Wiki as a Platform – Turning Dissemination into Collaboration

Craig S. Evans[1,2]

[1] Illinois Informatics Institute, University of Illinois at Urbana-Champaign

[2] Graduate School of Information and Library Science, University of Illinois at Urbana-Champaign

Abstract

In research projects, data collection and dissemination are considered as two discrete and independent activities. The focus is on the research question, and not on how to best collect, present and subsequently share data. Although most US funding agencies now require that researchers data share, the tools available to operationalize this requirement are lacking. We propose show how the open source MediaWiki system can provides a lightweight, collaborative, and inexpensive tool to support new data sharing practices. This note serves to illustrate how interactive data collection and dissemination supported by a Wiki server can be used by scientists both during the project and for subsequent dissemination.

Keywords: data sharing, information reuse, collaborative information systems, scholarly communications

Citation: Evans, C. S. (2014). Wiki as a Platform - Turning Dissemination into Collaboration. In *iConference 2014 Proceedings* (p. 820–826). doi:10.9776/14404

Acknowledgements: Project completed as part of Preserving Virtual Worlds 2 (PVW2). PVW2 was made possible through National Leadership Grant #LG-06-10-0160 from the Institute of Museum & Library Services. Thanks to Dr's Jerome McDonough (PVW2), Michael Twidale (Academic Advisor), and Catherine Blake for academic feedback, and editorial advice.

Contact: mailto:csevans2@illinois.edu

1 Introduction

Research does not get performed in a bubble. At some point, the findings and results of the research need to be published and made accessible. This is a basic tenet of the 'paradigms' or 'normal science' as defined by Kuhn (Kuhn, 2012). The traditional mechanism to achieve this goal is a published scientific article. As a distribution mechanism the journal, conference proceeding, and book have been the primary means of distributing knowledge. However, changes in scholarly practices such as FORCE11[1] (Bourne et al., 2012) which advocates new ways to work on an article, and nanopublishing (Sofronijević & Pavlović, 2013) are changing the way that scientists disseminate information after a project is complete.

In this paper we introduce a way to leverage the same platform used during the collaborative research process to concurrently create a collection suitable for external publishing.

A common requirement for grant funding agencies is that a "proposal budget may request funds for the costs of documenting, preparing, publishing or otherwise making available to others the findings and products of the work conducted under the grant." (National_Science_Foundation, 2013). For many projects this is a straightforward process, and can utilize an institutions online presence such as the IDEALS[2] centralized storage system, publish through a personal, project, or departmental webpage, or disseminate through journals and conference proceedings. However for some projects, a more interactive means for distributing data is required and that often takes the form of a Wiki.

While this is appropriate for disseminating results, it doesn't capture the growing need for interactive collection, analysis, and distribution of information during a project, and after a project it is complete.

---

[1] www.force11.org - "a community of scholars, librarians, archivists, publishers and research funders that has arisen organically to help facilitate the change toward improved knowledge creation and sharing. Individually and collectively, we aim to bring about a change in modern scholarly communications through the effective use of information technology."

[2] IDEALS – https://ideals.illinois.edu

Collaborative research is not new. Harrison and Dourish (Harrison & Dourish, 1996) make a distinction between space (the physical location where work occurs) and place (where work is done). Traditionally scientific collaboration required collocation (space), and by using well-defined or ad-hoc tools (place). Twidale and Nichols (Twidale & Nichols, 1996) showed that while collaborative systems exist that allow people to access common information sources, the designers of these systems do everything they can to make it feel like the user is using an individual resource. Scientists do work well together, but as reported by Blake (Blake & Pratt, 2006a) the process of collaborating would be enhanced with tools that allowed better sharing of information. Studies in Collaborative Information Seeking (Karunakaran & Reddy, 2012) as well as Blake's Collaborative Information Synthesis study have shown that there is a focus on sharing of documents and the integration of extracted facts, but that is all. It is proposed that collaboration should extend beyond documents, extracted data, and annotations to data collection and other activities that allow users to work together at different workplaces.

