Organic Data Science: A Task-Centered Interface to On-Line Collaboration in Science

Yolanda Gil,
Felix Michel, Varun Ratnakar
Information Sciences Institute
University of Southern California
Marina del Rey, CA 90292, US
{gil, felixm, varunr}@isi.edu

Matheus Hauder
Software Engineering for
Business Information Systems
Technical University Munich
Munich, BY 85748, DE
hauder@in.tum.de

ABSTRACT

Although collaborative activities are paramount in science, little attention has been devoted to supporting on-line scientific collaborations. Our work focuses on scientific collaborations that revolve around complex science questions that require significant coordination to synthesize multi-disciplinary findings, enticing contributors to remain engaged for extended periods of time, and continuous growth to accommodate new contributors as needed as the work evolves over time. This paper presents the interface of the Organic Data Science Wiki to address these challenges. Our solution is based on the Semantic MediaWiki and extends it with new features for scientific collaboration. We present preliminary results from the usage of the interface in a pilot research project.

Author Keywords

Collaboration interfaces; Organic Data Science; Semantic MediaWiki; Scientific collaboration

ACM Classification Keywords

H.5.3. Information Interfaces and Presentation (e.g. HCI): Group and organization interfaces.

INTRODUCTION

Science has become an increasingly collaborative endeavor and typical revolves around sharing instruments, shared database, shared software base, and shared scientific question (e.g., the Human Genome Project). Our work focuses on scientific collaborations that are driven by a shared scientific question that requires the integration of ideas, models, software, data, and other resources from different disciplines. These projects are particularly challenging because they require:

- *significant organization and coordination*, as people with diverse backgrounds are supposed to first discover one another and then find common ground to collaborate
- *retaining users over the long term*, since people need clear incentives to remain involved for the long period of time that such projects are active
- *incrementally growing the community* with unanticipated participants, as they bring in skills or resources needed as the project is fleshed out

For all these reasons, even though such scientific collaborations do occur they are not very common. Yet, they are needed in order to address major engineering and science challenges in our future (e.g., [9].)

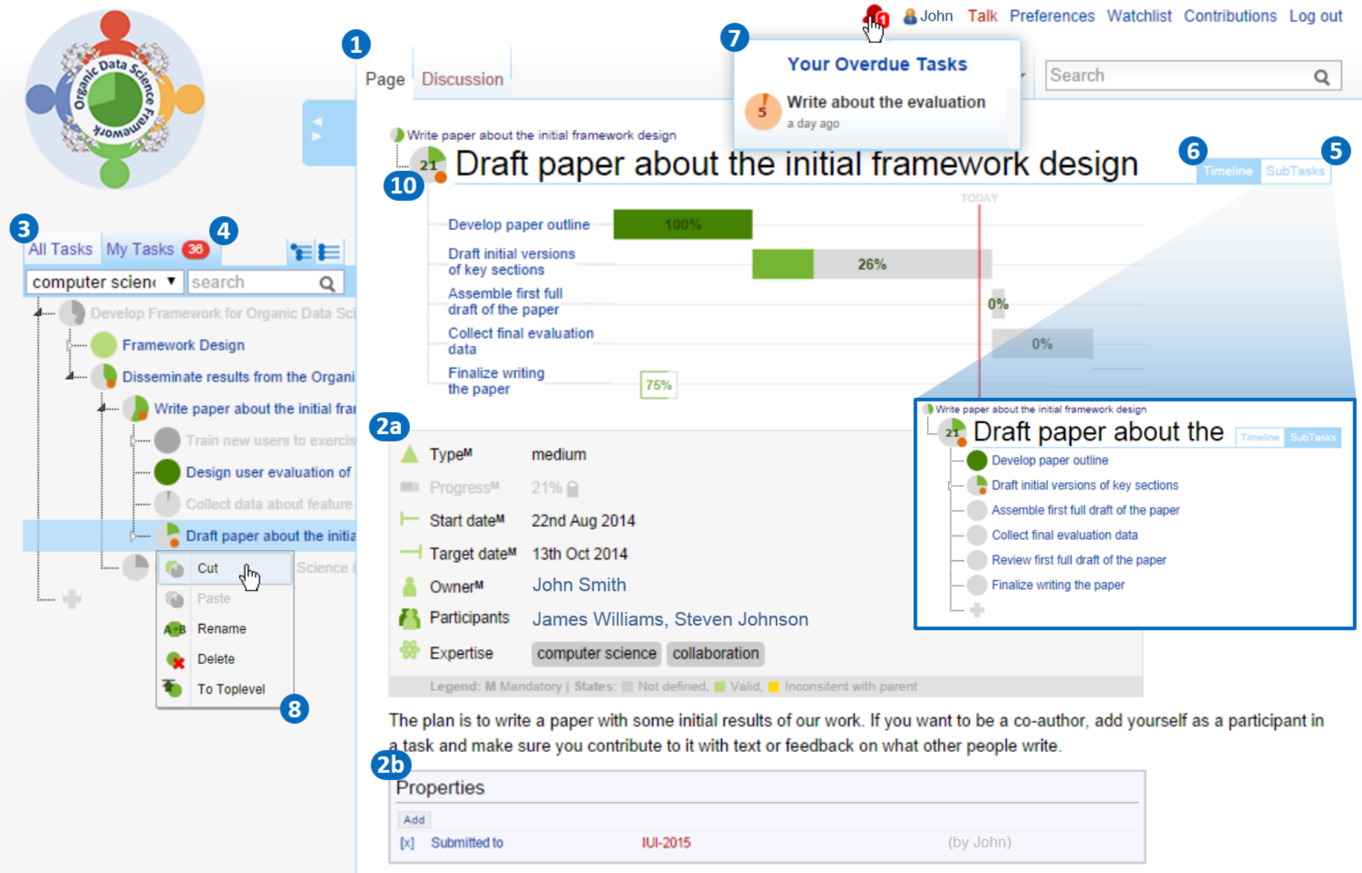
This paper presents an Organic Data Science framework to support scientific collaborations that revolve around complex science questions that require significant coordination to synthesize multi-disciplinary findings, enticing contributors to remain engaged for extended periods of time, and continuous growth to accommodate new contributors as needed as the work evolves over time.

