Open Research Online

The afterlife of 'living deliverables': angels or zombies?

Fridolin Wild, Thomas Ullmann

Knowledge Media Institute, The Open University, UK
{f.wild, t.ullmann}@open.ac.uk

Abstract: Within the STELLAR project, we provide the possibility to use living documents for the collaborative writing work on deliverables. Compared to 'normal' deliverables, 'living' deliverables come into existence much earlier than their delivery deadline and are expected to 'live on' after their official delivery to the European Commission. They are expected to foster collaboration. Within this contribution we investigate, how these deliverables have been used over the first 16 months of the project. We therefore propose a set of new analysis methods facilitating social network analysis on publicly available revision history data. With this instrumentarium, we critically look at whether the living deliverables have been successfully used for collaboration and whether their 'afterlife' beyond the contractual deadline had turned them into 'zombies' (still visible, but no or little live editing activities). The results show that the observed deliverables show signs of life, but often in connection with a topical change and in conjunction with changes in the pattern of collaboration.

Keywords: deliverables, wiki, collaboration, analysis, visualisation, #stellarnet

1 Introduction

In standard project management jargon, a 'deliverable' refers to a pre-defined, tangible, and verifiable work product such as a feasibility study or a prototype [1]. In research projects, deliverables often document process and outcomes of (more or less) systematic knowledge creation. They report on the progress against the tasks expected to be 'delivered' during a defined phase of the project. These documents sum up the focused work of a group or single person.

Within the STELLAR project, we provide the possibility to use living documents for the collaborative writing work on deliverables. They can be continuously updated and revised by all authors, even in parallel, using the popular wiki software MediaWiki (the software on which Wikipedia is based). Compared to 'normal' deliverables, 'living' deliverables come into existence much earlier than their delivery deadline and are expected to 'live on' after their official delivery to the European Commission. They are expected to foster collaboration in writing. Within this contribution we investigate, how these deliverables have been used over the first 16 months of the project. We will critically look at whether they have been successfully used for collaboration and whether their 'afterlife' beyond the contractual deadline had turned them into a 'zombie' (arguably still some sort of life, but not a really

welcome one). A zombie can still be seen, but does not show any signs of vital activity, whereas an angel cheerfully continues editing activities – but with the difference of being relieved from the duty of the mortal to deliver. It is clear that deadlines are typically drivers of activity, so also for angels, afterlife activity should be visibly less hectic and might focus on new or different areas of editing activity.

The analysis of the dynamics of wikis and their flagship Wikipedia is naturally a relatively young research field, since Wikipedia was created only back in 2001 – thereby making available a large public data-set of revision histories. Viegas et al. propose a method called 'history flows' for analysing the social dynamics expressed in the editing of Wikipedia articles [4]. They analyse the relationship between document revisions revealing cooperation and conflict patterns. Nunes et al. [3] use the revision history to visualize revision activity through sparklines in a timeline plot within their system 'WikiChanges', additionally supported by a 'tag-cloud'-like visualisation of term changes in the time frame selected (the font size is scaled by their changed frequency within the time window inspected). Arazy et al. [2] develop a series of glyphs to visualise contribution scores of authors in pages in order to ease the recognition of their work. Suh et al. [5] focus on identifying patterns of conflict with the help of so-called 'revert graphs', visualising the relation between authors of Wikipedia established through revisions that void previous edits. Baumgrass et al. [6] apply social network analysis in order to investigate corporate knowledge exchange processes in wikis. Closely related is also the work of Jesus et al. [7], within which network analysis is applied to study cluster-level collaboration between authors grouped by their work on related articles. Whereas [2,3,4,5] focus on the analysis of collaboration in individual pages, [6] and [7] deploy the same analytical technique – (social) network analysis –, but with a different focus of analysis [7] and in a different cultural and application setting [6].

All of them, however, share with our work the interest to shed light on the authorship relations documented in the revision histories. The user interface of the wikis is designed in a way, which centres the article and not so much the contributions of the single authors: its focus is on content and not authorship [2]. Making the authorship relation visible means extracting the relevant data from the revision histories of the pages and providing an easy to understand view of this data.

