An Initial Framework For Organic Data 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
An abstract.

Author Keywords
Keyword1; keyword2.

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 quest (e.g., the Human Genome Project). Our work focuses on scientific collaborations that revolve around complex science questions that require:

- *multi-disciplinary contributions*, so that the participants belong to different communities with diverse practices and approaches
- *significant coordination*, where ideas, models, software and data need to be discussed and integrated to address the shared science goals
- *unanticipated participants*, so that the collaboration needs to grow over time and include new contributors that may bring in new knowledge, skills, or data

Such scientific collaborations do occur but are not very common. Unfortunately, they take a significant amount of effort to pull together and to sustain for the usually long period of time required to solve the science questions. Yet, these kinds of collaborations are needed in order to address major engineering and science challenges ahead (e.g., http://www.engineeringchallenges.org). Our goal is to develop a collaborative software platform that supports such scientific collaborations, and ultimately make them significantly more efficient and commonplace.

This paper presents an **Organic Data Science framework** to support scientific collaborations that revolve around complex science questions that require multi-disciplinary contributions to gather and analyze data, significant coordination to synthesize findings, and grow organically to accommodate new contributors as needed as the work evolves over time. The key idea is to open science by exposing science processes declaratively to enable broader participation. Science processes describe the what, who, when, and how of the activities pursued by the collaboration. The framework is still under development, and it evolves to accommodate user feedback and to incorporate new collaboration features.

There is a significant body of work on studying 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.

There are a wide range of approaches that have been explored for collaboration, although they have not had much adoption in science practice [Introne et al 2013]. ADD MORE HERE.

The paper begins with a motivating scenario of a complex science task that we are currently pursuing using this framework. We then review prior work on social studies that discuss the nature and challenges of scientific collaborations, and on interfaces developed to support on-line collaboration.

MOTIVATING SCENARIO (~1 PAGE)

Text.

RELATED WORK (~1 PAGE)

We discuss related work in three categories: approaches to scientific collaboration, on-line collaboration systems, and task-centered user interfaces.

Scientific Collaboration

[Bos et al 2007] did a comprehensive multi-year study of scientific collaborations and propose seven types of collaboratories (MAYBE PUT THIS IN A TABLE??): *1) Shared Instruments*, where instruments or sensors are used by a community (e.g., National Ecological Observatory Network [cite NEON]); *2) Community Data Systems*, where a data resource is maintained and used by a community (e.g., the Protein Data Bank [cite PDB]), *3) Open Community Contribution Systems*, where tasks are carried out by a community including citizen scientists (e.g., the GalaxyZoo citizen science project for labeling galaxy images [cite Zooniverse]), *4) Virtual Communities of Practice*, where a community shares interest in specific research topics (e.g., the Global Lake Ecological Observatory Network [cite GLEON]), *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 [cite ENCODE], and *7) Community Infrastructure Projects*, where a community gets together to develop shared 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, since the project is jumpstarted by a multi-institutional collaboration, and is an open community contribution system but without the prescribed tasks typically found on those systems. Organic Data Science can be considered a new type of collaboratory, where the tasks are defined on the fly as the project progresses and the collaboration includes 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. 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

CITE HERE WORK FROM THE CSCW CONFERENCE.

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. In addition, climate researchers can be considered one discipline, and we are investigating the integration of multi-disciplinary research.

Task-Centered User Interfaces

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 is 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 tend to be tasks that have to do with interface use, rather than organizational or coordination tasks.

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 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 a workflow system is an area of future work.

APPROACH (2 PAGES)

Key features of our approach are:

1. **providing a task-oriented nexus driven by science goals** that connects scientists together, organizing tasks to help scientists track where they can contribute and when, as well as their past contributions
2. **incorporating principles from social sciences research on successful on-line collaborations**, including best practices for retention and growth of the community
3. **opening the science process in that the framework exposes all tasks and activities publicly**, so that all participants (especially newcomers) can immediately see what work is being done and what tasks they can contribute to

Task-Centered Collaborative Spaces

In practice, the contributors to the organic data science framework form an organization. We use tasks as an organizational mechanism for coordination. [Polanyi 1983] coined the terms and discussed differences between tacit as well as explicit knowledge of individuals in organizations. According to Polanyi an individual can have tacit knowledge without being able to explicitly express this knowledge in its essence. In contrast, explicit knowledge can be communicated in formal languages that can be processed by other persons. 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 tacit knowledge of researchers to resolve and formulate tasks in the science process through ad-hoc collaboration in an open framework. While we are focusing on science processes in this paper, Davenport also described the importance of processes and tasks for the productivity of knowledge workers in an organizational context [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].