THE ORGANIC DATA SCIENCE FRAMEWORK

We are developing an Organic Data Science framework to support task-oriented self-organizing on-line communities for open scientific collaboration. Our approach is implemented in the Organic Data Science Wiki (ODSW), which is an extension of Semantic MediaWiki [8]. Our overall goal is to reduce the coordination effort required and to lower the barriers to growing the community. Its key features are:

1. Self-Organization through User-Driven Dynamic Task Decomposition

Our system allows users to create tasks, describe them, and decompose them into smaller subtasks. Any user can do any of those actions on any task, whether they created it themselves or not. 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. Figure 1 shows a snapshot of the user interface illustrating how tasks are represented. A summary of all personal allocated tasks is provided on the person page illustrated in Figure 3. All tasks are structured along the time dimension, currently active tasks, furture tasks and completed tasks. User specific expertise is listed above the tasks, hovering over a certain expertise value fades out all tasks that are not associated with that expertise. Different features of the user interface are described in Figure 2.


1
2a
2b
3
4
5
6
7
8
10
Organic Data Science Framework
John Talk Preferences Watchlist Contributions Log out
Page Discussion
Search
Your Overdue Tasks
Write about the evaluation
a day ago
Write paper about the initial framework design
Draft paper about the initial framework design
Timeline SubTasks
TODAY
Develop paper outline 100%
Draft initial versions of key sections 26%
Assemble first full draft of the paper 0%
Collect final evaluation data 0%
Finalize writing the paper 75%
All Tasks My Tasks 36
computer scien
search
Develop Framework for Organic Data Sci
Framework Design
Disseminate results from the Organi
Write paper about the initial fra
Train new users to exercis
Design user evaluation of
Collect data about feature
Draft paper about the initia
Science i
Cut
Paste
Rename
Delete
To Toplevel
Write paper about the initial framework design
Draft paper about the
Timeline SubTasks
Develop paper outline
Draft initial versions of key sections
Assemble first full draft of the paper
Collect final evaluation data
Review first full draft of the paper
Finalize writing the paper

Type^M	medium
Progress^M	21%
Start date^M	22nd Aug 2014
Target date^M	13th Oct 2014
Owner^M	John Smith
Participants	James Williams, Steven Johnson
Expertise	computer science collaboration

Legend: M Mandatory | States: ▨ Not defined, 🟩 Valid, 🟨 Inconsitent with parent

The plan is to write a paper with some initial results of our work. If you want to be a co-author, add yourself as a participant in a task and make sure you contribute to it with text or feedback on what other people write.

Properties

Add

[x]	Submitted to	IUI-2015	(by John)

Figure 1. Organic Data Science Task Page.

⓿ **Welcome Page:** Describes clearly the science and technical project objectives, summarizes currently active tasks, and shows lead contributions (not shown).

❶ **Task Represntation:** Tasks have a unique identifier (URL), and are organized in a hierarchical subtask decomposition structure.

