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 and other wiki-style frameworks. 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. Argumentation interfaces facilitate the collaborative synthesis of diverse ideas [Conklin 1995], and have been used in the context of science [Filho et al 2010]. 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)

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
Some on-line collaboration tools have been developed to support 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-Oriented Collaboration Tools
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.

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

THIS SECTION MOTIVATES TASK-CENTERED COLLABORATION, SOME ROUGH TEXT HERE FOR NOW,

[Polanyi] 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 [H. Takeuchi and I. Nonaka]. In the organic data science 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 for the productivity of knowledge workers in an organizational context [Davenport, Thomas H].

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.

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].

Social Principles

Figure 1 summarizes the social principles that we are using in our approach.

EXPLAIN THE FIGURE HERE BRIEFLY .

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]. Social norms are found in other collaborations [Kraut and Resnick 2011; Birney 2013], and incorporate mechanisms for adjudication and credit.

Another project that has exposed best practices of a large collaboration is ENCODE [Birney 2012; Nature 2012].

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

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
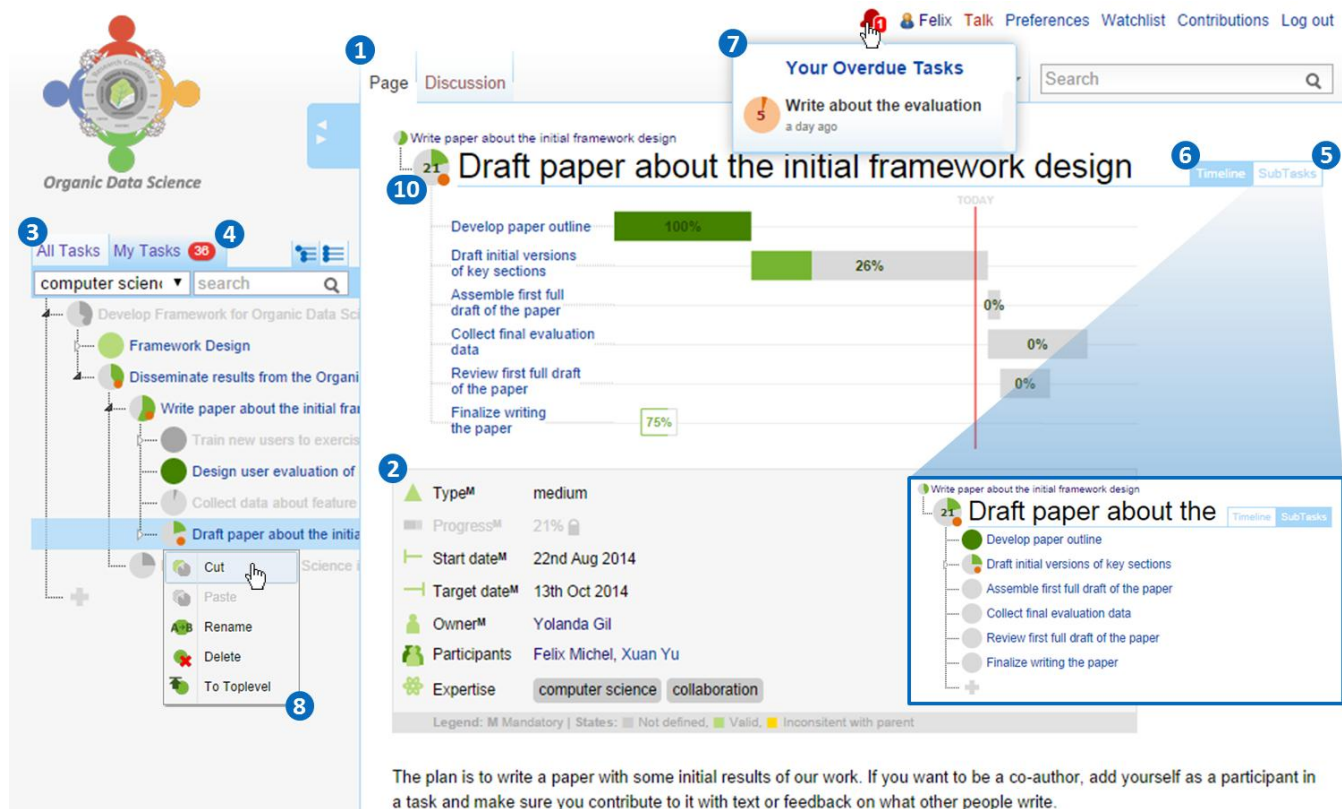

Figure 1: Organic Data Science Task Page.

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.