Temporal aspects of task achievement are also important. In project management, the duration estimates and resource selection have been found to be 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.

Social Principles

There are numerous studies about successful on-line communities [Kraut and Resnick 2011]. Many studies are focused on Wikipedia and other wiki-style frameworks, 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 across contributions [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]. These studies suggest a number of principles for the design of our on-line collaboration framework.

Figure 1 summarizes the social principles that we are using in our approach. We follow the organization used in [Kraut and Resnick 2011], but we focus here 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). The principles are written to be self-explanatory, and in the next section we will explain how they map to features in our user interface (marked with numbers on the right-hand side of the figure).

Opening Science Process

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. The collaboration is centered around tasks, that contributors create, decompose, reformulate, and resolve. This project 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 guidelines that serve as 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 potential contributors can see a way to contribute.

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.

1. **Starting communities**
    - 1.1. Carve a niche of interest, scoped in terms of topics, members, activities, and purpose
    - 1.2. Relate to competing sites, integrate content
    - 1.3. Organize content, people, and activities into subspaces once there is enough activity
    - 1.4. Highlight more active tasks
    - 1.5. Inactive tasks should have "expected active times"
    - 1.6. Create mechanisms to match people to activities
2. **Encouraging contributions through motivation**
    - 2.1. Make it easy to see and track needed contributions
    - 2.2. Ask specific people on tasks of interest to them
    - 2.3. Simple tasks with challenging goals are easier to comply with
    - 2.4. Specify deadlines for tasks, while leaving people in control
    - 2.5. Give frequent feedback specific to the goals ("immersive")
    - 2.6. Requests coming from leaders lead to more contributions
    - 2.7. Stress benefits of contribution
    - 2.8. Give (small, intangible) rewards tied to performance (not just for signing up)
    - 2.9. Publicize that others have complied with requests
    - 2.10. People are more willing to contribute: 1) when group is small, 2) when committed to the group, 3) when their contributions are unique
3. **Encouraging commitment**
    - 3.1. Cluster members to help them identify with the community
    - 3.2. Give subgroups a name and a tagline
    - 3.3. Put subgroups in the context of a larger group
    - 3.4. Make community goals and purpose explicit
    - 3.5. Interdependent tasks increase commitment and reduce conflict
4. **Dealing with newcomers**
    - 4.1. Members recruiting colleagues is most effective
    - 4.2. Appoint people responsible for immediate friendly interactions
    - 4.3. Introducing newcomers to members increases interactions
    - 4.4. Entry barriers for newcomers help screen for commitment
    - 4.5. When small, acknowledge each new member
    - 4.6. Advertise members particularly community leaders, include pictures
    - 4.7. Provide concrete incentives to early members
    - 4.8. Design common learning experiences for newcomers
    - 4.9. Design clear sequence of stages to newcomers
    - 4.10. Newcomers go through experiences to learn community rules
    - 4.11. Provide sandboxes for newcomers while they are learning
    - 4.12. 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).

5. **Best practices from Polymath**
    - 5.1. Permanent URLs for posts and comments, so others can refer to them
    - 5.2. Appoint a volunteer to summarize periodically
    - 5.3. Appoint a volunteer to answer questions from newcomers
    - 5.4. Low barrier of entry: make it VERY easy to comment
    - 5.5. Advance notice of tasks that are anticipated
    - 5.6. Keep few tasks active at any given time, helps focus
6. **Lessons learned from ENCODE**
    - 6.1. 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
    - 6.2. Written and publicly accessible rules to transfer work between groups, to assign credit when papers are published, to present the work
    - 6.3. Quality inspection with visibility into intermediate steps
    - 6.4. 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.

TASK CENTERED FRAMEWORK DESIGN

Our framework builds on top of the Semantic Media Wiki and uses the semantics to structure the task centered data. Normally users are scared by using semantic data structures. Therefore we provide a user interface which is easy to use and users not need to care about the semantics annotation. We designed all features based on the Social Design Principles. In the following all features explained in detail.

❶**Task Representation:** Every task is represented with an adapted wiki page. On the top the parent task is placed. In the next line the actual task name is written. Below that a subtask explorer navigation is provided. Alternatively a time line visualization of all subtask can be represented, this is illustrated in the Figure XX. After the subtask block a gray box represents all tasks metadata. Everything below this box is page content.