In the current research model, funding is predominantly for a fixed period of time, and while data management plans are increasingly asking for how data and research will be distributed and archived post-project, this is not the focus area of a project proposal and not an area where limited resources are typically allocated.

Adding complexity is the collaborative nature of projects, involving research from multiple public and private institutions, and the individual involvement of non-affiliated members of the general public. People from outside the Principle Investigators institution may not only be authoring or commenting on research in progress, but might be actively contributing edits, data, or annotations to the research. While ad-hoc tools are often brought together to facilitate collaboration, it is often as an after thought, and will often be a mix of online or cloud based solutions such as a Wiki or Google docs. The Wiki is used to share information in a consumptive manner, while the Google docs are used to collaboratively edit.

1.1 Motivating Example

The Preserving Virtual Worlds 2 Project[3] (McDonough et al., 2010) was a funded research initiative with a defined timeline. The outcomes from the project included a survey tool to continue gathering information, as well as an ability to disseminate information. The general structure of the data collection is along the lines of a survey. The obvious choice is to use a survey engine such as Survey Monkey … but who collects the data for analysis in the future? Ongoing maintenance of a system is not uncommon after a projects funding is complete. So how can this be achieved?

While the PVW2 project motivated this project, the Wiki as a platform concept was a very small part of the research and served the purpose of survey data collection only. While building the system, the author realized the potential for other research projects and has started to explore the use of Wikis as a collaborative dissemination mechanism.

2 Description

The use of Wikis in research is commonplace, and allows authors to publish information while allowing users to search and browse information and allows collaborative editing of text documents, as well as a modicum of discussion. However, this is largely a publication model, with little interaction with the data.

We needed a solution that was simple, lightweight, had low ongoing maintenance overhead, was secure, and could be sustained without the need for programming resources.

To satisfy these goals, along with the need for post project longevity, we turned to open source software that could be customized with minimal effort.

[3] PVW2 focuses on determining significant properties for a variety of educational games and game franchises in order to provide a set of best practices for preserving the materials through virtualization technologies and migration, as well as provide an analysis of how the preservation process is documented.

The platform we based our solution on was MediaWiki[4] – the same Wiki engine used by Wikipedia.

This platform was chosen because it presented a number of immediate benefits.

- **Lightweight**
  MediaWiki uses a lightweight LAMP (Linux, Apache, MySQL, PHP) style application stack. For our purposes, we used Apple's OS X operating system, with its internal web server (which is Apache based), a MySQL database, and PHP at the database interface language. Internal pages were developed using the standard Wiki markup language and HTML/CSS with JavaScript for the enhanced data entry pages.
- **Low Maintenance Overhead**
  The MediaWiki server is the basis of Wikipedia, and as such has been demonstrated to scale exceptionally well. As the underlying Wikipedia server, the software is actively supported and regularly patched. This has the advantage of reducing the burden of requiring a software engineer to maintain and upgrade code. Content is displayed using a combination of standard Wiki markup language as well as HTML/CSS and JavaScript via common plugins.
- **Secure**
  The inherent security model of MediaWiki was another attractive feature in that a dedicated systems administrator was not required for the server, and security was largely self managed by the system. Dedicated logins to internal resources were not required, with nominated users being able to grant/revoke user privileges. To enhance use in an academic setting, the user signup page can be readily customized to include wording for IRB acceptance as well.
- **Simplicity and familiarity**
  With those major functions addressed, the fourth criterion to be addressed was simplicity. The advantages of this approach were twofold:
  1. The navigation model was familiar to the user community;
  2. The Wiki markup schema is well established and well documented.

  Previous solutions were over engineered and added a level of complexity to the task of data collection that were inappropriate for the task and the target audience.
- **Extendibility**
  While not a standard feature, a readily available plugin for MediaWiki is an ability to use HTML and CSS within the pages. This allowed us to create flexible and dynamic survey forms that would normally require a webserver to host, and access to the file structure to maintain. By embedding HTML into MediaWiki, we are now able to update survey pages without the need to have backend server access.