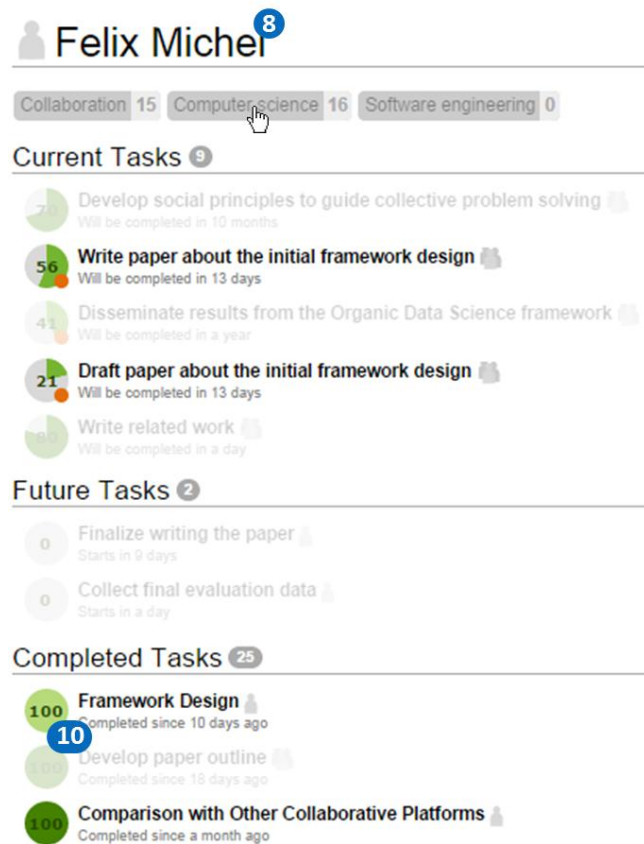❽**Task Actions:** We support ad-hoc collaboration this leads to an emerging task structure. After a certain time the

Figure 2: Organic Data Science Person Page.

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.

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.
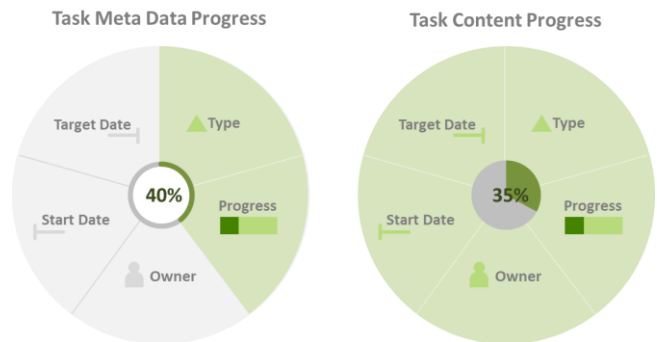

Figure 3: Conceptual Task Reprenetation.

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.
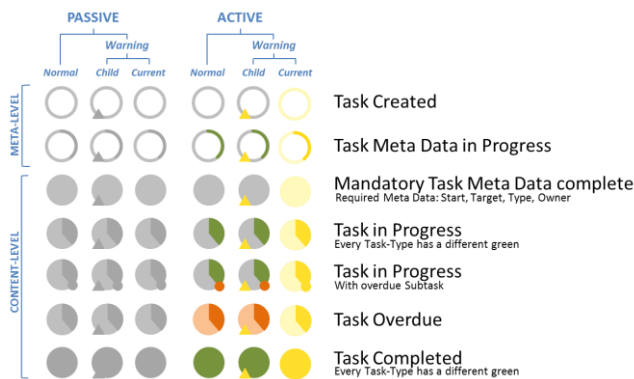
Figure 4: Task Sates Representations.

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.
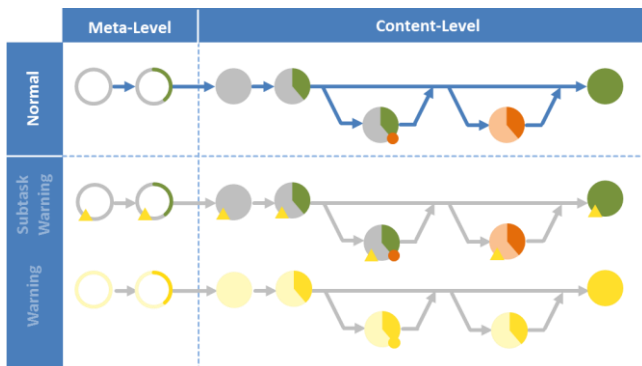
Figure 5: Task State Sample Sequences.

EVALUATION

We evaluated our collaborative framework approach with the organic data science wiki. Currently this wiki has 18 registered users, thereof 12 active users and contains 122 Tasks. Within 10 weeks we collected around 19,000 log entities. All task pages together have been accessed more than 2,900 times. All person pages together have been accessed 328 times.

