Organic Data Science:
Towards Task-Oriented Self-Organizing On-Line Communities for Open Collaboration in Science

Anonymized Authors
Affiliation
Address
e-mail address
Optional phone number

Anonymized Authors
Affiliation
Address
e-mail address
Optional phone number

Anonymized Authors
Affiliation
Address
e-mail address
Optional phone number

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 Organic Data Science framework to address these challenges with a collaborative user interface that supports: 1) *self-organization of the community* through user-driven dynamic task decomposition, 2) *on-line community support* by incorporating social design principles and best practices, 3) *an open science process* by capturing new kinds of metadata about the collaboration that provide invaluable context to newcomers. We present preliminary results with a pilot science project focused on the isotopic "age" of water in a ecosystem in a collaboration that spans hydrology, limnology, and environmental sciences.

Author Keywords
Collaboration interfaces, Organic Data Science.

ACM Classification Keywords
H.5.3. Information interfaces and presentation (e.g., HCI): Group and organization interfaces.

INTRODUCTION
Over the last hundred years, science has become an increasingly collaborative endeavor. Scientific collaborations, sometimes referred to as "collaboratories" and "virtual organizations", range from those that work closely together and others that are more loosely coordinated [Ribes and Finholt 2009; Bos et al. 2007]. Some scientific collaborations revolve around sharing instruments (e.g., the Large Hadron Collider), others focus on a shared database (e.g., the Sloan Sky Digital Survey), others form around a shared software base (e.g., SciPy), and others around a 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., [NAE 2014].)

This paper presents the **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. Our organic data science framework addresses these challenges with a collaborative user interface that supports: 1) *self-organization of the community* through user-driven dynamic task decomposition, 2) *on-line community support* by incorporating social design principles and best practices, 3) *an open science process* by capturing new kinds of metadata about the collaboration that provide invaluable context to newcomers. Users formulate science tasks to describe the what, who, when, and how of the smaller activities pursued within the collaboration. The interface is designed to entice contributors to participate and continue involved in the specific tasks they are interested in. The framework is in its early stages of development, and it evolves to accommodate user feedback and to incorporate new collaboration features.

The paper begins with a motivating scenario of a complex science task that we are pursuing using this framework. We then introduce our approach to support task-oriented self-organizing communities for open scientific collaboration. We present our implemented framework, and a preliminary evaluation with user data collected to date.

MOTIVATING SCENARIO

The recognition that the future health of the world depends on provisioning of ecosystem services provided by fresh waters, including quantity and quality available for consumption, agriculture and aquaculture, industry, recreation, and carbon sequestration, has motivated an array of research and advocacy initiatives [MEA 2005; ILEC 2007; Levin and Clark 2009]. The necessary expertise is represented in multiple disciplines, including hydrology [NRC: 2011, 2012], ecology [Foley et al. 2011], economics and security [Suweis et al. 2013]. Ongoing research programs dedicated to water are very fragmented, these include the Critical Zone Observatories (CZOs), the Global Lake Ecological Observatory Network (GLEON), the Consortium of Universities for the Advancement of Hydrologic Science (CUAHSI), and the Long Term Ecological Research (LTER). Many federal organizations have substantial efforts, such as the U.S. Geological Survey's Water Mission and the NASA Water Resources program. Despite great scientific advances and cross-connections among these collaborations, scientists still are challenged to quantify basic water and material fluxes that underpin aquatic ecosystems. We illustrate the challenges faced by collaborative science projects with a particular research agenda that we are supporting with our work.

The scientific research that we are pursuing focuses on theoretical and experimental aspects of the isotopic "age" of water in watershed-lake systems. In this context, "age" is defined as the time since the water entered the system as precipitation. Our hypothesis is that the watershed-lake "isoscape" provides an experimental basis for predicting flow paths, residence times and the relative age of water, and that understanding these spatiotemporal patterns provides a deeper understanding of fundamental biogeochemical processes (including carbon cycling) within the lake catchment system.

A great challenge is the amount of effort required to organize and coordinate the diversity of people, computational models, data, and other resources. Both the hydrology and limnology communities have developed methods for measuring isotope ratios of carbon, oxygen, and hydrogen, but their analyses are framed by very different science questions. In addition, LTER, CZO, GLEON, and CUAHSI all offer different kinds of data about different locations and areas of study. Achieving the science goals requires organizing people to reach agreements for the data and model standards necessary for interoperability and implement them. Although the work that people will do in the project is likely synergistic with their personal research agendas, a significant deterrent for participation is how much they will be taxed by the coordination overhead and how productivity diminishes in inverse proportion to the size and diversity of a group. In our experience, scientists are more willing to invest into small subgroups that are charged with manageable tasks that have clear outcomes of personal interest. Therefore, the challenge becomes supporting such an organization of work in an efficient manner.

Another important challenge is that the growth of the collaboration would take years to reach critical mass. Even if an initial group successfully collaborates to model the age of water for a few locations, the science results will be much stronger if more scientists join the collaboration and contribute models for the locations they typically study. Much of environmental science today goes through great pains to get results with an initial limited scope, but miss on the opportunity for broader science and stronger results. Many scientists could contribute, but the burden of figuring out how to run the proposed models in their site is too large. Many recent studies on reproducibility attest to the effort required to reuse models described in scientific publications [Begley and Ellis 2012; Garijo et al. 2012]. Our experience suggests that scientists are more willing to undertake this kind of work if there is a community around it ready to help them when difficulties or interesting questions arise. In our view, it is not the existence of the scientific problem itself but the existence of a functioning *community* around the problem that will attract new contributors.

Finally, a significant impediment to engaging in this kind of research is the effort required to be aware of the status of the collaboration. For newcomers, it is challenging to come up to speed on what has been done when all the necessary context is spread over personal emails and meeting notes. For others who are already contributing, it is hard to catch up after a week or a month of being away (e.g. due to travel or other activities. We believe that an on-line system that makes the collaborative activities more visible will be more approachable for outsiders.

APPROACH

We are developing an Organic Data Science framework to support task-oriented self-organizing on-line communities for open scientific collaboration. Its key features are:

1. **Self-organization of the work**, through an interface that supports scientists to organize joint tasks and to easily track where they can contribute and when
2. **Sustainable on-line communities**, through an interface that incorporates principles from social sciences research on successful on-line collaborations, including best practices for retention and growth of the community
3. **Open science processes that expose all tasks and activities publicly**, through an interface that captures new kinds of metadata about the collaboration so all participants (especially newcomers) can immediately catch up with the work being done

Our goal is to reduce the coordination effort required and to lower the barriers to growing the community.

Self-Organization

Our approach is to use tasks as an organizational mechanism for coordination, and to allow users to create joint tasks, decompose them into smaller subtasks, and easily track their status. Tasks can be seen as a shared tool for social cognition [Hutchins 1995], which considers that in collaborative settings the expertise is not only in the minds of individuals but in the organization of the tools and objects that they share. Processes and tasks have been shown to be a key to collaboration in science laboratories [Chandrasekaran and Nernessian 2015], to coordinate work in multi-agent systems [Grosz and Sidner 1990], and to the productivity of knowledge workers in an organization [Davenport 2013].

