Collaborative Development of Data Curation Profiles on a Wiki Platform: Experience from Free and Open Source Software Projects and Communities

Sulayman K. Sowe*
Information Services Platform laboratory.
NICT, 3-5 Hikaridai, Seika-cho,
Kyoto, 619-0289, Japan.
sowe@nict.go.jp

Koji Zettsu
Information Services Platform laboratory.
NICT, 3-5 Hikaridai, Seika-cho,
Kyoto, 619-0289, Japan.
zettsu@nict.go.jp

ABSTRACT

Wiki technologies have proven to be versatile and successful in aiding collaborative authoring of web content. Multitude of users can collaboratively add, edit, and revise wiki pages on the fly, with ease. This functionality makes wikis ideal platforms to support research communities curate data. However, without appropriate customization and a model to support collaborative editing of pages, wikis will fall sort in providing the functionalities needed to support collaborative work. In this paper, we present the architecture and design of a wiki platform, as well as a model that allow scientific communities, especially disaster response scientists, collaborative edit and append data to their wiki pages. Our experience in the implementation of the platform on MediaWiki demonstrates how wiki technologies can be used to support data curation, and how the dynamics of the FLOSS development process, its user and developer communities are increasingly informing our understanding about supporting collaboration and coordination on wikis.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; D.2.10 [**Software Engineering**]: Design—*methodologies, representation*

General Terms

Design, Human Factors, Management, Theory

Keywords

Wiki, MediaWiki, Open collaboration, FLOSS communities, Data curation, Data curation profiles, Cloud computing.

*Correspondence author: Dr. Sulayman K. Sowe. NICT, 3-5 Hikaridai, Kyoto, 619-0289, Japan.
Tel:+81-774-98-6871; Fax:+81-774-98-6960

WikiSym + OpenSym '13, August 05 - 07 2013, Hong Kong, China

1. INTRODUCTION

What makes Free/Libre/Open Source Software (FLOSS) or Open Source Software (FOSS) development more remarkable is not so much about the software itself, but rather the way the communities in various projects operate [40, 41]. Supported by collaboration and coordination tools (e.g. Mailing Lists, Forums, CVS/SVN, Bug Tracking Systems, Wikis), FOSS communities are typified by voluntary (some paid) contributions to FOSS projects, and exhibit some form of meritocratic governance. Extensive peer collaboration allows project participants to write code, debug, test, integrate software and components, and help newbies.

In the literature, many terms are used to illustrate how FOSS communities go about their work: wisdom of crowds [21], collective intelligence [21, 25, 14], peer production [1], open collaboration [36, 30]. The latter promise immense benefits, but is also a challenge to sustaining wiki content creators [9]. Open collaboration in FOSS development [36, 32] has been successful in producing a number of successful application including operating systems (GNU / Linux), servers (Apache), databases (MongoDB), browsers (Firefox), cloud computing infrastructure (Apache Hadoop, Cassandra), and some of the most widely used wiki technologies (MediaWiki, TWiki, TikiWiki, DokuWiki, etc.).

The capability of wikis such as Wikipedia to support collaborative authoring of content is well documented (e.g.[10, 21, 43, 1, 3]). The Wikipedia success story [34] demonstrates the intricate nature of open collaboration, and how we can build a knowledge repository from a variety of domains [21]; ranging from education [24], R&D in engineering labs [43], multilingual environments [10], business and law [12], to disasters response situations [45].

However, despite the compendium of research, designing wikis and employing effective mechanisms to support participants curate data [19], while collaboratively authoring wiki pages in specific context remains a fundamental challenge. Karen and Baker [19] recognized that data curation is not a static process, but a set of repeatable activities focusing on tending data and creating data products within a particular arena, such as disaster response. Data curation, according to [6], can be defined as "a means to collect, organize, validate, and preserve data". Lord, et al. [31] defined data curation as "the activity of managing and promoting the use of data from its point of creation, to ensure it is fit for contemporary purpose, and available for discovery and re-use". On the other hand, a data curation profile can be seen as an

"instrument that can be used to provide detailed information on particular data forms" [26]. As such, a researcher's data curation profile (DCP) documents all the actions that might be needed to ensure full utilization and sharing of a particular dataset used in a given research context.

Data curation on a wiki goes beyond collaboratively editing of data curation profiles alone. It also involves providing users the tools they would need to enable them search for relevant data, link other data sources to their DCP, upload or register data for their own DCP and for other curators, annotate texts, make data citations, track data provenance [8], and discuss or talk about their research or data curation profile(s). All or some these activities can be mapped to a user data curation profile in order to provide detailed information on particular data forms that are curated by the individual [26].

The challenge of extending the functionality of wikis to cater for research data needs have been addressed before [39, 5, 16, 46]. Nevertheless, Orlandi and Passant [29] described wikis as isolated systems. The authors posit that wikis introduce several drawbacks when users need to access and link heterogeneous data or information from another source such as external database archives. Ignat et al. [18] further argued that, because of their centralized architecture, wiki systems (e.g. MediaWiki) offer limited support for cooperative authoring, and they do not scale well.