How easy is it for new people to participate? Our approach to lower the participation barrier is to provide our users a trainings wiki with several tasks to train. The trainings tasks follow the documentation structure and guides the new users practicing relevant topics. To evaluate how valuable the documentation in combination with the training we evaluate: *How often is the documentation accessed after training?*

How often tasks are deleted short after creation by the same user? One user deleted in sum 3 tasks within five minutes after creation. This happened during he accomplished the trainings tasks in the trainings wiki.

What is the new user's total training time?

Ask Matheus, Hilary 1h+x, Xuan

How do users find relevant tasks? We analyzed the log data of our self-implemented java-script tracking tool. We measured how tasks pages are opened. Most users used the task explorer navigation to find their relevant pages. This is probably because users have overview over all tasks. At the same time it is possible to find relevant tasks quickly with a drill down or certain filters. The task alert not used that often but we expect that this feature will be more important a rising amount of tasks over time.
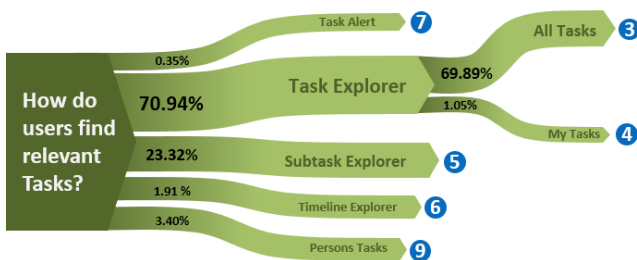

Figure 6: Finding Tasks via Features.

Where are collaboration touch points? To measure the collaboration we split this question into four sub questions. The system contains 122 tasks but the total number of tasks for the following questions can be higher due to the reason that tasks can be renamed or deleted. All results are illustrated in Figure XX.

A *Is more than one person viewing the task page?* Currently 47% of all task pages are accessed by only one person. The high vale is due to the reason that we consider also new created pages. The percentage of new create pages is in a new wiki comparable high.

B *Is more than one person signed up for each task?* For the evaluation we count the participants and add one person for the owner if set. With 29 percentage it is most common that 3 persons are assigned to a task. Tasks with two person

C *Is more than one person editing task metadata?* Currently 78 percentage of all tasks metadata is edited by only one percentage. The task creator often also adds the initial metadata which leads to such high percentage. Around 15 of the tasks metadata is edited by 2 persons.

D *Is more than one person editing the content of tasks?* The result of this question has a similarity with previous question. The tasks content is edited to 87 percentage by one person and around 10 percentage of two persons.
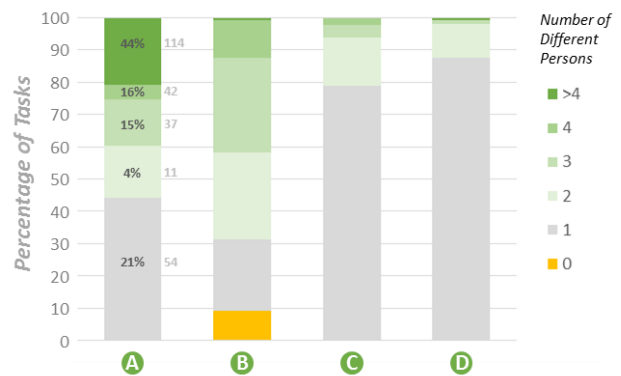

Figure 7: Task Collaboration Evaluation.

Is a person collaborating with more than on other person?

We used the tasks metadata attributes to evaluate this. Every task has a metadata attribute owner and participants. We created an artificial users set collaborators which combines the owner and participants. In the next step we created user-collaboration pairs and counted how often they collaborate on tasks. The result is illustrated in Figure XX. Users are represented as nodes and the number of tasks they have in common is expressed with the strength of edges. It is simply visible that there are basically two strong collaborations exist. This spread collaboration groups exist because we there are two main goals. The smaller collaboration group is developing this organic data science framework and the larger group represents the researchers which use this framework to accomplish their science goals.
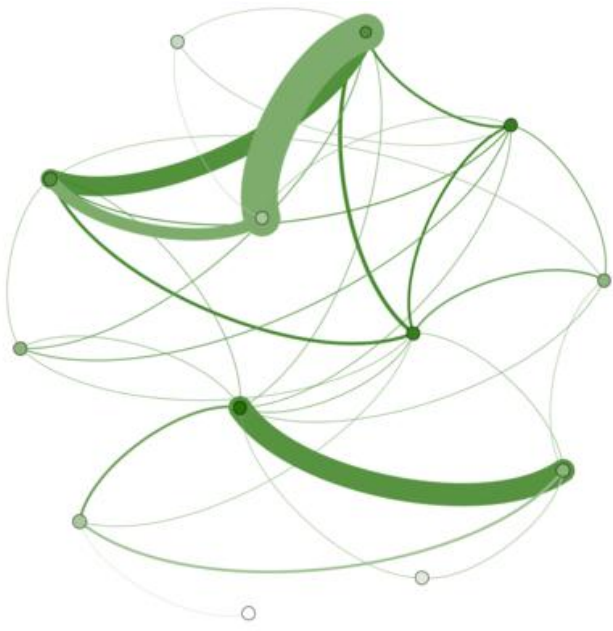
DISCUSSION
Text.