Decomposition of subtasks is an important aspect of describing tasks. Many explanations of procedures, including scientific and technical expositions, exhibit goal-oriented hierarchical structure [Britt and Larson 03]. The temporal relations among subtasks are also important [Pietras and Coury 94]. The user interface should be designed so users have some initial structure to express tasks. [Van Merrienboer 97] proposes the use of process worksheets to guide students through complex tasks. [Mahling and Croft 88] also found that the formulation of tasks is greatly improved through form-based interfaces.

Sustainable On-Line Communities

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 [Kraut and Resnick 2011], with topics as varied as the design of the editorial process [Spinellis and Louridas 2008], community composition and activities [Gil and Ratnakar 2013], incentives to contributors [Mao et al. 2013; Leskovec et al. 2010], critical mass of contributors [Raban et al. 2010], coordination [Kittur et al. 2009], group composition [Lam et al. 2010], conflict [Kittur et al. 2010], trust [McGuinness et al. 2006], and user interaction design [Hoffman et al. 2009].

Figure 1 summarizes the social principles that we are using in our approach. We follow the organization used in [Kraut and Resnick 2011], focusing in this paper on social principles that are relevant to early stages of the community. In the next section we explain how they map to features in our user interface.

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. [Polanyi 1983] coined the terms and discussed differences between tacit and explicit knowledge of individuals in organizations. While explicit knowledge can be communicated in formal languages that can be processed by other individuals, people have tacit knowledge that they cannot explicitly express. In their theory on organizational knowledge creation, Nonaka and Takeuchi described the transformation modes between tacit and explicit knowledge with socialization, externalization, internalization, and combination [Takeuchi and Nonaka 2004; Nonaka and Takeuchi 1995]. In our project, we aim at externalizing the tacit knowledge of scientists about the collaboration itself.

We find inspiration in the Polymath project, set up to collaboratively develop proofs for mathematical theorems [Nielsen 2011; Gowers 2009a], 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 [Nielsen 2013] with wiki pages that are used for write-ups of basic definitions, proof steps, and overall final publication [Gowers 2013]. Interactions among contributors to share tasks and discuss ideas are regulated by a simple set of social norms for the collaboration [Gowers 2009b]. The growth of the community is driven by the tasks that are posted, as tasks are decomposed into small enough chunks that contributors can take on.

Another project that has exposed best practices of a large collaboration is ENCODE [Birney 2012; Nature 2012]. In ENCODE, the tasks that are carved out for each group in the collaboration are formally assigned since there is funding allocated to the tasks. In addition the collaboration members are selected beforehand. Despite these differences with our project, we share the explicit assignment of tasks in service of science goals.

Figure 2 outlines the best practices and lessons learned from these two projects that are applicable to our work.

A. Starting communities

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

B. Encouraging contributions through motivation

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

C. Encouraging commitment

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

D. Dealing with newcomers

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

Figure 1. Selected social principles from [Kraut and Resnick 2011] for building successful online communities that can be applied to Organic Data Science. We focus on social principles that are relevant to early stages of the community, and leave out more advanced principles (e.g., for retention of members and for regulating behavior).

E. Best practices from Polymath

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

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
- F2. Written and publicly accessible rules to transfer work between groups, to assign credit when papers are published, to present the work
- F3. Quality inspection with visibility into intermediate steps
- F4. Export of data and results, integration with existing standards

Figure 2. Selected best practices from the Polymath [Nielsen 2011] project and lessons learned from ENCODE [Nature 2012] that can be applied to the initial design of our Organic Data Science framework.

THE ORGANIC DATA SCIENCE WIKI

In this section we describe the Organic Data Science Wiki (ODSW), our current implementation of the Organic Data Science framework. It is built as an extension of the Semantic Media Wiki platform [Krötzsch et al. 2011; Bry et al. 2012], and uses its semantic capabilities to structure the content of the site, including task properties and user properties. The semantic wiki provides an intuitive user interface that hides from users any formal semantic notation [Gil 2013; Bry et al. 2012]. We highlight here major features of the interface that implement our approach. These features are summarized in Table 1.

Self-Organization through User-Driven Dynamic Task Decomposition

ODSW 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 3 shows a screenshot of a task page, with the features highlighted in blue circles. Task pages follow a pre-defined structure that is automatically presented to the user when a new task is created (❶). Subtasks can be added to form a hierarchical task structure (❸,❺). Users are asked to specify metadata (2a), which are major properties of the task such as start and target dates. These metadata properties enable the ODSW to assist users to manage tasks by generating timelines (❻), a task state summarizing subtask progress (❿), and alerts when they are late (❼). Users can sign up for a task either as "owners", which makes them responsible for the task getting done, or "participants", which means they will contribute to the task.

We distinguish between several categories of metadata. *Pre-defined metadata* are properties of tasks that ODSW will use to assist users to manage tasks (❹,❼,❾,❿). Pre-defined metadata can be *required* or *optional*. Required metadata includes the start date, target date, task owner, and task type. Tasks whose required metadata is incomplete have special status in ODSW and are highlighted differently in the interface to alert users of their missing metadata. Optional task metadata includes the task participants and the task expertise indicating the kind of background or knowledge required to participate in the task. *Dynamically-defined metadata* (2b) allow users to create new properties on the fly that help group tasks with domain-specific features, for example tasks that are related to calibration of models or outreach tasks.

An important required metadata property is the task type (high, medium, and low level). The progress to date for low-level tasks is provided manually by their owners or participants, since the tasks have small duration. ODSW calculates the progress of higher-level tasks based on their subtasks and their start as well as target dates. The progress of a medium-level task is calculated as an average of the progress of its subtasks. For high-level tasks, we assume a linear progress based on the start and target date in relation to today's date. High-level task are colored in lighter green and lower-level tasks in darker green.

All task metadata is stored in the wiki as semantic properties of the task page. We use the query facilities of Semantic MediaWiki to create dynamic content of pages based on those properties. The query is embedded in the page and ODSW generates the page automatically [Krötzsch et al. 2011]. For example, special pages with dynamic content list all tasks that require specific kinds of expertise, or involve particular models or kinds of data.

Everything else in the task page is text and media that describes the actual work involved in doing a task as well as the results when it is completed.

Sustainable On-Line Communities through Best Practices

The user interface of ODSW is designed to support the formation of an on-line community and its growth. We follow the successful social design principles (see Figure 1) and combine it with best practices from projects such as Polymath and ENCODE (see Figure 2). We highlight here how the interface is designed to address some of these social principles.