❷ **Task Metadata:** a) Describes clearly the science and technical project objectives summarizes currently active tasks, and shows lead contributions. We distinguish between required metadata that is needed to progress a task and optional metadata. b) Optional user structured properties.

❸ **Task Navigation:** Tasks can expand until a leaf task is reached. Additionally users can search for task titles and apply an expertise filter.

❹ **Personal Worklis:t** The worklist contains the subset of tasks from the task navigation for which the user is owner or a participant. A red counter indicates the current number of tasks in the users worklist.

❺ **Subtask Navigation:** Subtasks of the currently opened task are presented. Filter and search options are not provided in this navigation.

❻ **Timeline Navigation:** All subtasks are represented based on their start, target times, and completion status in a visualization based on a Gantt chart.

❼ **Task Alert:** Signals when a task is not completed and the target date passed. A red counter next to the alert bell indicate the number of overdue tasks.

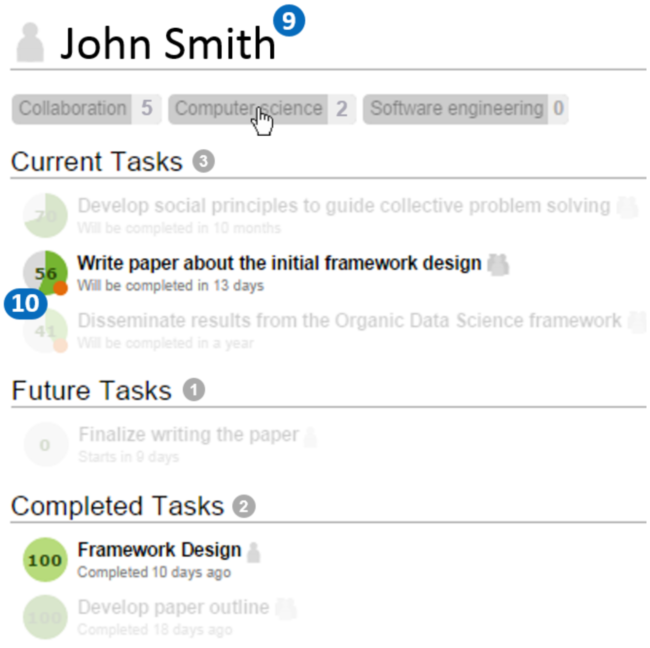
❽ **Task Management:** The interface supports creating, renaming, moving and deleting tasks. For usability reasons, all these actions can be reversed.

❾ **User Tasks and Expertise:** The interface allows users to easily see what others are working on or have done in the past. This creates a transparent process.

❿ **Task State:** Small icons visualize the state of each task intuitively. Tasks with incompleted required metadata are repesented with a cycle and tasks with completed required metadata are represented with a pie chart. The progess is indicated in green.

⓫ **Train New Members:** A separate site is used to train new users in a sandbox environment, where training tasks are explicit. The training is splited into two parts: 1) Users who participate on tasks and 2) User who own tasks (not shown).

Figure 2. Organic Data Science Core Features.


John Smith

Collaboration 5 | Computer science 2 | Software engineering 0

Current Tasks 3

70 Develop social principles to guide collective problem solving
Will be completed in 10 months

56 Write paper about the initial framework design
Will be completed in 13 days

41 Disseminate results from the Organic Data Science framework
Will be completed in a year

Future Tasks 1

0 Finalize writing the paper
Starts in 9 days

Completed Tasks 2

100 Framework Design
Completed 10 days ago

100 Develop paper outline
Completed 18 days ago

Figure 3. Organic Data Science Person Page.

2. Sustainable On-Line Communities through Best Practices

Our approach is to form and sustain communities around science goals, not simple collaborations. Numerous studies about successful on-line communities provide useful design principles for our framework [7], notably on Wikipedia. Our work builds on the social design principles uncovered by this research. Figure 4 gives an overview of the social principles used for this research, and how they map to the user interface features.

3. Opening Science Process

Our approach is to make the collaborative science processes explicit, so that everyone can examine the status of the collaboration and access the rationale of the current activities being pursued. These collaborative processes may be explicitly articulated but are never captured. For this, we find inspiration in the Polymath project, set up to collaboratively develop proofs for mathematical theorems [10], where professional mathematicians collaborate with volunteers that range from high-school teachers to engineers to solve mathematics conjectures. It uses common Web infrastructure for collaboration, interlinking public blogs for publishing problems and associated discussion threads with wiki pages that are used for write-ups of basic definitions, proof steps, and overall final publication. Another project that has exposed best practices of a large collaboration is ENCODE [2]. In ENCODE, as in many other science projects, specific tasks are carved out and assigned to smaller groups in the collaboration. Figure 4 highlights these best practices and lessons learned in items E and F, and indicates how they map to user interface features.