Despite the drawbacks and challenges, databases and wikis are said to have complementary strengths and weaknesses for use in collaborative data management and data curation situations [5]. While databases might not be ideal for supporting collaborative work like in wikis, they are scalable and have good query optimization features for storing and retrieving data and metadata for research purpose. On the other hand, the front-end of most wikis, by design, are easy to use, encourage sharing, and provide built-in support for archiving, tracking and annotating (mostly text) content [5].

1.1 Motivation and Research Contribution

Motivated by the complementary strengths and weaknesses of both wikis and databases, our intention is to develop an integrated community platform to support disaster response scientists collaboratively work on their data curation profiles by editing wiki pages, and at the same time be able to search and link datasets to enhance the value of their data curation profiles. This research presents our proof of concept (POC) in which we describe the architecture of a wiki platform for accomplishing data curation tasks, as well as a data curation model that allow disaster response scientists edit and append remote datasets to their data curation profiles.

In the architecture, we describe the components and subcomponents required for developing our wiki cloud platform. The "cloud", according to [13], consists of hardware, data storage, networks, APIs, and services that provide the means for users to access services on demand. The model for data curation and collaborative editing of wiki pages or DCPs benefits from our involvement and understanding of the way various FOSS communities and projects work.

The subject of data curation and data curation profiles is well known in the realms of library sciences, museums, and university faculties (e.g. [19, 25, 26, 31]). However, we are not aware of any organization or project that has leveraged wikis to support data curators collaboratively work on their profiles. Moreover, we could not find any framework, model, or best practices to help us support collaborative editing of DCPs. Thus, our research contributes towards understanding how best we can use some peculiarities in FOSS development process to help us develop and sustain a community of disaster response curators on a wiki platform.

The rest of the paper is structured as follows. In section 2, we briefly describe the information services platform community, profile the users or stakeholders, and present use case scenarios for the disaster response community. The architecture of the wiki platform is discussed in section 3. The model for data curation and collaborative editing of wiki pages is described in section 4. In the implementation part of our research, in section 5, we show screenshots of the wiki platform and describe the MediaWiki customizations and applications we have developed for the disaster response research community. Our experience report, highlighting best-practice guidelines for wiki platform designers and managers, is presented in section 6. In section 7, we summarize and conclude our research, and list future work we are undertaking.

2. THE ISP COMMUNITY

The information services platform (ISP) laboratory[1] benefits from a number of Virtual Machines (VM) servers, High Performance Computing (HPC) grid, and extensive networks supported by the Japan Gigabit Network (JGN-X[2]). The objective of the ISP lab is to build a website that will support service consumers and providers in their use and provision of information assets. We define an information asset as any piece of information that has value for the ISP community. This could be processed or unprocessed data, metadata, dataset, wiki page, or an amalgamation of one or all of these. The ISP community-based website caters for the needs of a broad range of e-science and data intensive science stakeholders. The community is loosely sub-divided into special interest groups that collaborate and work on topics related, but not limited, to Cyber-Physical systems, knowledge language grid applications, web archives and data citations, and disaster response information assets [42]. This research focuses on the latter group. In developing the wiki platform for this group, we evaluated five Open Source Content Management Systems: Joomla, Wordpress, Drupal, Ruby Nesta, and MediaWiki. We opted for Mediawiki, not only because it is the most installed wiki software [23], but because of its intuitiveness, flexibility and ease of customization, extensibility [35], and superiority in supporting collaborating editing of web content.

2.1 The Disaster Response Community

The disaster response community is a subcommunity [4] or group within the ISP community. The function of the disaster response information asset group is to create information assets for analyzing situations and responses relating to target disasters (e.g. typhoons, hurricanes, earthquakes, etc.) collaboratively. The user profile of the group consists of disaster response scientists, humanitarian relief organizations and volunteers, and the general public who are just interested in obtaining and understanding (shallow) disaster situations.

[1] http://www.nict.go.jp/en/univ-com/isp/index.html
[2] http://www.jgn.nict.go.jp/english/index.html

Table 1 shows ten use cases for the disaster response scientists and how we implemented the use cases of this user profile on the ISP wiki. Figure 1 depicts the relationships (dotted lines) between the use cases.

Table 1: Disaster response scientists Use Cases and their implementation.

Use Case	Implemented on ISP wiki
Create	Create new IA in the form of a DCP, upload or register a new datasets, or write a wiki page on a specific disaster.
Search	Search for wiki pages, DCPs, and datasets that are relevant to a specific disaster (e.g. Earthquake) using customized search systems.
Organize	Organize the retrieved IAs according correlations between the disaster (e.g. Earthquake) and other influences or factors (e.g. infrastructure damage, loss of life, etc.).
Browse	Browse the organized IAs (3) to understand a particular disaster situation.
Modify	Modify the organized IA (3) to improve its quality and usability.
Annotate	Annotate or comment on the quality and usability of IAs.
Provenance	Verify the provenance [8] of IAs.
Register	Register or upload data and metadata to the wiki website or save edited data curation profiles, wiki pages for future use.
Browse	Browse recently changed DCPs, DCPs with most citations, and most cited datasets to understand disaster situation.
Discuss	Discuss or talk about disaster response in a particular DCP page, make feature request, suggest improvements/enhancements, give pointers to data sources,etc.