For example, several social principles (A1-A6) address the formation of the community. They are most noticeable in the main page of the site. It describes clearly the science and technical objectives of the project, displays a summary of currently active tasks, and shows the leadership and major contributors (⓿). In geosciences, the models used in the project are important to anchor the work for newcomers, so they are also shown in the main page. ODSW automatically generates the model and contributor tables from the current contents with a semantic wiki query, so they are always up to date. Those tables highlight properties of note, which allow newcomers to match ongoing work to their personal interests.

Dealing with newcomers is another important aspect of creating an on-line community. Social principles D1-D12 address this. We set up a separate site[1] to train new users. This training site also uses ODSW. A new user is given a set of predefined training tasks, shown in Figure 4, each for learning and practicing a different feature of the interface. Training tasks (⓫) follow the structure of documentation pages, and allow new users to practice by using the same interface as they will use in the main site. As they complete the tasks, users can see the task status (❿) changing. The training is divided in two phases. The first phase trains them to contribute to existing tasks. The second phase trains

[1] Anonymized link to webpage.

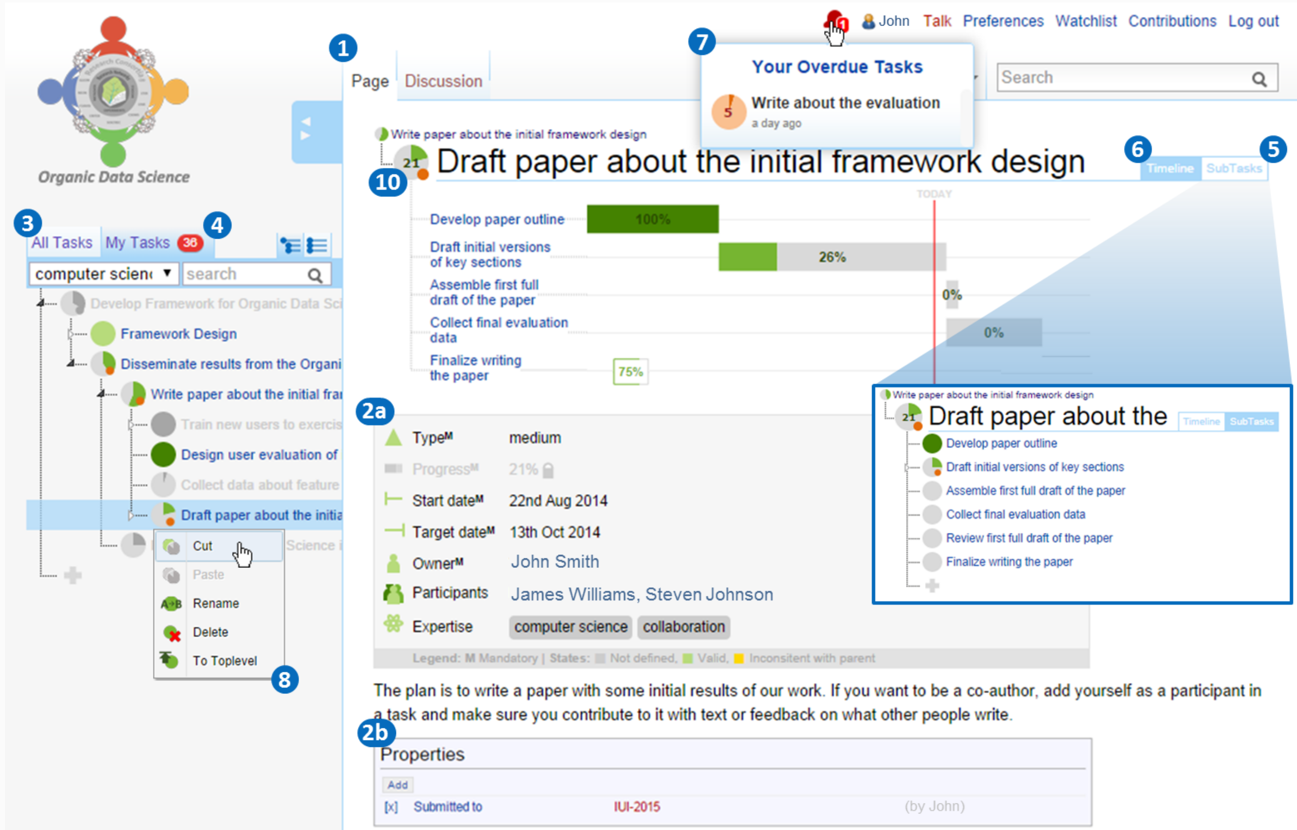

Figure 3: Organic Data Science Task Page.

Table 1: User interface features in Organic Data Science. Each feature is annotated with a description, and lists the objectives it serves: self-organizing the work (I), sustaining the on-line community (II), and opening the science process (III).

Feature	*Feature Description*	*Objectives* *I*	*II*	*III*	*Social Principles*
⓿ **Welcome Page**	Describes clearly the science and technical project objectives summarizes currently active tasks, and shows lead contributions.		✓	✓	A1, A2, A3, B7, D1, D5, D6, D7, E2, E6, F1, F2, E4
❶ **Task Representation**	Tasks have a unique identifier (URL), and are organized in a hierarchical subtask decomposition structure.	✓		✓	A3, A4, A6, B1, B3, B10, C2, C3, C4, C5, E1, E5, F3
❷ **Task Metadata**	a) Task metadata are properties of the task, such as start date and target date. b) User structured properties. All metadata is stored as semantic properties.	✓	✓	✓	A4, A5, A6, B1, B2, B4, B5, B6, C1, C2, C5, F3
❸ **Task Navigation**	Tasks can expand until a leaf task is reached. Additionally users can search for task titles and apply an expertise filter.	✓		✓	B1, B4, B10, C1, C2, C3, C4, C5, F3
❹ **Personal Worklist**	The worklist contains the subset of tasks from the task navigation for which the user is owner or a participant.	✓	✓	✓	A4, B1, B4, C3
❺ **Subtask Navigation**	Subtasks of the currently opened task are presented.	✓		✓	B1, B5, B9, B10 C5, F3
❻ **Timeline Navigation**	All subtasks are represented based on their start, target times, and completion status in a visualization based on a Gantt chart.	✓		✓	A4, A5, B1, B5, E5, F3
❼ **Task Alert**	Signals when a task is not completed and the target date passed	✓	✓		B1, B4
❽ **Task Management**	The interface supports creating, renaming, moving and deleting tasks. For usability reasons, all these actions can be reversed.	✓			A3, B3, B10, F3
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.	✓	✓	✓	B1, B2, B5, B8, B10, C1, C5
❿ **Task State**	Small icons visualize the state of each task intuitively.	✓	✓	✓	B1, B5, E5
⓫ **Training New Members**	A separate site is used to train new users in a sandbox environment, where training tasks are explicit.		✓	✓	D2, D3, D4, D8, D9, D10, D11, D12, E3, E4

them to create new tasks and to manage them as owners. One person in the collaboration is always assigned to help new users with their training, and is available by email to answer questions.