A. Starting communities

A1: Carve a niche of interest, scoped in terms of topics, members, activities, and purpose ⓿
A2: Relate to competing sites, integrate content ⓿
A3: Organize content, people, and activities into subspaces once there is enough activity ⓿❶❽
A4: Highlight more active tasks ❶❷❹
A5: Inactive tasks should have ”expected active times” ❷❻
A6: Create mechanisms to match people to activities ❶❷

B. Encouraging contributions through motivation

B1: Make it easy to see and track needed contributions ❶ - ❼❾❿
B2: Ask specific people on tasks of interest to them ❷
B3: Simple tasks with challenging goals are easier to comply with ❶❽
B4: Specify deadlines for tasks, while leaving people in control ❷-❹❼
B5: Give frequent feedback specific to the goals ❷❹❺❻❾❿
B6: Requests coming from leaders lead to more contributions ❷
B7: Stress benefits of contribution ⓿
B8: Give (small, intangible) rewards tied to performance (not just for signing up) ❾
B9: Publicize that others have complied with requests ❹
B10: People are more willing to contribute: 1) when group is small, 2) when committed to the group, 3) when their contributions are unique ❶❸❹❽❾

C. Encouraging commitment

C1: Cluster members to help them identify with the community ❷❸❾
C2: Give subgroups a name and a tagline ❶❷❸
C3: Put subgroups in the context of a larger group ❶❸❹
C4: Make community goals and purpose explicit ❶
C5: Interdependent tasks increase commitment a. reduce conflict ❶-❸❾

D. Dealing with newcomers

D1: Members recruiting colleagues is most effective ⓿
D2: Appoint people responsible for immediate friendly interactions ⓫
D3: Introducing newcomers to members increases interactions ⓫
D4: Entry barriers for newcomers help screen for commitment ⓫
D5: When small, acknowledge each new member ⓿
D6: Advertise members particularly community leaders, include pictures ⓿
D7: Provide concrete incentives to early members ⓿
D8: Design common learning experiences for newcomers ⓫
D9: Design clear sequence of stages to newcomers ⓫
D10: Newcomers go through experiences to learn community rules ⓫
D11: Provide sandboxes for newcomers while they are learning ⓫
D12: Progressive access controls reduce harm while learning ⓫

E. Best practices from Polymath

E1: Permanent URLs for posts and comments, so others can refer to them ❶
E2: Appoint a volunteer to summarize periodically ⓿
E3: Appoint a volunteer to answer questions from newcomers ⓫
E4: Low barrier of entry: make it VERY easy to comment ⓫
E5: Advance notice of tasks that are anticipated ❷❻❿
E6: Keep few tasks active at any given time, helps focus ⓿

F. Lessons learned from ENCODE

F1: Spine of leadership, including a few leading scientists and 1-2 operational project managers, that resolves complex scientific and social problems and has transparent decision making ⓿
F2: Written and publicly accessible rules to transfer work between groups, to assign credit when papers are published, to present the work ⓿
F3: Quality inspection w. visibility into intermediate steps ❶-❸❺❻❽❿
F4: Export of data and results, integration with existing standards ⓿

Figure 4. Selected social principles from [7] for building successful online communities and selected best practices from Polymath [10] and ENCODE [2].

EVALUATION

The current prototype has been evaluated with data of a scientific collaboration using the site for 10 weeks. Firgure 5 shows two heat maps of Organic Data Science task pages. The heat maps illustrate the proportion of user clicks in particular areas of a page. Here, they show most activity in the areas where the interface features for describing, finding, and managing tasks.

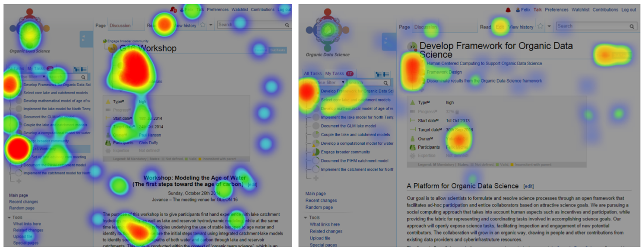

Workshop
Workshop: Modeling the Age of Water
Develop Framework for Organic Data Science
A Platform for Organic Data Science

Figure 5. Heat maps for task pages showing user clicks.