  To create an interactive server that was more than a publishing platform we also installed the following extensions:
  - Cite and SpecialCite – enhanced citation handling
  - SecureHTML – allows embedded HTML/CSS in wiki page
  - Vector – Adds the familiar Wikipedia UI
  - WikiEditor – Enhanced word processor style page editing
- **Cheap**
  While listed last, with limited resources in a budget, this is a major advantage. A Wiki based system is generally open source[5]. The system will run on low cost servers utilizing a LAMP (Linux, Apache, MySQL, PHP) approach that is well understood (no specialized training for staff to install, maintain, and use), has readily available support (online forums from a very active user community),

[4] www.mediawiki.org

[5] Commercial Wiki systems are available, however a majority of them are open source and free for academic use.

and has regular maintenance and upgrade release (minimizing developer overhead, and increasing security through user contributed patches). This adds up to being a small line item in a budget instead of a major undertaking.

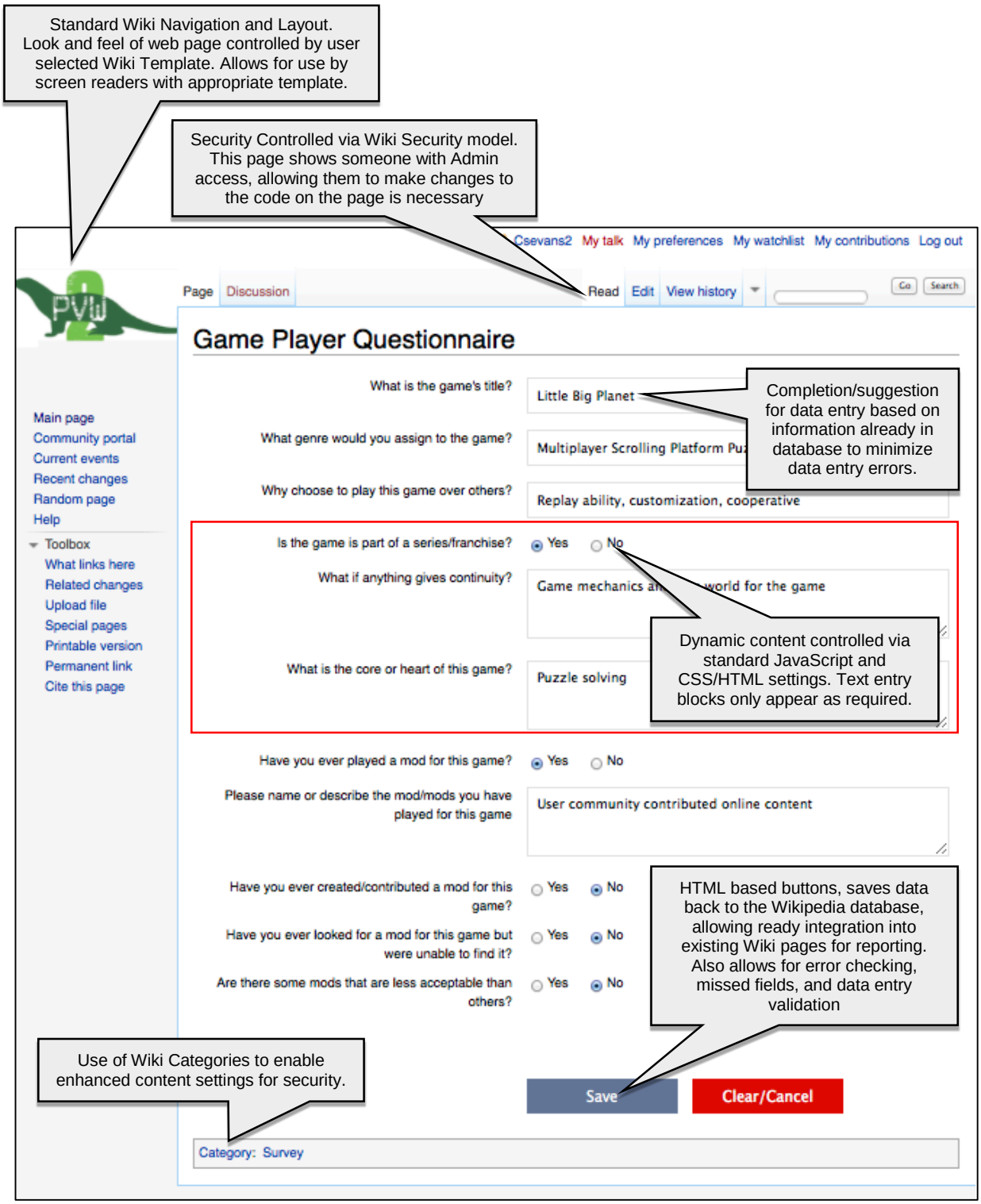

Figure 1: Example data collection form using HTML, CSS, and JavaScript

3 Discussion

The traditional use of a Wiki is that of a publishing platform that allows the content to be collaboratively edited and comments made, with an audit trail of changes. For many projects this provides an adequate level of control, and provides a means to consolidate changes within a collection of documents.

The HTML and CSS plugins for Wiki are intended for enhancing layout using standard web based techniques, but HTML is not purely for display. Adding in these plugins opens up options that are not generally considered for a Wiki based platform.

Using forms and server side scripting such as PHP, it is possible to create interactive websites. With an underlying MySQL database, forms can be written to the same database and tables that the Wiki is pulling from to display information. A further advantage is that data from the survey can write to the MediaWiki database, can dynamically pull information from the database (Wiki pages), and can use the MediaWiki interface to display data. By embedding these within a Wiki framework we now have a platform that is suitable for both the publishing and collection of data, and displaying dynamic content.

This combination of plugin and wiki server has a number of other advantages that can also suddenly be leveraged.

- **Dynamic Data Displays**
