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

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

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.

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

Personal Worklist: 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.

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

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

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.

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

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.

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

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

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: Specify deadlines for tasks, while leaving people in control
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 w. visibility into intermediate steps
F4: Export of data and results, integration with existing standards

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. Figure 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.

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 Sub-task 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 5. Heat maps for task pages showing user clicks.

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 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