Opening Science Processes through Explicit Metadata Capture

ODSW creates a transparent work process. Anyone can see the contents of the site, the process being followed by the whole community, and the tasks being undertaken by different subgroups are open and accessible. In order to edit the contents, users have to become contributors by getting a login and undergoing training (⓫).

Decomposing complex tasks into smaller manageable tasks also makes the science process more transparent. The ODSW interface allows to drill down into subtasks or drill up to the more general parent task (❸,❹,❺,❻,❽). Users who own small tasks can see the context and importance of their tasks.

Defining explicit task metadata (❷) such as a task type, progress, owner, participants or start and target date helps to make the process more transparent for other users. The interface exploits this metadata to help users find what is relevant to them. For example, ODSW groups the tasks for which the user is owner or a participant and forms personal task lists (❹,❾). This allows users to easily see what others are planning to work on or have worked on in the past. Another example is that hovering over a certain expertise value (e.g., "nutrients") fades out all tasks in the page that are not associated with that expertise. This helps new contributors find out relevant activities through the tasks that their colleagues are involved with.

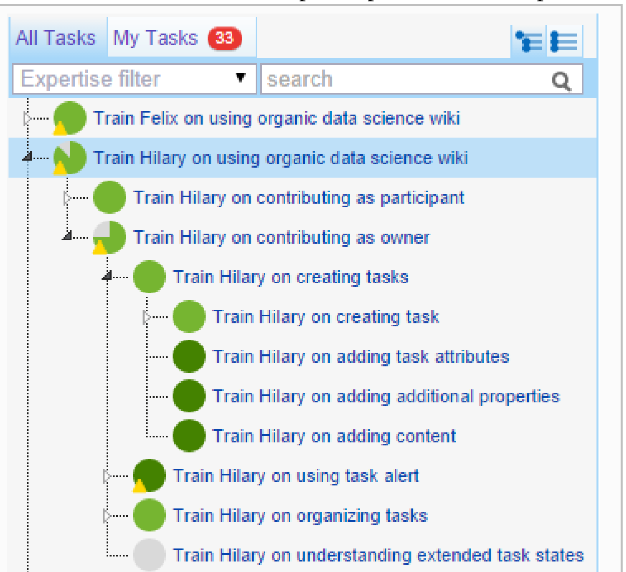

Figure 4: Tasks in the training site for a new member.

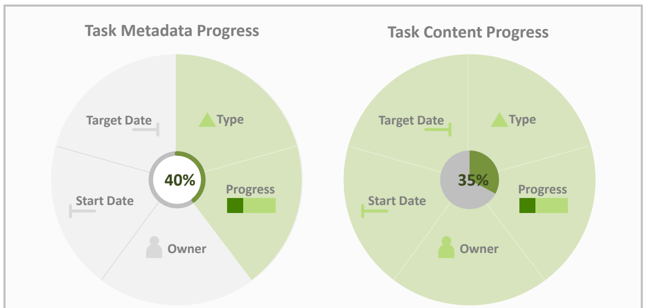

Figure 5: Conceptual task state estimation.

ODSW aggregates information from the required metadata properties and automatically generates visual task states as colored pie chart task icons (❿). Figure 5 illustrates how task states are generated from metadata fields. The left side illustrates the visualization when the required metadata is incomplete, and the right side when the user has provided all required metadata.

EVALUATION

We present an evaluation of our current implementation of the Organic Data Science framework. The site has been active since January 2014 and has been in use while new features were rolled out. The lead users have live discussions at biweekly telecons to discuss the design of the framework and the overall progress of the work, with all the resulting tasks captured in the wiki.

We instrumented the system and started to collect data once all the features described above were rolled out. In the first 10 weeks we collected around 19,000 log entries, which we used for the evaluation presented here.

Is the Framework Helping Users Organize their Work?

We analyzed how the framework helps users manage tasks.

Are users creating and using tasks? The site contains 122 tasks. In the 10-week time period all task pages together were accessed more than 2,900 times. Person pages were accessed 328 times in total. The tasks in the current ODSW site include: 1) tasks about the science of the age of water, 2) tasks about the development of the Organic Data Science approach and its implementation, 3) tasks about outreach such as an upcoming workshop about ODSW at the annual GLEON meeting. We organized and wrote this paper collaboratively using ODSW:

How do users find relevant tasks? Figure 6 shows what features are used by users to open task pages. Most users used the Task Navigation feature to find task pages. A probable explanation for this is the feature gives users an overview over all tasks, drill down quickly, and apply specific filters. The Task Alert feature was not used very often, but we expect that this feature will be more important as the group faces deadlines (such as the writing of this paper, an upcoming scientific workshop, etc.).

What features are used to manage tasks? Figure 7 shows heat maps for two task pages that illustrate in red where users click most. Every heat map represents the clicks on one single page. Most clicks occur in areas where many of our task-oriented interface features are located. *Are users* *creating subtask hierarchies?* Figure 8 shows data about the task hierarchies in terms of the depth (number of ancestors of tasks) and breadth (number of children). There are 13 top-level tasks, and the majority of tasks are at the next three levels of decomposition. As far as breadth, most tasks have no subtasks, and are either tasks small enough that they do not require further decomposition or tasks that will take place in the future and have not yet been fleshed out. Many tasks do have several subtasks.

Is the Framework Helping to Create Communities? We analyzed the logs to determine how many users were connecting in some way through the tasks in the site. We removed tasks with no participants, since they were created recently and did not even have an owner. We did not filter out data for tasks that were renamed or deleted. All results are illustrated in Figure 9.

How many tasks are viewed by more than one person? Figure 9(a) shows that 52% of the tasks are visited by two or more persons. Currently 48% of all task pages are accessed by only one person. This is a high number, but we believe that this is due to the many tasks that are planned but not yet worked on since the project is still in its first year. We expect this percentage to decrease as the project progresses, particularly as it gets closer to completion.

How many tasks have more than one person signed up? Figure 9(b) shows the total persons involved in tasks, including the participants and the owner. 72% of the tasks have two or more persons involved, and 46% have three or more. This is quite a high number of people sharing tasks.

How many tasks have more than one person editing task metadata? Figure 9(c) shows these results. Currently 81% of all tasks have their metadata edited by only one person. This is expected, since typically the task owner adds the initial metadata. But 19% of the tasks have their metadata edited by two or more persons. This indicates that non-owners have an interest in the management of the tasks.

How many tasks have more than one person editing their content? This is shown in Figure 9(d). 11% of the tasks have their content edited by two or more persons. The vast majority of the tasks have their content edited by just one person. This is a very low number, and we hope it will increase as more tasks are worked on and accomplished.

