participate as members of the scientific committee in the organisation of a track session on cloud computing at the IEEE CMBS 2012 conference.
- Cooperation with the EU FP7 IMREAL project (<http://www.imreal-project.eu>) on the use of social media content to augment the support for human tasks. (UOL)
- Cooperation with the UK Technology Strategy Board STRAPP project (<https://www.engineering.leeds.ac.uk/strapp>) on trusted digital spaces through timely reliable and personalised provenance. (UOL)

5 Intellectual property rights strategy

The IPR policy of the Dicode project is established with the objective of ensuring an efficient cooperation for carrying out the project and finding the right balance between protection of knowledge and dissemination to the benefit of the partners and for the interest of the European Commission.

Dicode's rules for the management of intellectual property within the frame of the project and for future exploitation of project's results are laid down in the contract between the Coordinator and the European Commission and the Consortium Agreement. In this section, the core aspects of Dicode's IPR strategy are discussed in depth.

5.1 General objectives and principles

Dicode's IPR strategy has been established with the objective of ensuring an efficient cooperation for carrying out the Dicode project and finding the right balance between protection of knowledge and dissemination to the benefit of the partners and for the interest of the European Commission. The associated rules concern the following:

- **Ownership:** Knowledge and know-how will be the intellectual property of the institution carrying out the work leading to that knowledge. Knowledge and results that are developed jointly will be jointly owned.
- **Scientific publications:** Academic partners will have the right to publish scientific results of the project having the obligation to explicitly mentioning the complete framework of the project. If these results are joint intellectual property, the parties involved will have the right to review drafts and have the opportunity to express their interest in the protection of specific results.
- **Types of know-how:** Knowledge and know-how can be abstract knowledge, general specifications, detailed specifications of data and processes, source code, and executable code.
- **Access rights to knowledge and pre-existing know-how:** All consortium partners will have access rights to knowledge arising from the project during the entire project period, for research and development purposes. Also, they will have access to pre-existing know-how of other partners explicitly listed in the consortium agreement, unless something is excluded or its access is restricted in the consortium agreement.
- **Dissemination of knowledge:** The consortium actively disseminates the Dicode project's knowledge beyond the consortium frame.

5.2 Licensing strategy

Dicode's IPR strategy follows an open source approach. All core components have to be based on open standards and open source software. Additional components, services and data sets can be published under a more restricted license.

5.2.1 Open Source Licensing for base components

During development of core components, proprietary software shall not be used. Core components shall be published under an Apache compatible open source license:

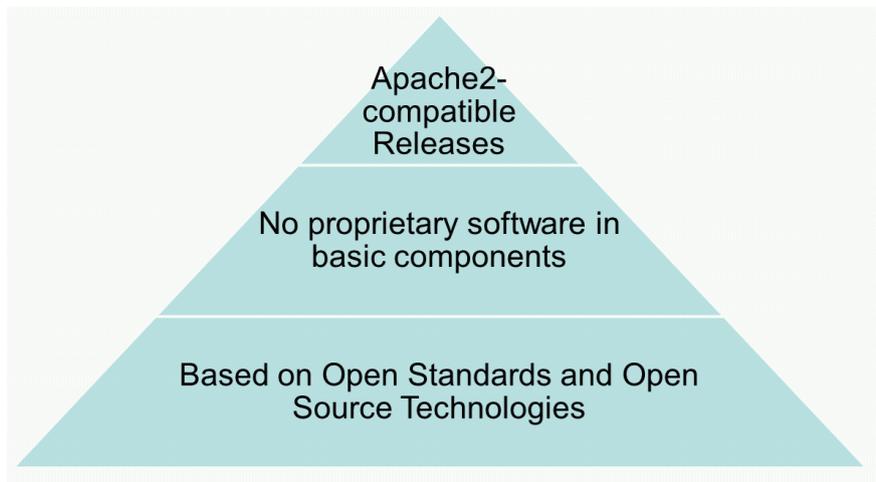


Figure 20: Dicode's basic open source principles

In the DoW, these principles are explained more detailed:

- “The Dicode platform will be developed using existing open source software, existing standards and existing custom web-technology.
- During the development of the basis technology, proprietary software, patented technology and other technology based on intellectual property rights will not be used.
- The basis software will be released under an open source license, which must be compatible with the Apache license⁹, as several predefined basis technologies, like Hadoop, Mahout and UIMA, already fall under this license.”¹⁰

The Dicode platform is developed following the idea of a Service Oriented Architecture.¹¹ Services developed within Dicode also have to conform to the open source principles defined above, as stated in the DoW: “Services to be developed and integrated in the context of the Dicode project will be released under an open source license.”¹²

Practically, Dicode's open source strategy implies the following:

- All presentation materials, for which this is appropriate, will be published on the project's web site under a Creative Commons license (<http://creativecommons.org/>) or another appropriate license.
- All open source software produced by the consortium will be published on publicly available software repositories such as <http://sourceforge.net/>, <http://github.com/> or <http://osor.eu/>.
- All data sets, for which this is appropriate, will be published on the project's web site under a Creative Commons license (<http://creativecommons.org/>) or another appropriate license.