Additionally all user activity is tracked and mapped to features. Figure 6 shows the usage distribution of the new feature that are used to find tasks. The features that improve the task based navigation such as the *Task Navigation* and *Subtask Navigation* are most used. We believe that the other features such as the *Task Alert*, *Timeline Navigation*, and *User Tasks and Expertise* will be used more as the number of tasks in the system increase over time..

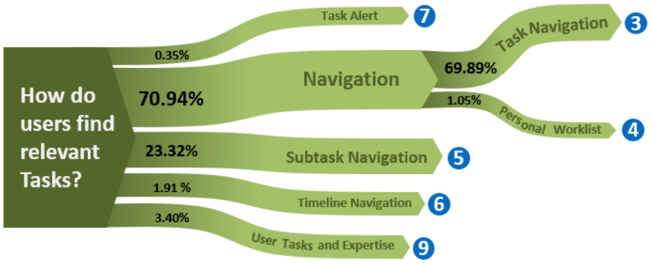


Figure 6. Finding tasks via feature.

RELATED WORK

Intelligent users interfaces to coordinate work has been addressed in prior research, from the underlying formal theories (e.g., SharedPlans [5]) to practical implementations of those theories [11, 12]. The work has focused either on human-computer dialogue or multi-agent coordination. In our case, the coordination is among humans. A promising area of future work is to investigate if these collaboration theories and frameworks could be incorporated into the design of our multi-human collaboration interface.

Some task-oriented collaboration systems have been developed for information seeking tasks [3]. Our goal is to support tasks that have interrelated subtasks and that involve collaboration among peers.

Task-oriented interfaces have been developed for scientific computing, where data analysis tasks are cast as workflows whose validation and execution are managed by the system [1, 4]. In our framework, tasks can be decomposed into more and more specific and well-defined tasks that can later be turned into workflows that can be executed for data analysis.

A wide range of users interfaces have been [6]), and workflow repositories [13]. However, their adoption remains limited. In contrast, popular collaborative Web frameworks are widely used in science, including code repositories, blogs, and wikis. Our approach shares some important features with these tools in tracking tasks, but is more focused on interrelated tasks.

REFERENCES

1. Chin, Jr., G., Leung, L. R., Schuchardt, K., and Gracio, D. New paradigms in problem solving environments for scientific computing. In *Proceedings of the 7th International Conference on Intelligent User Interfaces*, IUI '02, ACM (New York, NY, USA, 2002), 39–46.
2. ENCODE. Special issue on the encode project. vol. 489. Nature, September 2012.
3. Filho, F. F., Olson, G. M., and de Geus, P. L. Kolline: A task-oriented system for collaborative information seeking. In *Proceedings of the 28th ACM International Conference on Design of Communication*, SIGDOC '10, ACM (New York, NY, USA, 2010), 89–94.
4. Gil, Y., Ratnakar, V., Kim, J., Moody, J., Deelman, E., Gonzalez-Calero, P., and Groth, P. Wings: Intelligent workflow-based design of computational experiments. *Intelligent Systems, IEEE 26*, 1 (Jan 2011), 62–72.
5. Grosz, B. J., and Sidner, C. Intentions in communications: Plans for discourse. In *Cohen*, Morgan and Pollack, Eds. MIT Press, 1990, 417–444.
6. Huss, J. W., Lindenbaum, P., Martone, M., Roberts, D., Pizarro, A., Valafar, F., Hogenesch, J. B., and Su, A. I. The gene wiki: community intelligence applied to human gene annotation. *Nucleic Acids Research* (2009).
7. Kraut, E., and Resnick, P. *Building Successful Online Communities: Evidence-Based Social Design*. MIT Press, February 2011.
8. Krótzsch, M., and Vrandecic, D. *Semantic MediaWiki. Foundations for the Web of Information and Services*. Springer, 2011, 311–326.
9. National Academy of Sciences. *Grand Challenges for Engineering*. National Academy of Engineering, Boston MA, 2008.
10. Nielsen, M. *Reinventing Discovery: the new era of networked science*. Princeton University Press, 2011.
11. Rich, C., Sidner, C., Lesh, N., Garland, A., Booth, S., and Chimani, M. Diamondhelp: A collaborative task guidance framework for complex devices. In *AAAI*, vol. 20 (2005), 1700.
12. Rich, C., Sidner, C. L., and Lesh, N. Collagen: Applying collaborative discourse theory to human-computer interaction.(articles). *AI magazine 22*, 4 (2001).
13. Roure, D. D., Goble, C., and Stevens, R. The design and realisation of the virtual research environment for social sharing of workflows. *Future Generation Computer Systems 25*, 5 (2009), 561 – 567.