While a deliverable is the result of the edits of all authors, the revision history retains information about the contribution of each individual. This makes it easy to spot latest edits or compare changes with previous ones. It helps to keep track of the development of the pages contained in the living deliverable and, for example, make it easy to revert edits.

There are many ways of how to represent writing activity and collaboration of wiki pages. Within the rest of this paper, we first elaborate on our method of analysis used to make the collaborative writing process of living deliverables visible. With this, we analyse the data gathered within the STELLAR project so far: we visualize the overall co-authorship network; we outline the revision frequency over time to investigate if the living deliverables are indeed living; and we show how the collaboration network of authors and their contributions changes before and after a deadline. Finally, we conclude the paper with a summary and an outlook.

3 The data: Stellar's 'living deliverables'

The observed dataset consists of five living deliverables. They have been selected from the set of 14 wikis created so far for 19 project deliverables by excluding 'obvious zombies' and 'small group wikis' such as the coordination manual. Obvious zombies thereby relate to those wikis for which the group of collaborators did not use the offered wiki or abandoned it early in the writing process favouring different solutions to organise collaborative writing: these were mainly google docs and in several cases the exchange of word and excel files via mail with one or several editors consolidating tracked changes. The latter thereby being the main method used for the five management and evaluation deliverables that are much more clerical in nature and contain a lot of spreadsheet data – a task for which MediaWikis are hard to use.

Each living deliverable resides in its own MediaWiki instance. All wikis were initialized at the beginning of each deliverable writing period. While observing the process of the living deliverable evolution, we have to consider the fact that these documents served as input for the 'normal' deliverables (the type-set word or PDF file delivered to the European Commission), and the latter could then again feed back into the living deliverables.

The following Table 1 gives an overview of each of the investigated living deliverables. Among others, it outlines the number of authors, the number of pages contained in the wiki (and their number of page views), and – most notably – the number of edits these pages have received. All in all, the deliverables had an average number of 22.7 users, with a varying number of page views (in average 3,820). Some of them have received a substantial number of edits (such as the grand challenge document d1.1 and the science 2.0 mash-up deliverable d6.3, both earlier deliverables).

	Users	Total Views	Total Pages	Total Edits	Total Images	Pages/ Users	Edits/ Users
d1.1	78	14813	78	533	4	1	6.83
d1.2	9	1338	86	137	1	9.56	15.22
d6.1	4	677	39	152	28	9.75	38
d6.2	11	712	14	79	10	1.27	7.18
d6.3	21	2818	65	333	1	3.1	15.86
d7.1	13	2563	84	354	48	6.46	27.23

Table 1. Basic statistics of the investigated wikis.

4 Method of analysis: SNA of the collaboration networks

The revision history of the living deliverables is a chronologically sorted list of changes of pages, listing – amongst others – the editing user, the page, the amount of characters changed with the revision, and a timestamp expressing when the revision was applied. One example of this revision history can be found in the snapshot of a revision history visualisation widget we have created to support the work in the deliverables (Figure 1): it shows the revision of one living deliverable in a scrollable timeline, listing the title of the changed page, the date of the change, and the name of the editor (pop-up bubble).

While this way of exploring the revision data has its benefit for following latest changes or browsing through the history of all changes, it does not provide much insight into the nature and vitality of the underlying collaboration, nor much insight into the focus of collaboration.

Collaboration is expressed in the co-authorship relations and can be extracted from the revision history. Co-authorship relations in living deliverables, however, can be investigated in many ways. The simplest form would be a list of authors of the deliverable or a page in it. List-like representations, however, do not show the structure of collaboration between the authors of the living deliverable. This extra dimension of information can provide insights into the collaboration network structure. We used a co-authorship social network analysis, which shows the relations established between authors by editing the same page. Therefore, an incident matrix was constructed listing the pages as incidents in the rows, the authors in the columns, and their number of edits of the respective page in the matrix cells. By multiplying the matrix with its transpose, an undirected affiliation matrix can be constructed and visualised as a network (see **Figure 2**).