- IA = Information Asset, DCP = Data Curation Profile

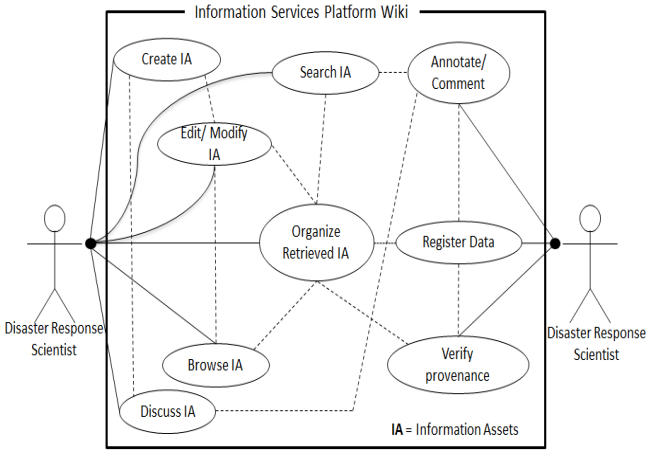

Figure 1: Relationships between Use Cases.

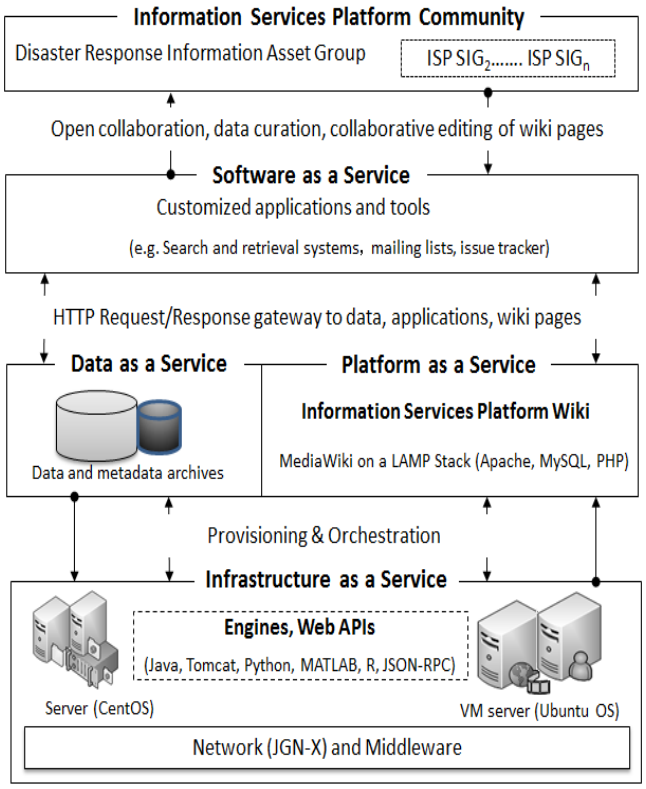

Figure 2: The Architecture of the wiki cloud platform.

3. THE WIKI PLATFORM ARCHITECTURE

In designing the architecture, we aim to make it easy for the disaster response community to curate data, and collaboratively edit wiki pages without them needing to worry about the underlying technical details or location of the data or services we offer. Furthermore, the community should be able to access our distributed data and services anytime, any-where, so long as they have access to a computer with a Web browser and Internet connectivity [20].

The wiki platform architecture shown in Figure 2 exhibits typical cloud service models [15, 42]; commonly referred to as infrastructure as a service (IaaS), software as a service (SaaS), and platform as a service (PaaS). In addition, we also offer data as a service (DaaS). Describing DaaS as a type of cloud computing service that provide data on demand, [44] pointed out that DaaS users can either download or query data from different data assets (through APIs), and they do not need to fetch and store giant data assets and search for the required information in the data asset. On top of the architecture sits the ISP community, with a specific focus on the disaster response information asset group. This subcommunity or group uses the infrastructure for open collaboration, data curation, and collaborative editing of DCP wiki pages. In the near future we plan to "plug into" the ISP community other special interest groups or ISP SIGs. We summarize how the cloud service models are loosely mapped to our wiki platform infrastructure.

3.1 Software as a Service - SaaS

This layer consists of MediaWiki customizations (see Sec-

tion 5) and distributed web-based application we have developed. These are accessible over the internet and made available as services on the wiki platform. The applications or software currently available to the disaster response community as service (SaaS) are the following:

- data and metadata search and retrieval systems for users to search for data and metadata for their research,
- a data citation tool that allow users to make citations and append the results to their wiki pages,
- data, metadata, and datasets registry or upload and download too,
- data curation tool for curating data. Users curate data by combining two or more datasets (e.g. Temperature and rain data) and import their curation results back to their wiki page.