How has the community grown? The site initially had four users, who started to create content and tasks. So far the community has grown by direct referral (per principle D1 of Figure 1). Within 3 months a handful of additional users were brought in to help with specific tasks. In the last few weeks, a few more have been added. The site currently has 18 registered users, which include computer scientists, hydrologists, ecologists, limnologists, and geoinformaticians. At the end of October, a first outreach invitation-only workshop will be held at the GLEON annual meeting which has already 40 registered participants interested in the age of water. All workshop activities are being managed using ODSW by the organizers and by the participants themselves, and we expect some fraction of them to remain involved. We are cautiously expanding the community since ODSW is still under development. So far new users have been added painlessly. New users report that the training tasks take around one hour. Our logs show that they did not access the documentation once training

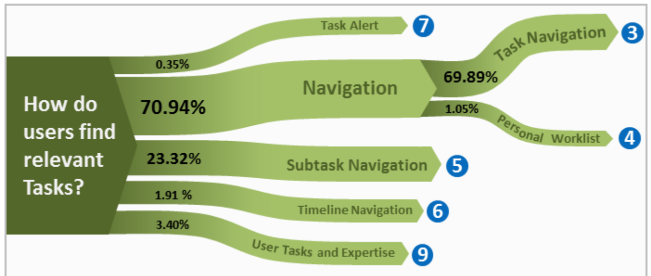

Figure 6: Finding Tasks via Features.

Figure 7: Heat maps for task pages showing user clicks.

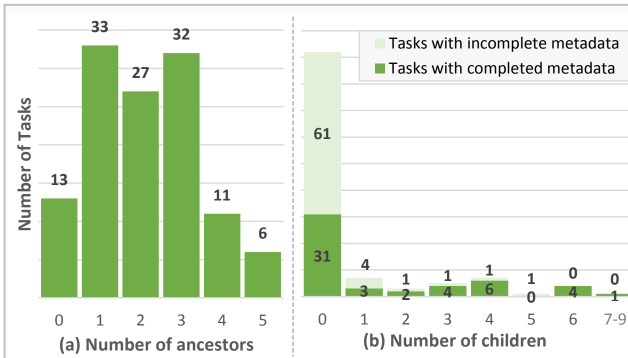

Figure 8: Subtask Hierarchies.

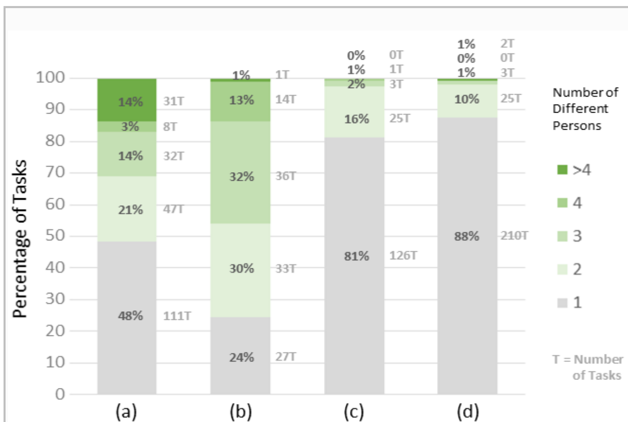

Figure 9: Task Collaboration Evaluation.

was completed. New users are creating tasks and participating in them, and the logs also show they are not undoing any of these actions.

What does the current social network of collaborators look like? We created a network by using task metadata properties about owners and participants in tasks. Users are represented as nodes in the network, and each edge between two nodes represents that the two users are signed up for the same task one or more times. The number of tasks they have in common is expressed by the strength of edges. The result is illustrated in Figure 10. One interesting observation is that there are edges among most of the existing users, indicating collaboration activities across all participants. There are two major connected components in the graph, which are apparent at the top and the bottom of the network, indicating two strong collaboration communities. Users developing the ODSW software are at the bottom, while users working on the age of water are at the top. There are many links across these groups, as both are involved in the design of the overall approach to Organic Data Science.

Is the Framework Helping to Open the Science Processes?

This aspect of our approach is hard to evaluate, particularly since the community is still small. For now, we present initial relevant data.

Is the open science process helping new users understand the status of the collaboration? New users report informally that it is easy to browse the wiki and understand what tasks are currently active, why are they being pursued, who is involved, and what their scope and goals are. In the future, we plan to conduct surveys with users about the utility of the framework to help them understand the status of the collaboration.

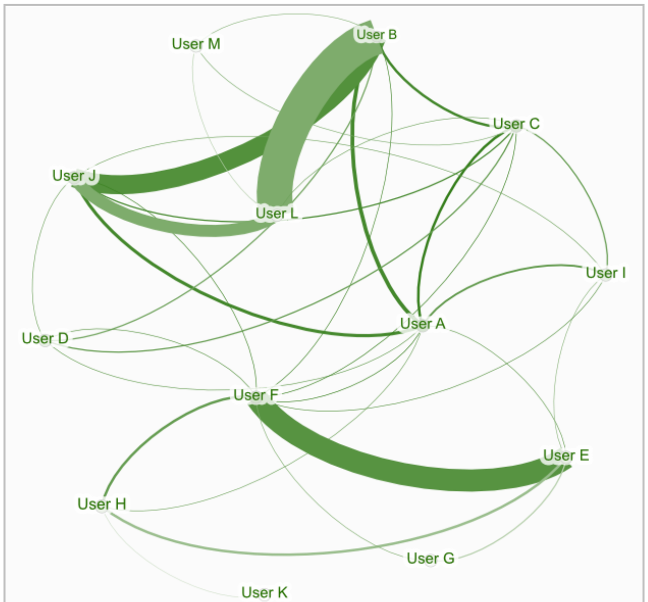

Figure 10: Organic Data Science Collaboration Graph.

How much metadata have contributors added? Since Semantic MediaWiki represents properties using the W3C RDF standard [Brickley and Guha 2014], we tallied the amount of triples of the form <object property value> represented in RDF. We found that so far the amount is a total of 1047 RDF triples. The vast majority is pre-defined metadata properties, as we have not yet emphasized this aspect of ODSW when we train new users. As the site grows in content, we expect that these properties will be most useful in organizing information and exposing other thematic dimensions for tasks, people, and other resources in the site.

RELATED WORK

We discuss related work in scientific collaboration, collaboration systems, and task-centered user interfaces.

Scientific Collaboration

[Bos et al. 2007] did a comprehensive study of scientific collaborations and propose seven types:

1) Shared Instruments, where instruments or sensors are used by a community (e.g., National Ecological Observatory Network [NEON 2014]); *2) Community Data Systems*, where a data resource is maintained and used by a community (e.g., the Protein Data Bank [Berman et al. 2000]), *3) Open Community Contribution Systems*, where tasks are carried out by a community including citizen scientists (e.g., the GalaxyZoo project for labeling galaxy images [Lintott 2010]), *4) Virtual Communities of Practice*, where a community shares interest in specific research topics (e.g., the Global Lake Ecological Observatory Network [GLEON 2014]), *5) Virtual Learning Communities*, where the purpose is to learn through the collaboration (e.g., the VIVO research network [Krafft et al. 2010]), *6) Distributed Research Centers*, where several institutions collaborate in a funded project (e.g., the ENCODE genomics project [Nature 2012], and *7) Community Infrastructure Projects*, where a community shares computing and software infrastructure (e.g., the Community Surface Dynamics Modeling System [Peckham et al. 2013]).