Figure 1. Timeline widget (visualizing the revision history of D6.3).

Since the central jump page ('home') of wikis is edited very often and by almost everyone (to, e.g., add links to new sub pages), it may be excluded from analysis in order to expose the clusters of collaborating authors more clearly (see Figure 3).
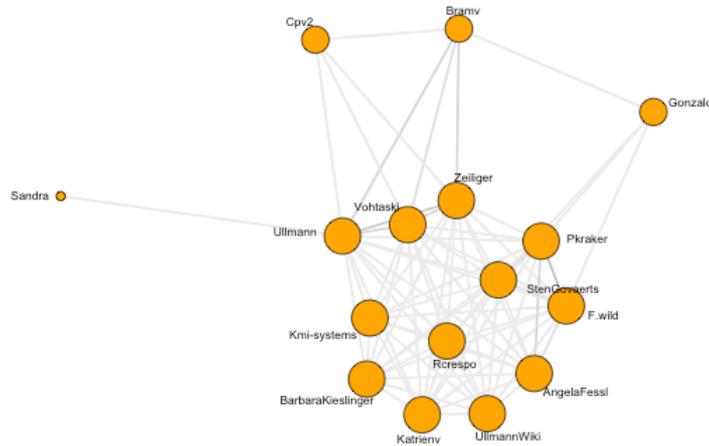
Figure 2. Collaboration network including edits of the central home page (D6.3).

The graph shows, a cluster of authors who contribute to a shared article. On the periphery of the cluster, the less connected authors are shown. By removing the central home page, two clusters can be seen, which are connected only through shared contributions of two authors. On the periphery there are four authors, who only wrote contributions to the main page or only on pages not edited by others, but not on any of the pages co-edited by the authors in the two clusters.

Figure 3. Collaboration network excluding the central home page.

This co-authorship visualisation has its benefit in showing who collaborated with whom. It does not, however, show the evolution of the living deliverable over time and it lacks information about the content on which the authors collaborated. This can

be extended by adding pages as nodes to the network and introducing directed editing relationships pointing from the authors to the pages they have changed. With that, authoring relations on particular pages become more salient.

Additionally, the development of the overall number of non-minor edits over time provides information on the vitality of the wiki and complements the analysis.

5 Discussion: Is there an afterlife after the deadline?

The deadline of regular deliverables marks the end of the writing process. After the deadline, the official writing process ends and there is no formal requirement to modify them anymore. As mentioned above, the purpose of living deliverables is to allow for more continuous collaboration beyond delivery deadlines. The assumption behind living documents is that knowledge construction processes are continuous and deliverables are artefacts of an underlying, continuous collaboration process. By turning these artefacts into living documents, they better reflect the dynamic structure of project work, which is somewhat artificially subjected to a project framework in order to allow for efficient and effective management. Not only in networks of excellence, where a consortium faces additionally the challenge to re-organise an open research network beyond the partnership, but also in other research project types, interdependencies of tasks naturally create feedback loops that should inform already 'delivered' work (such as from validation to conceptual design), thus creating an opportunity to update them.

To test whether or not the documents were subject to editing activity also after the submission deadline, we gathered the revisions of each deliverable and cumulated the amount of revisions for each deliverable for each project month. The following line chart shows on the y-axis the amount of revisions and on the x-axis the time frames (16 project months). One deliverable already exists since 13 months, while others are in use for shorter periods of time. The vertical lines at month 3, 6, 9, and 12 represent the submission deadlines.