❷**Task Metadata:** Our approach distinguish between mandatory and nice to have metadata. Mandatory metadata attributes are needed e.g. to calculate the task progress or link it to a person. Tasks are most relevant in a certain time interval. Therefore every task has a defined start and target date. Start and target date must be in the scope of the parent task. This time interval is used to calculate the progress of tasks depending on the task type. We introduced three different task types. High-level tasks have a high abstraction grade and a high uncertainty in the estimation of the task completion. E.g. a task on a project level. Medium-level tasks have a medium uncertainty in estimation of the task completion. E.g. represents an activity within a project and is may split into several subtasks. Low-level tasks have a low uncertainty in estimation of the task completion. E.g. small well defined tasks which can be accomplished in a short time period. We indicate the task type with different green colors in the task icon. High-level task is light green and a low-level tasks is dark green. The progress of every task is calculated dependent on the task type. For high-level tasks we estimate a linear progress based on the start and target date in relation to the today's date. The progress of Medium-level tasks is calculated as average of the subtasks progress. Low-level tasks are estimated by users because they know it best. To provide simple user feedback we created icons for each metadata attribute the color of this icon indicates the state of this attribute. Not defined attributes are gray, valid defined attributes are green and attributes which are inconsistent with attributes of the parent task are yellow.

❸**Task Explorer:** Similar to well-known hierarchical folder navigation we provide a hierarchical task navigation. The nested task structure can be expanded until the leaf is reached or the searched task is found. Additionally we provide a task title search and a task expertise filter. All tasks which does not match with filter are hidden. Except parent tasks which have matching subtasks are represented

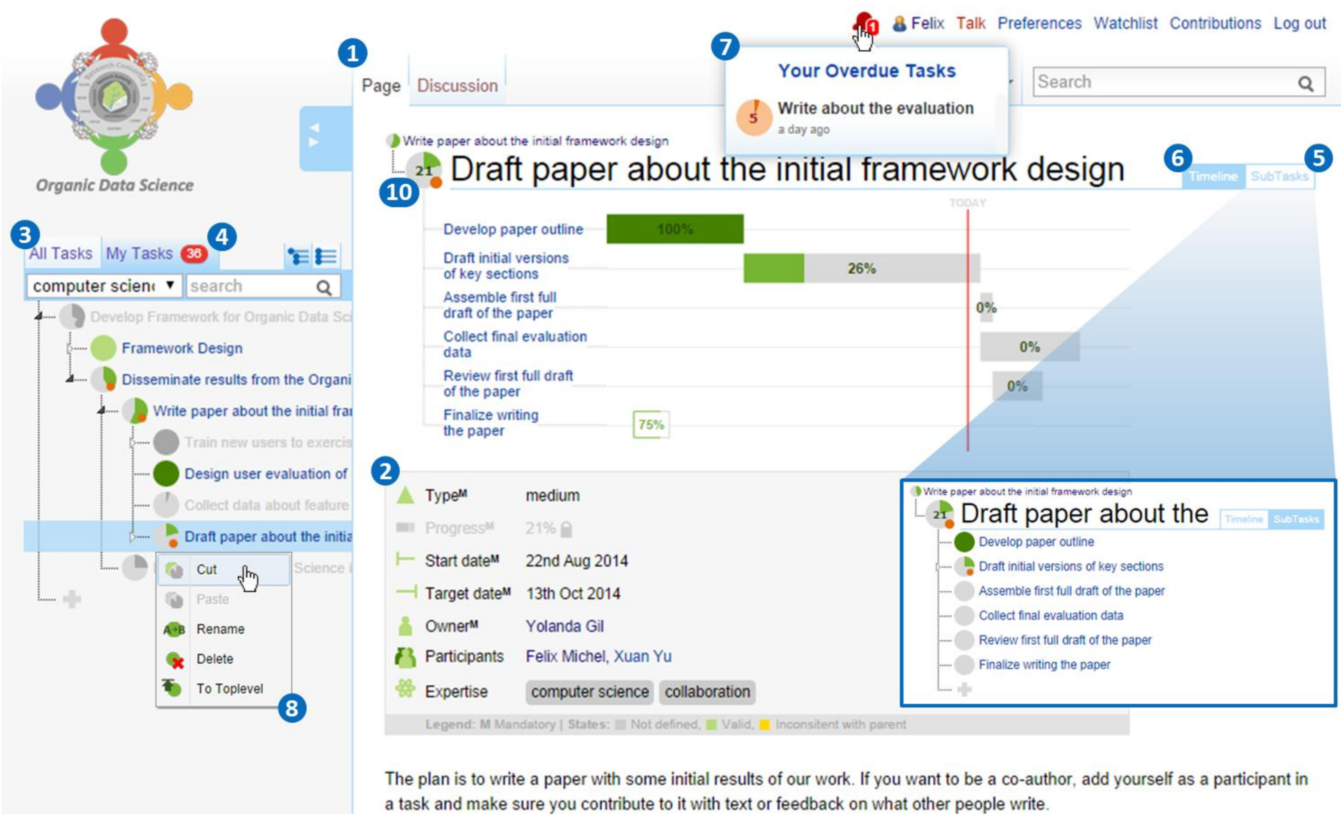

Figure 1: Organic Data Science Task Page.

fade out to provide context.

