A Semantic, Task-Centered Collaborative Framework for Science

Yolanda Gil[1], Felix Michel[1], Varun Ratnakar[1], Matheus Hauder[2]

[1]Information Sciences Institute, University of Southern California,
Marina del Rey, CA 90292, US
`{gil, felixm, varunr}@isi.edu`

[2]Software Engineering for Business Information Systems, Technical University Munich,
Munich, BY 85748, DE
`hauder@in.tum.de`

Abstract. This paper gives an overview of the Organic Data Science framework, a new approach for scientific collaboration that opens the science process and exposes information about shared tasks, participants, and other relevant entities. The framework enables scientists to formulate new tasks and contribute to tasks posed by others. The framework is currently in use by a science community studying the age of water, and is beginning to be used by others.

Keywords: Organic Data Science, Semantic Wiki, collaborative web platforms

1 Introduction

The Web was originally developed to support collaboration in science. Although scientists benefit from many forms of collaboration on the Web (e.g., blogs, wikis, forums, code sharing, etc.), most collaborative projects are coordinated over email, phone calls, and in-person meetings.

We are interested in supporting scientific collaborations where joint work occurs on a concrete problem of interest, with many participants, and over a long period of time. Although the Web may be used to share information, there is no explicit support for the shared tasks involved. These tasks are discussed through email, phone calls, and occasional face-to-face meetings.

2. Organic Data Science

We are developing an Organic Data Science framework based on a task-centered organization of the collaboration, and that includes principles from social sciences for successful on-line communities. Figure 1 illustrates the representation of a task, which includes properties such as the owner, participants, start and end times, and expertise required. Users can create additional task properties, as is typical in semantic wikis. They can add subtasks, and sign up as participants in tasks created by others. These tasks capture the what, who, when, and how of the activities pursued by the collaboration, and capture a novel form of science processes that has not been explicitly captured before.

Our Organic Data Science framework is implemented as an extension of a semantic wiki, in particular the Semantic MediaWiki platform [1]. Users can add properties to tasks as needed, and can describe any entity of interest to the collaboration (datasets, software, papers, etc.) using semantic properties of the wiki. Every task has its own page, and therefore a unique URL, which gives users a way to refer to the task from any other pages in the site as well as outside of it. Semantic wikis provide an easy-to-use interface where users can define structured properties, which are then represented in RDF and exported as linked data. The framework is still under development, and it evolves to accommodate user feedback and to incorporate new collaboration features.

We view the scientific collaboration as an on-line community, and have designed the Organic Data Science framework following social design principles uncovered by research on successful on-line communities [2]. The community design aspects are described in [3]. The semantic aspects of the design are described in [4].

The Organic Data Science framework captures science processes that are not made explicit in publications, supports the formation of ad-hoc groups to work on tasks of interest, enables anyone to contribute to tasks that match their interests, and advertises ongoing work to potential newcomers.

Fig. 1: Collaboration is organized around tasks in the Organic Data Science framework, represented through RDF properties in the underlying semantic wiki. These semantic properties are used by the system to generate content for other pages, such as the user page shown in the right side based on the tasks that this user is participating in. Status icons, shown as pie charts next to tasks, are also derived from the semantic properties that specify task type and deadlines.

Table 1 illustrates the features of the Organic Data Science framework (shown at the bottom), compared other collaborative tools on the Web that scientists use. Our framework is the only one that is designed to support on-line communities, is organized around tasks, and captures semantic structures for the entities involved in the collaboration.

Table 1: An overview of features supported by existing online collaboration tools.

Category	Tool	Community support	Task-oriented organization	Semantic structures	Link (last accessed on March 2015)
Forum	V-Bulletin Forum	Yes	No	No	http://www.vbulletin.com/forum
	WordPress bbPress	No	No	No	http://bbpress.org/
	WordPress p2 Theme	No	No	No	http://wordpress.org/themes/p2
	Stackoverflow	Yes	No	No	http://stackoverflow.com/
	Mathematics	No	Yes	No	http://math.stackexchange.com/
Wiki	MediaWiki	Yes	No	No	http://www.mediawiki.org/
	Semantic MediaWiki	Yes	No	Yes	https://semantic-mediawiki.org/
Enterprise Knowledge Management	Confluence	Yes	Yes	No	https://www.atlassian.com/de/ software/confluence
	Connections	No	Yes	No	http://www-03.ibm.com/ software/products/en/conn
	Jive	No	Yes	No	http://de.jivesoftware.com/
	MS SharePoint	No	Yes	No	http://office.microsoft.com/en-us/sharepoint/
	Communote	No	No	No	http://www.communote.com/ homepage/
	Yammer	No	No	No	https://www.yammer.com/
	Redmine	Yes	Yes	No	http://www.redmine.org/
ToDo Lists	Trello	No	Yes	No	https://trello.com/
	Keep	No	Yes	No	https://keep.google.com/keep/
	Todoist	No	Yes	No	http://todoist.com/
Repositories	GitHub	Yes	No	No	https://github.com/
	BitBucket	Yes	No	No	https://bitbucket.org/
	Google Code	Yes	No	No	https://code.google.com/
Organic Data Science	Organic Data Science Framework	Yes	Yes	Yes	http://organicdatascience.org/

3. Supporting Scientific Collaboration

The major user of the Organic Data Science framework is a community of hydrologists and limnologists that are studying the age of water in an ecosystem. Other communities are beginning to use the framework for neuroscience and geosciences research.

Table 2 shows the tasks defined so far in the age of water collaboration. In a 10-week time period started on 1st of October 2014, all task pages together were accessed more than 2,900 times. Person pages were accessed 328 times. We logged in total more than 19,000 events as users interacted with the system. Users have defined a total of 1047 RDF triples.

Table 2: Owner and participant's distribution of task types.

Type of task	Total tasks	Total owners	Total participants
High-level	7	7	8
Medium-level	110	109	82
Low-level	237	237	24
Unspecified	24	20	7

The Organic Data Science framework provides live collaboration data evaluation for every community with a dashboard. A community dashboard includes a collaboration graph and statistical task metadata. The oldest ODS community is the age of water community[1], followed by the Organic Data Science framework development community, geosciences community[2] and the private neuroscience community.

Acknowledgements

References

1. M. Krötzsch, D. Vrandecic, M. Völkel, H. Haller, and R. Studer. "Semantic Wikipedia." Journal of Web Semantics, 5(4), 2007.
2. R. E. Kraut and P. Resnick. "Building Successful Online Communities." IT Press, 2011
3. F. Michel, Y. Gil, V. Ratnakar and M. Hauder. "A Task-Centered Interface for On-Line Collaboration in Science." In Proceedings of the ACM International Conference on Intelligent User Interfaces, Atlanta, GA, 2015.
4. Y. Gil, F. Michel, V. Ratnakar, J. Read, M. Hauder, C. Duffy, P. Hanson, and H. Dugan "Supporting Open Collaboration in Science through Explicit and Linked Semantic Description of Processes." In Proceedings of the European Semantic Web Conference (ESWC15), Portoroz, Slovenia, 2015.

---

[1] http://www.organicdatascience.org/ageofwater/index.php/Special:WTDashboard
[2] http://www.organicdatascience.org/gpf/index.php/Special:WTDashboard