Furthermore, platform participants can discuss their wiki pages using the customized wiki âĂIJTalkâĂİ page. More collaboration and coordination tools such as forums, mailing lists, and issue trackers will be made available to the community.

3.2 Platform/Data as a Service - PaaS/DaaS

This layer consists of the information services platform (ISP) Wiki installed on MediaWiki (version 1.20.2) and data and metadata archives. The two provide the environment for data curation and the development of wiki contents in the field of disaster response. The HTTP Request/Response gateway between the ISP wiki and the SaaS layer gives the disaster response community on-demand access to resources, data and metadata, applications, visualizations, data curation profiles, and templates of wiki pages. With our data as a service (DaaS) offering, the disaster response community have on demand access to data in the form of HTML, XML, audio/video, text, CSV, etc.

3.3 Infrastructure as a Service - IaaS

Our supporting IT infrastructure consists of the JGN-X network node connected to our research lab, physical and virtual machine (VM) servers with their respective operating systems (CentOS 6 on one of the physical machines and Ubuntu 12.04 OS on another virtual machine). A scalable and high-performance virtualization VMWare ESX Server allows us to install and maintain multiple applications on a single physical platform. Physical machines are sandwiched between firewalls and linked directly to the JGN-X network. On top of the CentOS6 is the engine that handles visualization, simulation, processing and other service requests to the data and metadata stored in archives databases.

4. THE DATA CURATION MODEL

The data curation and collaborative editing of wiki pages model describes the FOSS community governance mechanisms [40, 7, 28] or processes we adopted to support the disaster response community in their activities. The model also describes how the community can create their products, while at the same time being able to request data from a cloud storage, and append or link that data to their wiki pages. The model, shown in Figure 3, consists of the three interrelated layers described below:

4.1 Process layer

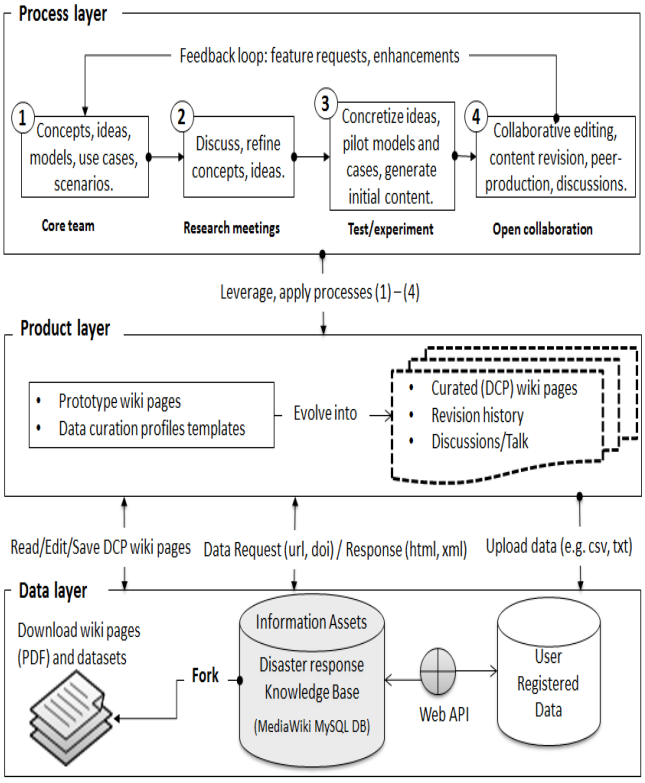

Figure 3: A model for data curation and collaborative editing of wiki pages.

This layer exemplifies our R&D motivation to develop the wiki platform for the disaster response scientific community. The community governance mechanisms in this layer are analogous to the *scratching an itch* metaphor [33], and the way core teams operate in FOSS projects [27, 7, 28]. The concept or idea is spearheaded by a small core team of ISP researchers.

In consultation with a team of experts in the domain of disaster response, the *core team* developed models, use case scenarios, prototypes of disaster response data curation wiki pages, and data curation profiles (1). Concepts and ideas were further discussed and refined during weekly *research meeting* (2). This generated consensus and gave the core team clear understanding of the tasks ahead. The resulting use cases, prototyped wiki pages, and data curation profiles were then made available on the platform for testing and experimentation (3). A small number of testers we called the "seed community" were involved in the testing and experimentation phase. This community generated the initial wiki contents that could then be used to solicit further contributions from a larger community of users (3). The seed community revised and discussed the prototyped contents, and collaboratively edited the sample data curation profiles (developed by the core team) (4).