Our work has some of the properties of a distributed research center (6), and is an open community contribution system (3) but without the prescribed tasks typically found there. Organic Data Science can be considered a new type of collaboratory, where tasks are defined on the fly as the project progresses and the collaboration involves unanticipated contributors.

[Ribes and Finholt 2009] analyze the challenges of organizing work in four scientific collaborations: GEON (Geosciences Network), LEAD (Linked Environments for Atmospheric Discovery), WATERS (Water and Environmental Research Systems), and LTER (Long-Term Ecological Research). They found that major challenges for organizing work were: 1) the tension between planned work, with its work breakdown structures with deadlines, versus emergent organization as new requirements and unknowns are uncovered, 2) the tradeoff that participants face between doing basic research and contributing to the technical development in support of the research, and 3) the desire to incorporate innovations while needing a stable framework to do research. Other studies have uncovered similar needs [Steinhardt and Jackson 2014]. Organic Data Science is poised to offer the flexibility of easily incorporating emergent tasks and people, and the enticement to participants through acknowledgement of contributions so that uneven support from particular contributors is properly exposed.

On-Line Collaboration Systems

Despite the wide range of approaches that have been explored for collaboration, they have not had much adoption in science practice. A study on Electronic Lab Notebooks shows the benefits of structuring knowledge in an ad-hoc and simple manner [Oleksik et al. 2014]. Other studies have demonstrated the benefits of using a shared communication board to facilitate collaborative decision making for patient care [Kane et al. 2013]. A study of MathOverflow shows how the quality of answers can be improved collaboratively [Tausczik et al. 2014]. Collaborative user interfaces that have been used in science include semantic wikis (e.g., [Huss et al. 2010]), workflow repositories [De Roure et al. 2009], and argumentation systems (e.g., [Introne et al. 2013]). However, their adoption remains limited. In contrast, popular collaborative Web frameworks are widely used in science, including code repositories, blogs, and wikis. For example, issue tracking tools are popular to coordinate programmer teams, and can be used for managing other kinds of tasks. Our approach shares some important features with these tools in tracking tasks. However, our approach is better positioned to address social issues such as incentives, motivation, and enticing newcomers.

There have been many studies of on-line communities [Kraut and Resnick 2011], notably on Wikipedia. Our work builds on the social design principles uncovered by this research. However, our belief is that scientific work is best organized around tasks, not topic pages. An analysis of Wikipedia shows a continuously increasing readership and a decreasing contribution since 2007, pointing to the need to better coordinate work [Morgan et al. 2014].

Argumentation interfaces facilitate the collaborative synthesis of diverse ideas [Buckingham-Shum 2006], and have been used in the context of science. [Introne et al. 2013] describe the Climate CoLab, a collaborative environment for climate research. It offers argumentation structures, where evidence and hypotheses from different scientists can be compared and integrated to create a common view on climate research. This work, however, does not focus on supporting science research tasks while they are being carried out, only on organizing results of scientific work done elsewhere.

Task-Centered User Interfaces

The coordination of work has been a focus of formal theories (e.g., SharedPlans [Sidner and Grosz 1990]) and practical implementations of those theories [Rich et al. 2005; Rich et al. 2001]. 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 (e.g., Web search). An example is Kolline [Filho et al. 2010], which supports the collaboration between inexperienced users that need help from more advanced users. Our goal is to support tasks that have interrelated subtasks and that involve collaboration among peers.

Other work on managing tasks in on-line environments addresses tasks for remote workers, such as microtasks in Amazon Mechanical Turk [Park et al. 2014; Kamar et al. 2012]. The workers are not explicitly coordinating the work, and the tasks are pre-defined for them and tend to be repetitive across workers.

User tasks are sometimes inferred from their use of the interface [Steichen et al. 2013]. These tasks concern interface use, rather than coordination. Other studies suggest how to design visual feedback to increase the task resumption rate with less stress [Liu et al. 2014].

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 [Chin et al. 2002; Gil et al. 2011]. In our framework, tasks can be decomposed into more and more specific and well-defined tasks that can be turned into workflows that can be executed for data analysis. The interface between our framework and workflows is an area of planned work.

CONCLUSION

We have presented a novel on-line collaboration framework to support organic data science. The main features of this framework are a task-centered organization, the incorporation of social design principles, and the open exposure of scientific processes.

We continue to collect data about the on-line activities of the project. We have specific hypotheses about how the maturity of the project will affect the management of tasks, about how the growth of the community will affect the amount of on-line coordination that occurs, and about the task structure as the scope of the work increases. Future work includes analyzing the evolution of the community in quantitative terms.

REFERENCES

[Auer et al. 06] "OntoWiki - A Tool for Social, Semantic Collaboration." Soren Auer, Sebestian Dietzold and Thomas Riechert. 5th International Semantic Web Conference, 2006.

[Auer et al. 07] "DBpedia: a nucleus for a web of open data." Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, Zachary Ives. Proceedings of the 6th international semantic web conference, 2007.

[Begley and Ellis 2012] "Drug development: Raise standards for preclinical cancer research." C. Glenn Begley and Lee M. Ellis. Nature 483, 531–533, 29 March 2012.

[Berman et al. 2000] "The Protein Data Bank." H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne. Nucleic Acids Research, 28: 235-242, 2000.

[Birney 2013] "Lessons for big data projects." Ewan Birney. Nature, Special Issue on the ENCODE project, 6 September 2012.

[Bos et al. 2007] "From Shared Databases to Communities of Practice: A Taxonomy of Collaboratories." Nathan Bos, Ann Zimmerman, Judith S. Olson, Jude Yew, Jason Yerkie, Erik Dahl, Gary M. Olson. Journal of Computer-Mediated Communication 12(2): 652-672 (2007).

[Brickley and Guha 2014] "RDF Vocabulary Description Language 1.0: RDF Schema." D. Brickley and R.V. Guha. February 2014. Available from http://www.w3.org/TR/2014/REC-rdf-schema-20140225/

[Britt and Larson 2003] "Constructing Representations of Arguments." M. A. Britt and A. A. Larson. Journal of Memory and Language, 48, 2003.

[Bry et al. 2012] "Semantic Wikis: Approaches, Applications, and Perspectives." François Bry, Sebastian Schaffert, Denny Vrandečić and Klara Weiand. Lecture Notes in Computer Science, Reasoning Web. Semantic Technologies for Advanced Query Answering, Volume 7487, 2012.