All deliverables continue their life also after their formal deadline. Even when considering a phase of two months after the deadlines (taking into account possible delays in delivery), still three of the deliverables show lively activity. According to the revision counts, the official deadline raised the number of revisions, while after a deadline the amount of revisions increases mostly less steep. The three deliverables d6.2 (blue), d6.3 (purple), and d1.2 (yellow) show a very steady increase over time, whereas particularly the early deliverables d7.1 (orange) and d1.1 (green) experience their most busy editing processes around the time of their deadline.
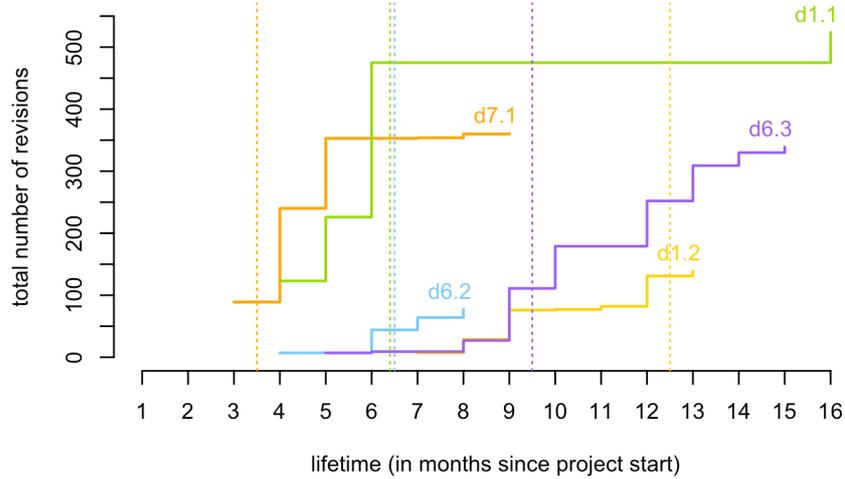
Figure 4. Total number of edits (cumulated) for each living deliverable.

While the line chart visualisation only shows the frequencies of the revisions over time, it does not provide information about the themes of collaboration and the collaboration network created in the co-editing activity – and how they have changed from before to after the deadline.
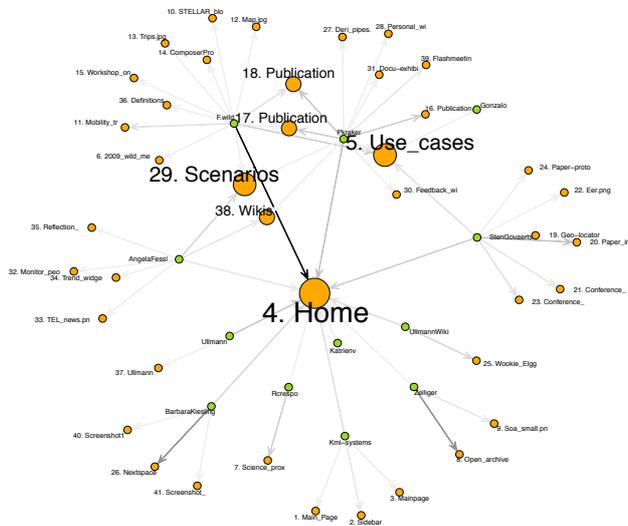

Figure 5. Authors (green) and their contributions to pages (orange):
before the submission deadline.

Figures Figure 5 and Figure 6 show the network of authors and their contributions to pages in d6.3 before and after the submission deadline. While the focus before the deadline is clearly on 'use cases', 'scenarios' and the main page of the deliverable, the

figure for the network after the deadline shows a change towards more technical topics, like 'Tools', 'Services', and 'Widgets'.

Figure 6. Authors (orange) and their contributions to pages (green): after the submission deadline.

The other deliverables show similar patterns of activity: d7.1 again exposes a larger network of pages (but with a smaller number of contributors), where as d1.1 is significantly reduced in the number of contributors (but still showing a larger number of edits). The deliverable d6.2 shows a star pattern of authors editing the main page and d1.2 ceased its activity with its delivery deadline.

6 Conclusion and outlook

With the analysis presented, the conclusion can be drawn that there definitely is an afterlife for most of the living deliverables. With only one zombie exception, this afterlife is more like a blitheful continuation of activities – relieved of the duty of having a deadline. At least for the one deliverables we have analysed this in more depth and collaboration beyond the deadline exposes a large co-authorship network, accompanied by shift in focus.

As stated the data are extracted from the public revision histories of the living deliverables, made available by MediaWiki. They can be used to show whether wikis show any signs of editing activity and to further investigate the collaboration network structure expressed in these revisions. It is possible to inspect who is collaborating on particular pages. In large projects, like STELLAR, these visualisations can help to make activities more transparent which can create more awareness and accountability – and ultimately offers triggers for new activity.