The requests for improvements and enhancement filed in by the seed community provide a feedback loop in our community development process. We posit that this feedback loop was very effective because the community suddenly became very active in the collaborative editing of the resulting wiki pages and data curation profiles, after seeing their re-

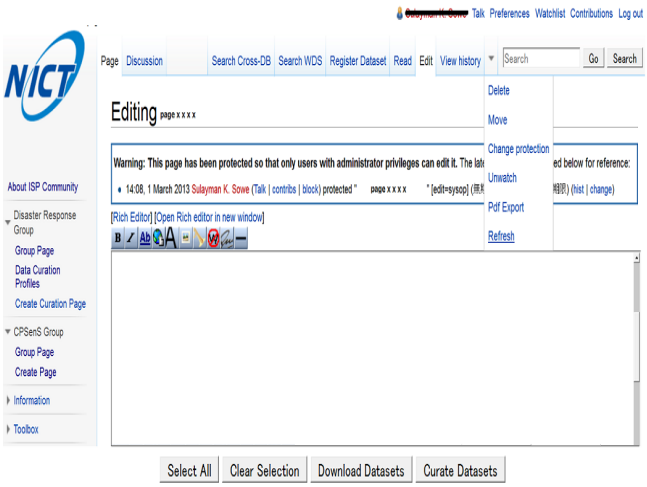

Figure 4: The wiki cloud platform showing the major customizations added to MediaWiki.

quests incorporated by the core team.

4.2 Product layer

The product layer describes the products the platform community produced collaboratively. Prototype wiki pages developed by the core team were improved upon by the seed community and evolved into full-fledge, peer-produced curated wiki pages. The community had access to a guide on data curation profile to help facilitate their collaborative editing of the sample profiles.

4.3 Data layer

The data layer contains data and metadata the wiki platform community may need for their curation activities. Custom APIs, PHP and JAVA scripts allow platform users execute request-response transaction with the data layer. In addition to registering their own data, users can also download and "own" or fork datasets from the MediaWiki MySQL database. Forking, in the positive sense of the ISP community, can be described as a situation where new products are created by cloning the existing one [37]. In FOSS, however, forking is often taken to mean starting a parallel project from the same code base [7].

5. IMPLEMENTATION

The implementation the wiki cloud platform website is shown in Figure 4. The screenshot shows the main front page and navigation menus for the group. The MediaWiki customizations ("Search Cross-DB", "Search WDS", "Register Dataset", "Download Datasets", "Curate Datasets") are shown at the top and bottom of Figure 4, respectively. With these customization tabs the group can discuss or "Talk" about their data curation profiles or research activities in the designated wiki pages, search the wiki using both the Cross-DB systems and the WDS portal, and insert the search results into the wiki pages they are editing, register or upload/download datasets for their research, and curate data.

By access the data curation tool through the "Curate Datasets" button (bottom left in Figure 4), for example, data curators can select and combine different data assets such a "Radioactivity Sensing Data" and "Twitter" messages data, and visualize the result of their combination. They can also see a map view, which is an overlay showing the features and attributes of two or more data sets that are combined. The user can then import the curation results back to his/her data curation profile without leaving the wiki platform.

Furthermore, on the front page, visitors can see (omitted from Figure 4 for simplicity) recently changed or curated profiles pages, rank of data curation profile pages with most citations, and a rank of the most cited datasets in each data curation profile.

5.1 Search Systems

The main function of the search systems is to help disaster response curators find metadata and datasets that are relevant to their research. Figure 5 shows the Cross-DB systems and the WDS portal[3] search systems as part of the MediaWiki customization process. The Cross-DB search system was developed in-house. The system uses ontological correlations based on evolutionary computing to query and retrieve the best correlated datasets for a particular query input. The World Data System or WDS portal contains datasets in the fields of social and environmental sciences that are relevant to the disaster response community. With both systems, search results can be inserted back into the wiki page being edited.

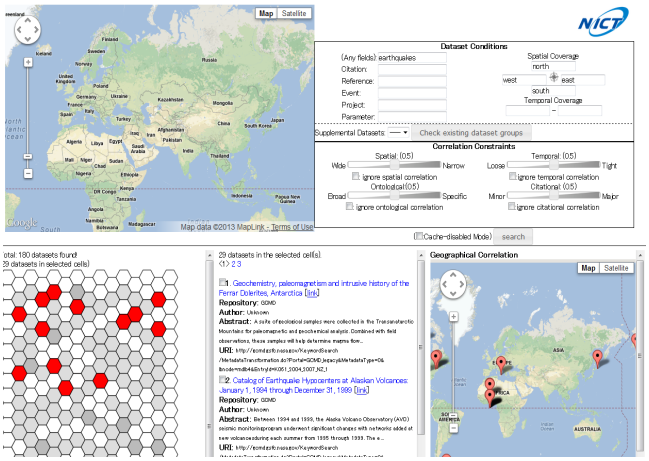

Figure 5: The Cross-DB datasets search system .

6. EXPERIENCE REPORT

During the design and implementation of the wiki platform for the disaster response research community, we experienced high as well as low points that could be best-practice guidelines for the data curation community, wiki platform designers and managers, and other systems that encourage collaborative authoring of web content. The following points highlight our experience:

- *Eliciting functional and non-functional requirements*: Seek contributions from domain experts in the main theme of your wiki platform. This could be, as in our case, organizing a mini-workshop or seminar for disaster response researchers and discussing with them the main processes, use case scenarios, data, and tools

[3] `http://www.icsu-wds.org/services/data-portal`

they may need to support their collaboration in a wiki environment. Feedback from such events can be useful in helping help you refine the functional and non-functional requirements for your wiki, as well as address some usability issues. For example, in our first workshop, participants suggested that in addition to our archived data, we should add the WDS as a data source. We also receive requests to replace the default MediaWiki WYSIWYG editor.