❹**Personal Worklist:** The worklist contains a subset of tasks form the task explorer. All tasks which contain the logged in users as owner or as participant are part the worklist. A red counter indicates the current number of tasks in the worklist.

❺**Subtask Explorer:** Subtasks of the currently opened task are presented. The navigation works similar to the Task Explorer. No filters and search options are provided.

❻**Timeline Explorer:** All subtasks are represented regarding the time context comparable to a Gantt chart. Tasks on a meta-level are represented with an empty rectangle. The green part of the border shows the percentage of completed metadata. Content level tasks are basically illustrated with a gray rectangle. The green part represents the progress of the task. Start and target date define the position. Navigating via timeline works similar to the subtask explorer.

❼**Task Alert:** A task alert occurs when a task is not completed until target date is reached. Only the task owner get this alert notification. The owner responsible to complete the task or get other users involved completing the task.

❽**Task Actions:** We support ad-hoc collaboration this leads to an emerging task structure. After a certain time the task structure need to be adapted. E.g. introducing a new top level task. This frameworks supports actions like creating, renaming, moving and deleting tasks. To ensure a good usability all actions requires reversibility. Subtasks can be created via subtask explore with the plus button below the last task. Root tasks can be crated in the task explorer on the left. If an expertise is selected or/and the "my task" tab is used the new task is created with the selected metadata. All other actions are accessible via context menu. Deleting a task means deleting the task itself and all their subtasks. Tasks can be moved with the cut and paste operation. Moving a task to the root level works only with the "To Toplevel" action. All moving actions can cause inconsistent task hierarchies. E.g. the time interval of the pasted task does not fit into the time interval of the new parent task. The same problem can occur with the tasks type. All tasks which have an inconstant state are highlighted in yellow and the parent tasks indicate this inconsistency with a small yellow triangle.

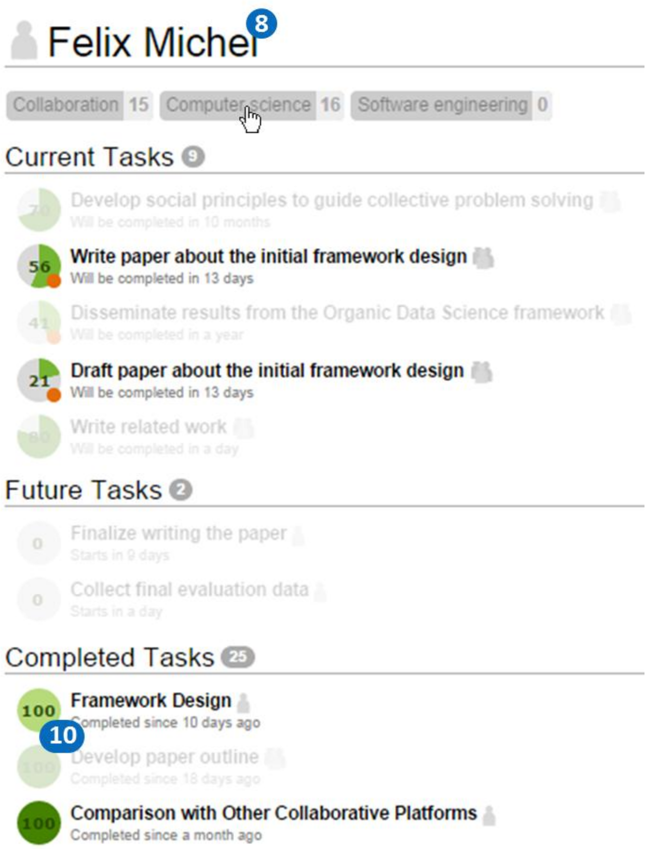

Figure 2: Organic Data Science Person Page.

❾**User related Tasks and Expertise:** Allows users to easily see what other users planning to work on, they recently working on and on what they have worked in the past. This creates a transparent working process. This makes it easy for newcomers to browse tasks of topic related users and help finding important tasks for themselves. The top of every user page contains a user icon followed by the user name. Users are individual persons and every user has expertise in a certain field. Hovering over a certain expertise fades out all not related tasks.