[Buckingham-Shum 2006] "Hypermedia Support for Argumentation-Based Rationale: 15 Years on from gIBIS and QOC." Simon J. Buckingham Shum, Albert M. Selvin, Maarten Sierhuis, Jeffrey Conklin, Charles B. Haley and Bashar Nuseibeh. In "Rationale Management in Software Engineering", Allen H. Dutoit, Raymond McCall, Ivan Mistrik, and Barbara Paech (Eds). Springer-Verlag, 2006.

[Chandrasekaran and Nersessian 2015] "Building Cognition: The Construction of Computational Representations for Scientific Discovery." Chandrasekharan, S. & Nersessian, N.J. To appear in Cognitive Science, 2015. Available from http://www.cc.gatech.edu/aimosaic/faculty/nersessian/papers/Building-cognition_Cogsci_Final_manuscript.pdf

[Chin et al. 2002] "New paradigms in problem solving environments for scientific computing." George Chin, Jr., L. Ruby Leung, Karen Schuchardt, Debbie Gracio. IUI 2002.

[Davenport 2013] "Thinking for a living: how to get better performances and results from knowledge workers." Davenport, Thomas H. Harvard Business Press, 2013.

[De Roure et al. 2009] De Roure, D.; Goble, C.; Stevens, R, The design and realization of the myexperiment virtual research environment for social sharing of workflows, Future Generation Computer Systems 25 (5), 2009.

[Filho et al. 2010] "Kolline: a task-oriented system for collaborative information seeking." Fernando Marques Figueira Filho, Gary M. Olson, Paulo Lício de Geus. SIGDOC, 2010.

[Foley et al.. 2011] Foley J a., Ramankutty N, Brauman K a., Cassidy ES, Gerber JS, Johnston M, et al. "Solutions for a cultivated planet." Nature. 2011 Oct 12. From www.nature.com/doifinder/10.1038/nature10452. Last accessed October 7, 2014.

[Garijo et al. 2012] "Quantifying Reproducibility in Computational Biology: The Case of the Tuberculosis Drugome." Garijo, D.; Kinnings, S.; Xie, L.; Xie, L.; Zhang, Y.; Bourne, P. E.; and Gil, Y. PLOS ONE, 2013.

[Gil 2013] "Social Knowledge Collection." Gil, Y. In Handbook of Human Computation, P. Michelucci (Ed). Springer, 2013.

[Gil et al. 2011] "Wings: Intelligent Workflow-Based Design of Computational Experiments." Gil, Y.; Ratnakar, V.; Kim, J.; Gonzalez-Calero, P. A.; Groth, P.; Moody, J.; and Deelman, E. IEEE Intelligent Systems, 26(1). 2011.

[Gil and Ratnakar 2013] "Knowledge Capture in the Wild: A Perspective from Semantic Wiki Communities." Gil, Y.; and Ratnakar, V. In Seventh ACM International Conference on Knowledge Capture (K-CAP), Banff, Canada, 2013.

[GLEON 2014] GLEON. The Global lakes Ecological Observatory Network. From http://www.gleon.org. Last accessed October 7, 2014.

[Grosz and Sidner 1990] "Plans for Discourse." B. J. Grosz and C. Sidner. In Cohen, Morgan and Pollack (eds.), Intentions in Communications, MIT Press, Chapter 20, pp. 417-444. 1990.

[Huss et al. 2010] "The Gene Wiki: community intelligence applied to human gene annotation." Huss JW, 3rd, Lindenbaum P, Martone M, Roberts D, Pizarro A, Valafar F, Hogenesch JB and Su AI. Nucleic Acids Res., 38:D633-D639, 2010.

[Hutchins 1995] “How a Cockpit remembers its speeds”, E. Hutchins, Cognitive Science, 19, 1995.

[ILEC. 2007] "Integrated Lake Basin Management: An Introduction." International Lake Environment Committee Foundation: Kusatsu, Japan. 2007.

[Introne et al. 2013] “Solving Wicked Social Problems with Socio-computational Systems.” Joshua Introne, Robert Laubacher, Gary M. Olson, Thomas W. Malone. KI - Künstliche Intelligenz 27(1): 45-52 (2013).

[Kamar et al. 2012] “Combining human and machine intelligence in large-scale crowdsourcing.” Ece Kamar, Severin Hacker, Eric Horvitz. AAMAS 2012.

[Kane et al. 2013] "Shared Decision Making Needs a Communication Record." Bridget Kane, Pieter Toussaint and Saturnino Luz. Computer Supported Cooperative Work and Social Computing (CSCW), San Antonio, Texas, February 2013.

[Kittur et al. 08] Aniket Kittur, Bongwon Suh, Ed H. Chi. “Can you ever trust a wiki? Impacting perceived trustworthiness in Wikipedia.” Proceedings of the ACM conference on Computer supported cooperative work, 2008.

[Kittur and Kraut 08] Aniket Kittur, Robert E. Kraut. “Harnessing the wisdom of crowds in Wikipedia: Quality through coordination.” Proceedings of the ACM conference on Computer supported cooperative work, 2008.

[Kittur et al. 09] Aniket Kittur, Bryant Lee and Robert E. Kraut. Proceedings of the 27th international conference on Human factors in computing systems, 2009.

[Kittur and Kraut 10] Aniket Kittur, Robert E. Kraut. “Beyond Wikipedia: coordination and conflict in online production groups.” Proceedings of the 2010 ACM conference on Computer supported cooperative work.

[Krafft et al. 2010] “VIVO: enabling national networking of scientists.” D Krafft, N Cappadona, B Caruso, J Corson-Rikert, M Devare, B Lowe, and VIVO Collaboration. Conference on Web Science (WebSci), Raleigh, NC, April 2010.

[Kraut and Resnick 2011] “Building Successful Online Communities: Evidence-Based Social Design.” Robert E. Kraut and Paul Resnick. MIT Press, 2011.

[Krötzsch et al. 07] Markus Krötzsch, Denny Vrandecic, Max Völkel, Heiko Haller, Rudi Studer. “Semantic Wikipedia.” Journal of Web Semantics, 5(4), pages 251-261, December 2007.

[Krötzsch et al. 2011] Markus Krötzsch, Denny Vrandecic: Semantic MediaWiki. Foundations for the Web of Information and Services 2011: 311-326

[Kuhn 09] Tobias Kuhn. “AceWiki: A Natural and Expressive Semantic Wiki.” Proceedings of the Fifth International Workshop on Semantic Web User Interaction (SWUI 2008), CEUR Workshop Proceedings, Volume 543, 2009.

[Lam et al. 10] Shyong (Tony) K. Lam, Jawed Karim, John Riedl. “The effects of group composition on decision quality in a social production community.” Proceedings of the 16th ACM international conference on supporting group work, 2010.

[Leskovec et al. 10] J. Leskovec, D. Huttenlocher, J. Kleinberg. Governance in Social Media: A case study of the Wikipedia promotion process. Proceedings of the AAAI International Conference on Weblogs and Social Media (ICWSM), 2010.