- *Customizing wisely*: There is no need to try and provide additional customizations when MediaWiki extensions or plugins can help you accomplish your task. Extensions are much easier to activate and deactivate than custom code. However, there is also danger of encountering dependencies or conflicts between extensions. For examples, we installed the MediaWiki Google translator extension to provide language translation for the platform community. However, when we customized the "Talk" PHP to provide a threaded discussion forum, we encountered a conflict between the two extension. To solve this issue, we ended up extending and using the NICT iPhone multi-language translator, called TextTra[4]. What to customize and for what purpose will depend on your target user group, who should be the first point of contact before you even start the customization process.
- *Community dynamics*: Get a core-team of dedicated or trusted lieutenants who will develop prototype wiki pages or data curation profiles that could be improved upon by your future community. Expand your core-team by incorporating a small group of initial testers, like our seed community. Build confidence and trust with your seed community, no matter how small or large. Encourage discussion and feedback from them. Get a wiki facilitator or "*wkilitator*" to facilitate discussion with your seed community, who might be your future collaborators.

7. SUMMARY AND CONCLUSIONS

In this paper, we presented and discussed the architecture and design of a wiki platform that supports scientific communities working in the domain of disaster response. The architecture combines the flexibility of cloud computing infrastructure with extensive customization of the generic MediaWiki core technology. We described a model for data curation and collaborative editing of wiki pages and used the FLOSS development metaphor of "scratching an itch" to help use illustrate how our disaster response community work. We further provided snapshots of the wiki cloud platform and discussed the customizations we made to generic MediaWiki core so that we are able to provide the functionalities our disaster response community needed.

In our experience report, we used practical examples to highlight some best-practices we adopted during the development of both the platform and community. We conjecture that these guidelines will kindle the debate among data curation specialists and wiki platform designers about what should and shouldn't be done to support open collaboration on wiki environments. This research sheds light, especially in the model description section, on the importance of continuously learning and using experience from FLOSS projects and communities to help us better design systems for collaborative authoring of web content.

Generalizability, transferability: According to [17], data curation practices are highly specific to individual communities and, therefore, may vary widely across communities such as disaster response, library and information sciences, and condense matter physics. However, while the context of this research is unique to our research environment and focused community (section 2), we hope that similar projects (e.g. 4C - `http://4cproject.net/`, DryadUK - `http://wiki.datadryad.org/DryadUK`, UniProt - `http://www.uniprot.org/`, the Daphnia Genomics Consortium - `https://wiki.cgb.indiana.edu`) using wiki environments to support open collaboration and data curation could benefit from our presentation. In particular, we posit that the conceptualization and relationship between the process, product, and data layers in the data curation model can be easily transferred in other contexts to support curation activities.

Threat to validity: There are number of issues relating to supporting communities involved in collaborative editing of wiki pages and/or data curation activities that we have not discussed in this research. Some of these issues include: intellectual property [38] and provenance of the registered and curated data [11], quality of the curated data and wiki pages, structure of the disaster response community or sub-communities that may emerge [4], licensing of curated or research Data [2], and how to motivate disaster response scientists to be involved in the data curation process.

7.1 Future Work

The development and implementation of the ISP wiki platform experience leaves us with future research directions. In particular, we are further exploiting the semantic wiki [22, 39] functionality using the Semantic MediaWiki (SMW) extension to improve the annotation of the data curation profiles and other wiki pages [22]. We are also improving our data and metadata harvesting techniques and harvest more diverse and heterogeneous datasets in the domains that are of interest to the disaster response community. Lastly, our proof of concept will not be complete without a comprehensive testing and evaluation strategy. We currently piloting the platform with a group of disaster response research scientists, including our original seed community.

ACKNOWLEDGMENT

The authors wish to acknowledge Mr. Yuhei Akahoshi, technical expert with the Information Services Platform Laboratory of NICT, for his valuable contribution to the Wiki Cloud Platform architecture and the development of some of the customized applications. The authors also wish to extend sincere thanks to the four anonymous reviewers for their comprehensive comments that have helped use to improved the earlier version of this paper.

8. REFERENCES

[1] J. Antin, E. H. Chi, J. Howison, S. Paul, A. Shaw, and J. Yew. Apples to oranges?: comparing across studies of open collaboration/peer production. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration*, WikiSym '11, pages 227–228, New York, USA, 2011. ACM.

[4] `http://www.nict.go.jp/en/press/2010/08/05-1.html`

[2] A. Ball. How to license research dataAf. dcc how-to guides. edinburgh: Digital curation centre. http://www.dcc.ac.uk/resources/how-guides, 2012.

[3] Z. A. Bhatti, S. Baile, and H. M. Yasin. The success of corporate wiki systems: an end user perspective. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration*, WikiSym '11, pages 134–143, New York, USA, 2011. ACM.