  The web has become interactive. Data is now dynamically graphed, charted, plotted and manipulated through web based interfaces using a number of techniques such as the Data Driven Documents library (d3.org). These interfaces can be readily incorporated into a Wiki page. This has the advantage that the data presented in a wiki is no longer static, but is a living entity that can be visualized by the user in ways that were not originally conceived.
- **External data sources**
  A Wiki is typically limited to the data that is contained within its database. A web based solution does not have those limitations, and can pull data from sources outside the Wiki database. While not searchable as part of the Wiki index, this information could be used to enhance the data within a wiki page to provide context or supplemental data.
- **Template Based Displays**
  The advantage of this is two fold. The wiki is a primarily a publishing platform. Content is displayed according to the deployment of templates. This template driven approach has a number of advantages.
  1. **Mobile Sites**: By detecting the browser, different template options can be loaded. This allows for the same content to be displayed in formats that are appropriate to both desktop style clients as well as mobile clients. With the relevant HTML embedded within the template code for a Wiki page, information can be displayed to mobile devices in a form that is more appropriate to that platform, while desktop computers can have a richer presentation.
  2. **Accessibility**: By deploying a template that adheres to accessibility standards, those with visual or mobility disabilities will be able to access and participate using the same underlying data as others within a project.

  By allowing a user to choose the appropriate template for their usage needs (standard, mobile, accessible) the system can be used by a wider audience.
- **Data Verification and Quality Control**
  Surprisingly, a standard Wiki installation has little in the way of form controls. By using a security models we can limit who can enter data (verified project team members), who can edit data (verified project team members), who can comment on the data (users with an account), and who can read the data (anyone). Data entered and made visible in this form allows for a peer review and

verification, and importantly, because changes are logged, a edit history of changes is also maintained.

It should also be noted that beyond peer review, simpler systems such as using Javascript to perform edit checks prior to saving data, or Ajax to look up a field in a database and autocomplete a response (minimizing transcription errors) are also available once we incorporate an interactive interface into the Wiki system.