5.2.2 Proprietary licenses for additional modules

Dicode restricts the development of proprietary software to “additional modules”. Whereas the basic components have to be free software, additional modules can be published under

⁹ Compatible licences are for example the MIT and BSC licences.

¹⁰ Dicode Grant Agreement, Annex 1 Description of Work, Part B, p. 94.

¹¹ See D3.1.1 – The Dicode Data Mining Framework for conceptual details and D5.4.1 – Integrated Dicode Services for a possible integration platform.

¹² Dicode Grant Agreement, Annex 1 Description of Work, Part B, p. 3

more restrictive licenses. This allows the Dicode consortium to financially profit through a variety of revenue streams, such as:

- Licensing of additional modules, which extends the functionality and capability of the platform. These modules will be proprietary, and will not be under an open source license. For example, better monitoring and profiling modules for the system;
- Licensing of data and text processing modules. For instance, a person names recognition module in different languages or a sophisticated sentiment detection module for the German language. These modules will also be proprietary, and not be under an open source license;
- The development of additional applications based upon the core functionality of the Dicode platform which would benefit and extend the consortium members' own line of business;
- Data as a Service (DaaS): Cloud application developers can use existing data which is provided by other resellers. Data can be licensed as a basis for applications. For example, a company can harvest "news" via web-spidering, process and annotate it with additional meta-data such as detected named entities, geographical information etc. This annotated news feed can then be subscribed to and licensed by other companies which, for example, develop applications for "political trend detection" based on media observation;
- Data sources, such as new feeds which are enriched with meta-data.

6 Outlook

From its very beginning, Dicode put much effort into planning for dissemination and exploitation. This strategy has paid off. During the first half of the project, Dicode had the chance to disseminate the project results to a broad audience. Dicode used the opportunity to get feedback from the software and the user as soon as possible. Scientific results were presented to the respective community right away, which resulted in fruitful discussions and improvements of Dicode's approach.

The following graph shows the expected relative impact of Dicode for research, open source communities and industry. While the impact in research grows steadily, the impact both in open source development and in industry increases slowly in the beginning. For year three, a big leap in growth in both areas can be expected for several reasons: In year three open source developers will be able to use Dicode's comprehensive software frameworks. Software developers and system integrators will be able to develop solutions based on the Dicode framework and the Dicode suite of services.

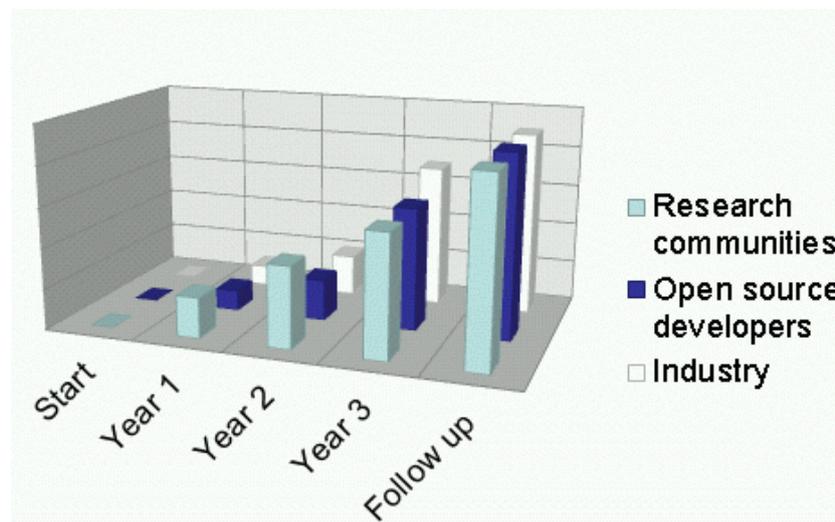


Figure 21: Estimated timeline of Dicode's impact

Dissemination and exploitation activities will slightly shift towards open source communities and the industry in the second half of the project. In general, Dicode will stick to the dissemination strategy laid out in the DoW.

Figures and Tables

Figure 1: Dicode's users	7
Figure 2: Distribution of person month in WP7.....	8
Figure 3: Dicode's main research areas.	10
Figure 4: Fields of research in Dicode.....	10
Figure 5: Dicode targets a wide audience	12
Figure 6: The Dicode web page	14
Figure 7: Connect with Dicode	14
Figure 8: Blog posts	15
Figure 9: List of publications, presentations and software contributions	15
Figure 10: Dicode leaflet, outer side.....	16
Figure 11: Dicode leaflet, inner side.....	17
Figure 12: Selection of Dicode's followers on Twitter.....	18
Figure 13: Dicode's Wall at Facebook.....	18
Figure 14: Dicode at international conferences	20
Figure 15: Dicode's publication word cloud.....	22
Figure 16: Dicode at Open Source Conferences	23
Figure 17: Atom Feed showing subversion commits of Dicode	24
Figure 18: A pull request in GitHub	25
Figure 19: GitHub network graph.....	25
Figure 20: Dicode's basic open source principles	32
Figure 21: Estimated timeline of Dicode's impact.....	34
Table 1: Tasks in WP7	8
Table 2: Deliverables in WP7	8
Table 3: Dissemination activities matrix	13