[4] C. Bird, D. Pattison, R. D'Souza, V. Filkov, and P. Devanbu. Latent social structure in open source projects. In *Proceedings of the 16th ACM SIGSOFT International Symposium on Foundations of software engineering*, SIGSOFT '08/FSE-16, pages 24–35, New York, NY, USA, 2008. ACM.

[5] P. Buneman, J. Cheney, S. Lindley, and H. Müller. Dbwiki: a structured wiki for curated data and collaborative data management. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of data*, SIGMOD '11, pages 1335–1338, New York, USA, 2011. ACM.

[6] S. Choudury. Data curation: An ecological perspective. *College & Research Libraries News*, 71(4):194–196, April 2010.

[7] K. Crowston, K. Wei, J. Howison, and A. Wiggins. Free/libre open-source software development: What we know and what we do not know. *ACM Comput. Surv.*, 44(2):7:1–7:35, Mar. 2008.

[8] S. Davies. Still building the memex. *Magazine, Communications of the ACM*, 54(2):80–88, 2011.

[9] L. De Alfaro, A. Kulshreshtha, I. Pye, and B. T. Adler. Reputation systems for open collaboration. *Commun. ACM*, 54(8):81–87, Aug. 2011.

[10] A. Désilets, L. Gonzalez, S. Paquet, and M. Stojanovic. Translation the wiki way. In *Proceedings of the 2006 international symposium on Wikis*, WikiSym '06, pages 19–32, New York, USA, 2006. ACM.

[11] L. Di, Y. Shao, and L. Kang. Implementation of geospatial data provenance in a web service workflow environment with iso 19115 and iso 19115-2 lineage model. *Geoscience and Remote Sensing, IEEE Transactions on*, PP(99):1–1, 2013.

[12] U. Egli and P. Sommerlad. Experience report - wiki for law firms. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration*, WikiSym '09, pages 1–4, New York, NY, USA, 2009. ACM.

[13] K. Ferguson-Boucher. Cloud computing: A records and information management perspective. *Security Privacy, IEEE*, 9(6):63–66, Nov.-Dec. 2011.

[14] D. G. Gregg. Designing for collective intelligence. *Commun. ACM*, 53(4):134–138, Apr. 2010.

[15] B. Grobauer, T. Walloschek, and E. Stocker. Understanding cloud computing vulnerabilities. *Security Privacy, IEEE*, 9(2):50–57, March-April 2011.

[16] B. Hanrahan, G. Bouchard, G. Convertino, T. Weksteen, N. Kong, C. Archambeau, and E. H. Chi. Mail2wiki: low-cost sharing and early curation from email to wikis. In *Proceedings of the 5th International Conference on Communities and Technologies*, C&T '11, pages 98–107, New York, USA, 2011. ACM.

[17] C. C. Hinnant, B. Stvilia, S. Wu, A. Worrall, K. Burnett, G. Burnett, M. M. Kazmer, and P. F. Marty. Data curation in scientific teams: an exploratory study of condensed matter physics at a national science lab. In *Proceedings of the 2012 iConference*, iConference '12, pages 498–500, New York, NY, USA, 2012. ACM.

[18] C.-L. Ignat, G. Oster, P. Molli, M. Cart, J. Ferrie, A. M. Kermarrec, P. Sutra, M. Shapiro, L. Benmouffok, J.-M. Busca, and R. Guerraoui. A comparison of optimistic approaches to collaborative editing of wiki pages. In *Collaborative Computing: Networking, Applications and Worksharing, 2007. CollaborateCom 2007. International Conference on*, pages 474–483, Nov.

[19] L. Y. Karen S. Baker. Data stewardship: Environmental data curation and a web-of-repositories. *International Journal of Digital Curation*, 4(2):12–27, 2009.

[20] C. Kiddle, A. R. Taylor, J. Cordes, O. Eymere, V. Kaspi, D. Pigat, E. Rosolowsky, I. Stairs, and A. G. Willis. Cyberska: an on-line collaborative portal for data-intensive radio astronomy. In *Proceedings of the 2011 ACM workshop on Gateway computing environments*, pages 65–72, New York, USA, 2011.

[21] A. Kittur and R. E. Kraut. Harnessing the wisdom of crowds in wikipedia: quality through coordination. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, CSCW '08, pages 37–46, New York, USA, 2008. ACM.

[22] M. Krötzsch, D. Vrandečić, M. Völkel, H. Haller, and R. Studer. Semantic wikipedia. *Journal of Web Semantics*, 5(4):251–261, 2007.

[23] R. M. Lerner. Installing and customizing mediawiki. *Linux J.*, 2006(144):3–, Apr. 2006.

[24] A. Lund and O. Smørdal. Is there a space for the teacher in a wiki? In *Proceedings of the 2006 international symposium on Wikis*, WikiSym '06, pages 37–46, New York, USA, 2006. ACM.