CONCLUSION
Text.

REFERENCES

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

2. [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.

3. [Chklovski 05] Timothy Chklovski. Designing interfaces for guided collection of knowledge about everyday objects from volunteers. Proceedings of IUI 2005.

4. [Chklovski and Gil 05a] Tim Chklovski and Yolanda Gil. "Towards Managing Knowledge Collection from Volunteer Contributors," Proceedings of the 2005 AAAI Spring Symposium on Knowledge Collection from Volunteer Contributors (KCVC), Stanford, CA, March 2005.

5. [Chklovski and Gil 05b] Tim Chklovski and Yolanda Gil. "An Analysis of Knowledge Collected from Volunteer Contributors," Proceedings of the Twentieth National Conference on Artificial Intelligence (AAAI-05), Pittsburgh, PA, July 9-13, 2005.

6. [Colorsquet et al 09] Corlosquet S, Delbru R , Clark T, Polleres A, and Decker S. Produce and Consume Linked Data with Drupal. 8th International Semantic Web Conference (ISWC 2009); October 25-29, 2009, Washington DC.

7. [Haase et al 09] Peter Haase, Daniel Herzig, Mark Musen, Thanh Tran. "Semantic Wiki Search." Proceedings of the 6th European Semantic Web Conference, 2009.

8. [Happel 10] Hans-Jörg Happel. "Woogle4MediaWiki: from searchers to contributors." Proceedings of the 6th International Symposium on Wikis and Open Collaboration, 2010.

9. [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.

10. [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.

11. [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.

12. [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.

13. [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.

14. [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.

15. [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.

16. [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.

17. [Weld et al 08] Daniel S. Weld, Fei Wu, Eytan Adar, Saleema Amershi, James Fogarty, Raphael Hoffmann, Kayur Patel, Michael Skinner. "Intelligence in Wikipedia." Proceedings of the 23rd national conference on Artificial intelligence, 2008.

18. [Wu et al 08] Fei Wu, Raphael Hoffmann, Daniel S. Weld. "Information extraction from Wikipedia: moving down the long tail." Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, 2008.

19. [Wu and Weld 07] Fei Wu, Daniel S. Weld. "Autonomously semantifying Wikipedia." Proceedings of the sixteenth ACM conference on information and knowledge management, 2007.

20. [Wu and Weld 08] Fei Wu, Daniel S. Weld. "Automatically refining the Wikipedia infobox ontology." Proceedings of the 17th international conference on World Wide Web, 2008.

MORE REFERENCES:

1. BRITT, M. A., AND A. A. LARSON. "Constructing Representations of Arguments", Journal of Memory and Language, 48, 2003.

2. MAHLING, D.E. AND W.B. CROFT. "Relating human knowledge of tasks to the requirements of plan libraries." International Journal of Human-Computer Studies, Vol. 31, 1988.

3. MAHLING, D.E. AND W.B. CROFT. "Acquisition and Support of Goal-Based Tasks." Knowledge Acquisition, Vol. 5, 1993.

4. PIETRAS, C. M. AND B. G. COURY. "The Development of Cognitive Models of Planning for Use in the Design of Project Management Systems." International Journal of Human-Computer Studies, Vol 40, 1994.

5. SMITH, E. AND L. GOODMAN. "Understanding Written Instructions: The Role of an Explanatory Schema." Cognition and Instruction, 1(4), 1984.

6. VAN MERRIËNBOER, J.J.G., KIRSCHNER, P.A., & KESTER, L. "Taking the load of a learners' mind: Instructional design for complex learning." Educational Psychologist, 38(1), 2003.

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

8. Joshua Introne, Robert Laubacher, Gary M. Olson, Thomas W. Malone: Solving Wicked Social Problems with Socio-computational Systems. KI 27(1): 45-52 (2013)

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

10. Fernando Marques Figueira Filho, Gary M. Olson, Paulo Lício de Geus: Kolline: a task-oriented system for collaborative information seeking. SIGDOC 2010: 89-94.

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

12. [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.

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

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