❿**Task State Representation:** The state of every task is summarized within the task icons (See Figure 3). Basically we distinguish between tasks on meta-level and tasks on content level. Tasks on meta-level are represented with a cycle, the green part indicates the percentage of completed metadata. All tasks which have completed all mandatory metadata are content level tasks. Content level tasks are represented by a pie chart, the green part indicates the competed part. Different green are used to express the task type.

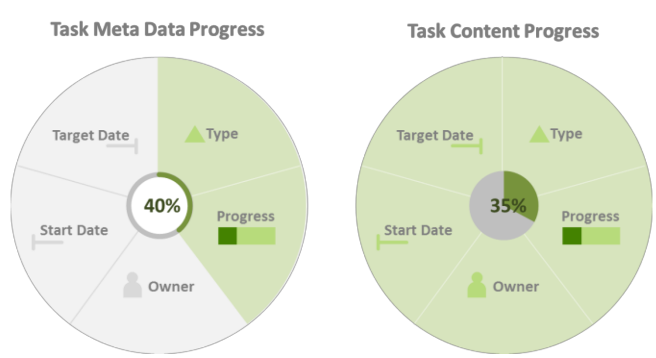

Figure 3: Conceputal Task Reprenetation.

We extended the explained states representation with the states illustrated in Figure xx to cover all task states. All representations can be distinguished along two dimensions meta- or content-level and active or passive. Tasks which provides context but not match an applied filter are passive tasks. This tasks are fade out versions of the active tasks. The progress of normal tasks is shown in different greens depending on the task type, this is not visible in the Figure XX. The progress over overdue tasks are indicated with an orange pie chart. A small orange point indicates that at least on subtask of any level is an overdue task. This is especially helpful to help users finding overdue via task explorer. Yellow colored icons indicate inconsistent tasks caused by move operations. The yellow triangles indicate an inconsistent subtask similar to the overdue subtasks. All task icons existing in three different size. Figure XX shows the most used medium icons. Large icons additional contain the progress in percentage as text.

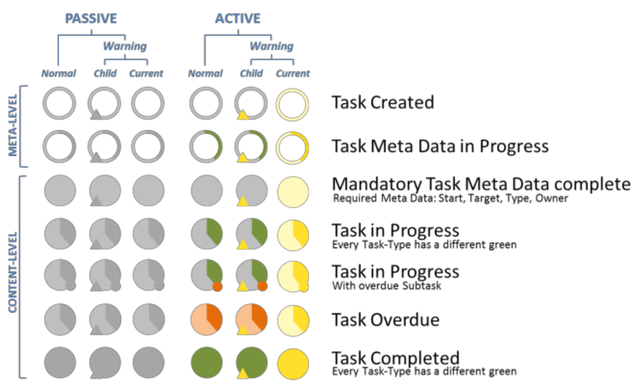

Figure 4: Task Sates Representations.

A task state changes many times until a task is completed. Figure XX represents simplified task states sample sequences. Many more task state changes are possible. E.g. it is possible to switch task states vertically and continue in another sequence horizontally.

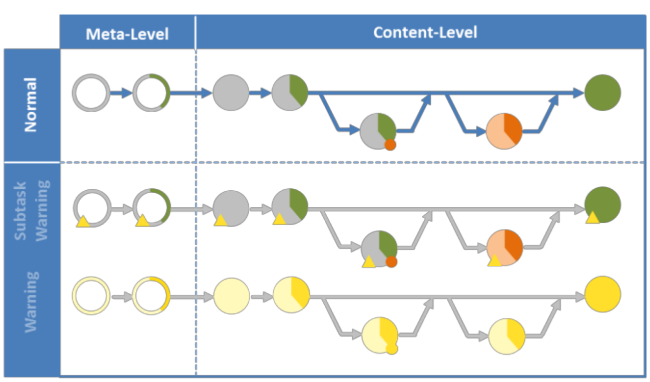

Figure 5: Task State Sample Sequences.

DISCUSSION
Text.

CONCLUSION
Text.

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

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

[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).

[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] François Bry, Sebastian Schaffert, Denny Vrandečić and Klara Weiand. "Semantic Wikis: Approaches, Applications, and Perspectives." 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, Bashar Nuseibeh. In "Rationale Management in Software Engineering", Allen H. Dutoit, Raymond McCall, Ivan Mistrik, and Barbara Paech (Eds). Springer-Verlag, 2006.

[Chin 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.

[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.

[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.

[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.

[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, Robert E. Kraut. "Coordination in collective intelligence: the role of team structure and task interdependence." 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.

[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.

[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.

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

[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.

[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.

[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.

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

[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.

The columns on the last page should be of approximately equal length.
Remove these two lines from your final version.