[Lintott et al. 10] Chris Lintott, Kevin Schawinski, Steven Bamford, An"e Slosar, Kate Land, Daniel Thomas, Edd Edmondson, Karen Masters, Robert C. Nichol, M. Jordan Raddick, Alex Szalay, Dan Andreescu, Phil Murray, Jan Vandenberg. “Galaxy Zoo 1: data release of morphological classifications for nearly 900,000 galaxies”. Monthly Notices of the Royal Astronomical Society, 2010

[Liu et al. 2014] "Supporting Task Resumption Using Visual Feedback." Yikun Liu, Yuan Jia, Wei Pan and Mark Pfaff. Computer Supported Cooperative Work and Social Computing (CSCW), Baltimore, Maryland, February 2014.

[Levin and Clark. 2009] "Report from Toward a Science of Sustainability Conference." Levin, S, and W Clark. Toward a Science of Sustainability, Airlie Center ~ Warrenton, Virginia, 2009.

[Mahling and Croft 1988]. “Relating human knowledge of tasks to the requirements of plan libraries.” D. E. Mahling and W. B. Croft. International Journal of Human-Computer Studies, Vol. 31, 1988.

[Mahling and Croft 1993]. “Acquisition and Support of Goal-Based Tasks.” D. E. Mahling and W. B. Croft. Knowledge Acquisition, Vol. 5, 1993.

[Mao et al. 2013] “Volunteering Versus Work for Pay: Incentives and Tradeoffs in Crowdsourcing.” Andrew Mao, Ece Kamar, Yiling Chen, Eric Horvitz, Megan E. Schwamb, Chris J. Lintott, Arfon M. Smith. HCOMP 2013.

[McGuinness et al. 2006] Deborah L. McGuinness, Honglei Zeng, Paulo Pinheiro da Silva, Li Ding, Dhyanesh Narayanan, and Mayukh Bhaowal. “Investigations into Trust for Collaborative Information Repositories: A Wikipedia Case Study.” Proceedings of the Workshop on Models of Trust for the Web, 2006.

[MEA 2005] "Ecosystems and Human Well-being: Synthesis." Millennium Ecosystem Assessment. Island Press, Washington, DC. 2005.

[Morgan et al. 2014] "Editing Beyond Articles: Diversity & Dynamics of Teamwork in Open Collaborations." Jonathan T. Morgan, Michael Gilbert, David W. McDonald and Mark Zachry. Computer Supported Cooperative Work and Social Computing (CSCW), Baltimore, Maryland, February 2014.

[NAE 2014] "NAE Grand Challenges in Engineering". National Academy of Engineering. Available from http://www.engineeringchallenges.org.

[Nature 2013] Nature, Special Issue on the ENCODE project, 6 September 2012.

[NEON 2014] NEON. National Ecological Observatory Network. From http://www.neoninc.org. Last accessed October 7, 2014.

[Nielsen 2011] "Reinventing Discovery." Nielsen M. Princeton University Press, 2011.

[Nonaka and Takeuchi 1995] Nonaka and H. Takeuchi, The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation. New York: Oxford University Press, 1995.

[NRC 2011] "Global Change and Extreme Hydrology: Testing Conventional Wisdom." The National Academies Press. Washington, D.C. 2011.

[NRC 2012] "Challenges and Opportunities in the Hydrologic Sciences." National Research Council, National Academy Press 161 pp, Washington D.C. 2012.

[Oleksik et al. 2014] "Study of an Electronic Lab Notebook Design and Practices that Emerged in a Collaborative Scientific Environment." Gerard Oleksik, Natasa Milic-Frayling and Rachel Jones. Computer Supported Cooperative Work and Social Computing (CSCW), Baltimore, Maryland, February 2014.

[Park et al. 2014] "Toward crowdsourcing micro-level behavioral annotations: the challenges of interface, training, and generalization." S. Park, P. Shoemark, and L.-P. Morency. IUI 2014.

[Pietras and Coury 1994] "The Development of Cognitive Models of Planning for Use in the Design of Project Management Systems." C. M. Pietras and B. G. Coury. International Journal of Human-Computer Studies, Vol 40, 1994.

[Polanyi 1983] M. Polanyi, Tacit Dimension. Gloucester, Mass.: Peter Smith Publisher Inc, 1983.

[Raban et al. 2010] Daphne R. Raban, Mihai Moldovan, Quentin Jones. "An empirical study of critical mass and online community survival." Proceedings of the ACM conference on Computer supported cooperative work, 2010.

[Rich et al. 2001] "COLLAGEN: Applying Collaborative Discourse Theory to Human-Computer Interaction." C. Rich, C. L. Sidner and N. Lesh. AI Magazine 22, 2001.

[Rich et al. 2005] "DiamondHelp: A Collaborative Task Guidance Framework for Complex Devices." C. Rich, C. L. Sidner, N. Lesh, A. Garland, Shane Booth and Markus Chimani. AAAI 2005.

[Ribes and Finholt 2009] "The long now of infrastructure: Articulating tensions in development." Ribes, D. and T. A. Finholt. Journal for the Association of Information Systems (JAIS): Special issue on eInfrastructures 10(5): 375-398, 2009.

[Smith and Goodman 1984] "Understanding Written Instructions: The Role of an Explanatory Schema." E. Smith and L. Goodman. Cognition and Instruction, 1(4), 1984.

[Spinellis and Louridas 2008] Diomidis Spinellis and Panagiotis Louridas. "The Collaborative Organization of Knowledge." CACM, August 2008.

[Steichen et al. 2013] "User-adaptive information visualization: using eye gaze data to infer visualization tasks and user cognitive abilities." B. Steichen, G. Carenini, and C. Conati. IUI 2013.

[Steinhardt and Jackson. 2014] Stephanie B. Steinhardt and Steven J. Jackson. "Reconciling Rhythms: Plans and Temporal Alignment in Collaborative Scientific Work." Computer Supported Cooperative Work and Social Computing (CSCW), Baltimore, Maryland, February 2014.

[Suweis et al. 2013] Suweis S, Rinaldo A, Maritan A, D'Odorico P. "Water-controlled wealth of nations." Proceedings of the National Academy of Sciences of the United States of America. 2013 Mar 12. From www.ncbi.nlm.nih.gov/pmc/articles/PMC3600477/. Last accessed October 7, 2014.

[Takeuchi and Nonaka 2004] H. Takeuchi and I. Nonaka, Hitotsubashi on Knowledge Management. Singapore: John Wiley & Sons (Asia), 2004.

[Tausczik et al. 2014] "Collaborative Problem Solving: A Study of MathOverflow." Yla R. Tausczik, Aniket Kittur and Robert E. Kraut. Computer Supported Cooperative Work and Social Computing (CSCW), Baltimore, Maryland, February 2014.

[Van Merrienboer et al. 2003] "Taking the load of a learners' mind: Instructional design for complex learning." Van Merriënboer, J.J.G., Kirschner, P.A., & Kester, L. Educational Psychologist, 38(1), 2003. I.