4 Future Work

Grudin's Eight Challenges for Groupware Developers (Grudin, 1994) posited that there was a "disparity in work and benefit … often requir(ing) additional work from individuals who do not perceive a direct benefit from the use of the application." At the moment the people who are doing the work are the scientists, and the people who get the reward are scientists in other groups. By repurpose existing software in order to support collaboration during the project and subsequent dissemination, those creating the content get the benefits during the project, and the results are already in a form suitable for public dissemination.

While the PVW2 project is completed, the Wiki as a Platform concept has been included in a number of upcoming research initiatives. We are hoping that by using the MediaWiki system in combination with the described plugins, we will have a platform that we can further tailor with minimal effort to not only act as an interactive portal for the collection of data, and the presentation and explorations of data, but will also provide a platform to gauge usage patterns, and ascertain which pages are the most important to different types of users.

The projects under review will make use of the Data Driven Documents paradigm, creating visualizations that will help scientists to explore the data they are contributing, and will have a template driven interface so that they display appropriately on mobiles devices as well as desktop computers without the need for a dedicated mobile client.

5 Conclusion

Collaboration should mean working together. The location is immaterial provided that the tools and resources at hand allow people to contribute to the common project, comment on each others work, and share results with both those in their project as well as disseminate the findings to a wider audience.

The use of the Wiki framework presented also has the advantage of allowing scientists multiple views of the same data, and a level of transparency in data entry that could also help to ensure data quality. Streamlining the process and integrating the research and publication will make it easier for scientists to collaborate during a project and disseminate their work after.

6 References

Blake, Catherine, & Pratt, Wanda. (2006a). Collaborative information synthesis I: A model of information behaviors of scientists in medicine and public health. *Journal of the American Society for Information Science and Technology, 57*(13), 1740-1749.

Blake, Catherine, & Pratt, Wanda. (2006b). Collaborative information synthesis II: Recommendations for information systems to support synthesis activities. *Journal of the American Society for Information Science and Technology, 57*(14), 1888-1895.

Bourne, PE, Clark, T, Dale, R, de Waard, A, Herman, I, Hovy, E, & Shotton, D. (2012). Improving Future Research Communication and e-Scholarship. *Force 11 Manifesto.*

Grudin, Jonathan. (1994). Groupware and social dynamics: eight challenges for developers. *Communications of the ACM, 37*(1), 92-105.

Harrison, Steve, & Dourish, Paul. (1996). *Re-place-ing space: the roles of place and space in collaborative systems.* Paper presented at the Proceedings of the 1996 ACM conference on Computer supported

cooperative work, Boston, Massachusetts, United States. http://dl.acm.org/citation.cfm?doid=240080.240193

Karunakaran, Arvind, & Reddy, Madhu. (2012). *The role of narratives in collaborative information seeking*. Paper presented at the Proceedings of the 17th ACM international conference on Supporting group work, Sanibel Island, Florida, USA.

Kuhn, Thomas S. (2012). *The structure of scientific revolutions*: University of Chicago press.

McDonough, Jerome, Olendorf, Robert, Kirschenbaum, Matthew, Kraus, Kari M, Reside, Doug, Donahue, Rachel, . . . Rojo, Susan. (2010). Preserving virtual worlds final report.

National Science Foundation. (2013). *Proposal and Award Policies and Procedures Guide: Part I - Grant Proposal Guide*. (OMB Control Number: 3145-0058). Retrieved from http://www.nsf.gov/publications/pub_summ.jsp?ods_key=gpg.

Sofronijević, Adam, & Pavlović, Aleksandra. (2013). *Applicability of the nano-publication concept for fostering Open Access in developing and transition countries*.

Twidale, Michael, & Nichols, David. (1996). *Collaborative browsing and visualisation of the search process*. Paper presented at the Aslib Proceedings.

7 Table of Figures