For living deliverables as such, it provides a way to check for signs of life, especially when their delivery deadline has passed.

There are several limitations this study has. Most notably, collaboration in co-authoring wiki pages cannot be mistaken for the overall collaboration on the (printed) report delivered to the European Commission. All wikis had phases close to the deadline, where an export of the Wikipages into a Word-file served the final polishing and further elaboration. All the deliverables were embedded into collaborative activities of other nature, such as presence and virtual meetings (flashmeetings), reviews (with separate reports), and other forms of collaboration that left no traces in the wikis. Still they are part of the process of creating their content.

Moreover, we have so far looked at only a small number of living deliverables in a limited time period. It will be very interesting to see, whether our findings will be confirmed when repeated in the future with more data and a longer time frame. Not to mention that it will be interesting to see, whether there is an afterlife of the deliverables beyond the runtime of the project.

It is an open question, whether the analysis method used can be matured into a self-explaining visualisation that does not require any insider knowledge about the collaboration in order to correctly read it. Or in other words: an evaluation of usability and accuracy is pending. This might also be helpful further what (wiki-wise) the difference between a living and living dead deliverable is. And it might help to identify driving factors: is it the medium, the collaborators, or the content?

In its current form, the co-editing network plots depict only a holistic view of all contributions. A more flexible approach would be to let the user interactively choose time windows, thereby providing means to investigate collaboration patterns before and after significant events. An animation of the graph change over time would additionally help to understand the development of a living deliverable, emphasizing the process dimension further.

A more fine-grain distinction of the types of contributions and their drivers would serve further analysis: writing passages, proofreading, enhancing with links and media, discussing, altering, and deleting text are all important for the quality of an article, but possibly not all of them trigger further activity by collaborators. This would be equally interesting for life and afterlife of the deliverables.

Additional evidence sources are available to further investigate collaboration among the researchers outside the living deliverable. It would be very interesting to see whether collaboration patterns differ when looking at the accompanying virtual meetings, e-mail exchange, or presence meetings. Does the medium foster certain styles of collaborations or do they converge?

From a project oriented view the proposed type of analysis could serve as a feedback mechanism making achievements visible. This could help to activate discussion about research collaboration.

Acknowledgment

Reference

[1] Duncan, W.: Developing a project-management body-of-knowledge document: the US Project Management Institute's approach, 1983-94, In: International Journal of Project Management Vol. 13, No. 2, pp. 89-94, 1995

[2] Arazy, O., Stroulia, E., et al.: Recognizing contributions in wikis: Authorship categories, algorithms, and visualizations. Journal of the American Society for Information Science and Technology. 61, 6, 1166-1179 (2010).

[3] Nunes, S., Ribeiro, C., David, G.: Wikichanges-exposing wikipedia revision activity. Procceedings of the 2008 International Symposium on Wikis (WikiSym) Porto, Portugal. (2008).

[4] Viégas, F.B., Wattenberg, M., Dave, K.: Studying cooperation and conflict between authors with history flow visualizations. Proceedings of the CHI 2004, pp. 575-582, ACM, Vienna, Austria (2004).

[5] Suh, B., Chi, E., Pendleton, B.A., Kittur, A.: Us vs. Them: Understanding Social Dynamics in Wikipedia with Revert Graph Visualizations, In: IEEE Symposium on Visual Analytics Science and Technology 2007, IEEE, Sacramento/CA, USA, 2007.

[6] Baumgrass, A., Mueller, C., Meurath, B.: Analyzing Wiki-based Networks to Improve Knowledge Processes in Organizations, In: Journal of Universal Computer Science, Vol. 14, No. 5, pp. 526-545, ISSN 0948-695x, 2008.

[7] Jesus, R., Schwartz, M., Lehmann, S.: Bipartite Networks of Wikipedia's Articles and Authors: a Meso-level Approach, In: WikiSum'09, ACM, Orlando/Florida, USA, 2009