[25] M. Memmel, M. Wolpers, and E. Tomadaki. An approach to enable collective intelligence in digital repositories. In J. Luca and E. R. Weippl, editors, *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2008*, pages 1803–1811, Vienna, Austria, June 2008. AACE.

[26] D. S. B. M. H. C. Michael Witt, Jacob Carlson. Constructing data curation profiles. *International Journal of Digital Curation*, 4(3):93–103, 2009.

[27] A. Mockus, R. T. Fielding, and J. D. Herbsleb. Two case studies of open source software development: Apache and mozilla. *ACM Trans. Softw. Eng. Methodol.*, 11(3):309–346, July 2002.

[28] K. Nemoto, P. Gloor, and R. Laubacher. Social capital increases efficiency of collaboration among wikipedia editors. In *Proceedings of the 22nd ACM conference on Hypertext and hypermedia*, HT '11, pages 231–240, New York, NY, USA, 2011. ACM.

[29] F. Orlandi and A. Passant. Semantic search on heterogeneous wiki systems. In *Proceedings of the 6th International Symposium on Wikis and Open Collaboration*, WikiSym '10, pages 4:1–4:10, New

York, USA, 2010. ACM.

[30] K. Panciera, M. Masli, and L. Terveen. "how should i go from - to - without getting killed?": motivation and benefits in open collaboration. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration*, WikiSym '11, pages 183–192, New York, USA, 2011. ACM.

[31] L. L. D. G. Philip Lord, Alison Macdonald. From data deluge to data curation. Technical report, JISC (the Joint Information Systems Committee) and the UKĄfs e-Science Core Programme, 2003.

[32] P. K. Raj and K. Srinivasa. Empirical studies of global volunteer collaboration in the development of free and open source software: analysis of six top ranked projects in sourceforge.net. *SIGSOFT Softw. Eng. Notes*, 37(2):1–11, Apr. 2012.

[33] E. S. Raymond. *The Cathedral and the Bazaar*. OĄfReilly & Associates, Inc., Sebastopol, CA, USA., 1st edition, 1999.

[34] A. J. Reinoso, F. Ortega, J. M. González-Barahona, and I. Herraiz. A statistical approach to the impact of featured articles in wikipedia. In *KEOD*, pages 420–423, 2010.

[35] J. Reyes. How to evaluate what cms to use, Nov 24 2009.

[36] D. Riehle, J. Ellenberger, T. Menahem, B. Mikhailovski, Y. Natchetoi, B. Naveh, and T. Odenwald. Open collaboration within corporations using software forges. *Software, IEEE*, 26(2):52–58, 2009.

[37] J. Rubin, A. Kirshin, G. Botterweck, and M. Chechik. Managing forked product variants. In *Proceedings of the 16th International Software Product Line Conference - Volume 1*, pages 156–160, New York, USA, 2012. ACM.

[38] C. D. Santos, Jr., M. B. Cavalca, F. Kon, J. Singer, V. Ritter, D. Regina, and T. Tsujimoto. Intellectual property policy and attractiveness: a longitudinal study of free and open source software projects. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, CSCW '11, pages 705–708, New York, NY, USA, 2011. ACM.

[39] S. Schaffert, F. Bry, J. Baumeister, and M. Kiesel. Semantic wikis. *Software, IEEE*, 25(4):8–11, 2008.

[40] S. K. Sowe, I. Stamelos, and L. Angelis. Understanding knowledge sharing activities in free/open source software projects: An empirical study. *Journal of Systems and Software*, 81(3):431–446, March 2008.

[41] S. K. Sowe, I. G. Stamelos, and I. M. Samoladas, editors. *Emerging Free and Open Source Software Practices*. IGI Global, Covent Garden, London., 2008.

[42] S. K. Sowe, K. Zettsu, and Y. Murakami. A model for creating and sustaining information services platform communities: Lessons learnt from open source software. In *The 5th ITU Kaleidoscope conference on Building Sustainable Communities, Kyoto, Japan, 22-25 April 2013*, pages 13–20, Kyoto University, Japan, 2013. ITU Telecommunication Standardization Sector.

[43] B. Tansey and E. Stroulia. Annoki: a mediawiki-based collaboration platform. In *Proceedings of the 1st Workshop on Web 2.0 for Software Engineering*, pages 31–36, New York, USA, 2010. ACM.

[44] Q. H. Vu, T.-V. Pham, H.-L. Truong, S. Dustdar, and R. Asal. Demods: A description model for data-as-a-service. In *Advanced Information Networking and Applications (AINA), 2012 IEEE 26th International Conference on*, pages 605–612, March 2012.

[45] D. Yates and S. Paquette. Emergency knowledge management and social media technologies: a case study of the 2010 haitian earthquake. In *Proceedings of the 73rd ASIS&T Annual Meeting on Navigating Streams in an Information Ecosystem - Volume 47*, pages 1–9, Silver Springs, MD, USA, 2010. American Society for Information Science.

[46] L. Zhu, I. Vaghi, and B. R. Barricelli. A meta-reflective wiki for collaborative design. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration*, WikiSym '11, pages 53–62, New York, USA, 2011. ACM.