M3.11 – Potential mechanisms to increase user uptake

Making invisible work visible

Date: January 13, 2012 / V130112
Work package: WP3 'Training, outreach and community support'
Partner: VU

Reporters: Peter van den Besselaar & Daphne Duin (VU) together with: Christos Arvanitidis (HCMR), Irina Brake (NMH), David King (OU) and Isabella Van de Velde (RBINS)

Making invisible work visible

The purpose of information systems, or sometimes an unintended result, is that they make work visible that was once invisible. In information systems "behind the scene work" is rationalized in the design and made accessible for (all) users of the system. This is also the case for ViBRANT tools and services that aim to support biodiversity researchers in doing their work. The tools archive and link data and people in biodiversity research in a way and on a scale that was not possible before.

Hence, infrastructures like ViBRANT make pieces of work and the organization of work visible to a large audience, were before these were only visible to colleagues sharing the same office or otherwise working in close proximity to each other. These work activities and work settings include among others: (informal) communication and brainstorming; sharing work in progress (annotated texts) and visibility of network relations. Tools like forums, bug/issue trackers, groupware, access to use-metrics (e.g. number of visits, downloads), and social network applications, make that today in science not only the final research article or product is shared, but also how the work settings are organized, what steps are taken to get to the final output, as well as how often the output is used.[1] If facilitated well, these tools help researchers in advancing and communicating their work and interactions, and will give a more detailed picture of the impact and use of biodiversity research in general.

During ViBRANT's first year we found evidence that the mechanisms which help researchers to make invisible parts in their work visible, are also likely to give them more control over their work and so increase their autonomy[2]. The COMBER project shows that respect is another important element to make users engage.

Below first the ViBRANT activities are described that explicitly addressed the engagement of users in Year 1. We conclude this report with a short discussion on the ViBRANT outreach and communication strategy.

Potential mechanisms

All activities carried out under the ViBRANT umbrella take in hand the engagement of users. However, when and how this is integrated in the development process differs for each project. In this section we will discuss four different projects that had the engagement of users up front during Year 1.

1. User studies

The VU explored two approaches that may contribute to make invisible research work visible. Both studies are based on the use and users of Scratchpads.

Scope of the audience

The first study looked at who is using the Scratchpads. More precisely, can we identify the different audiences that access the data published by the virtual communities? Scratchpads are visited by registered and unanimous users. The registered users, if logged in, can also access information that is not 'open' and may contribute content. Unanimous users have only access to

[1] For a critical discussion on consequences of making invisible work visible see: Star & Strauss (1999) and Suchman (1995)

[2] Autonomy is generally considered an important element of professionalism (cf. Engel, 1970).

the public pages of the Scratchpad sites, sometimes they land on a site because of a search term they used in a search engine.

Uncovering who is visiting the Scratchpad sites will help to get a better view of the impact of the service and where this impact is most likely: inside or outside of the domain of biodiversity research, inside or outside the specialty, inside or outside academia, and where in case of non-academic users. We worked on a method to reveal the scope of the audiences of Scratchpads by analyzing incoming web traffic to the Scratchpad web domain 'myspecies.info'. Most internet statistical reports show the numbers of visitors, the country where visitors access the sites from and the time they spent on the sites. Hence, internet statistics hold also other information that is worth exploring. We used the web reports of the web domain myspecies.info to study the Internet Service Providers (ISPs) that visited the Scratchpads.

The first analysis on the ISP's names that we run showed that the principle of using ISP's names works and helps to identify and cluster visitors. In this first attempt the VU identified the following categories of users: Research & Education; Government; Business; Non-Profit; Health; Art/culture/media; and Travel.

Theoretically, with a more robust method, the maintainers of a Scratchpad site should be able to apply the method of clustering ISPs as a tool to demonstrate the (social-economic) impact of their sites. Facilitating reporting on usage of the work done online will increase users' autonomy over the communication of the impact of their work to the bodies that are important for them (e.g. peers, department, funding bodies). However, the results also showed that the current method needs further work (see appendix). For this the VU has sought computing expertise from ViBRANT partner Open University (OU, David King) to further improve the filter that was used. The results of this activity will be published during ViBRANT year 2.

The data discussed were presented at the Altmetrics Workshop of the ACM Web science conference 2011 and are further summarized in the appendix of this report.

More information:
Abstract Altmetrics workshop: http://altmetrics.org/workshop2011/

Multiplex networks: co-author networks and Scratchpad membership

A second line of study that was carried out by the VU aims to make two types of relationships visible wherein researchers in biodiversity research are connected. In this study the VU compared the traditional connections of co-author relations between researchers to a new way of connecting to peers i.e.: Scratchpad membership (Duin & Van den Besselaar 2011). The idea behind the study is to investigate a possible added value of Scratchpads compared to the traditional type of connections in academia. VU proposed theory and methods to study and tested some first ideas on multiplex networks in biodiversity research on a small set of network data. This type of research wants to make visible the network relations of biodiversity researchers, provides arguments on how their traditional and new type of relations might effect the creation of new knowledge.

One of the applications of mapping multiplex networks of Scratchpad users could be the design of a social navigation system. Social navigation systems are IT applications showing other people's actions directly or indirectly. A popular example of social navigation on the web is the customer service "people who bought this also bought…"

Wu & Bowles (2010) studied social navigation in collaborative systems. According to them:

> *The values of social navigation support in collaboration are: discovery of new features; predicting the consequence of certain actions and decisions based on what other people have done previously; and conveying cultural context to meet the expectations of other members of the collaborative space (p. 1).*

Showing users of Scratchpads how other users are linked to each-other through different relational networks – will give them tools to (more strategically) link to people, ideas and information.

2. Co-learning between developers and users

As argued by Brake et al. (2011), in ViBRANT a platform like the Scratchpads benefits from co-learning between developers and users. Traditional support mechanisms in an IT environment are emails to a help desk or a downloadable text manual. In the last several years another help support service has been widely adopted, the issue tracker. An issue tracker is an online system to report bugs and feature requests of a specific IT service. The Scratchpad platform uses this tool for users *and* developers to report bugs, to request support or to make feature requests. The issue tracker openly lists the issues that are reported and shows the "status" of the report. By going to the issue tracker users can immediately see on which issues developers are working, what other users and developers already reported on and read the solutions that have been proposed in the past. So, the issue tracker facilitates the communication and interaction between users and developers and among developers in the development team. Making this communication visible contributes to the integration and active participation of users in bug tracking and will result in bugs being fixed faster and more efficiently (Breu et al. 2010). The issue tracker is not only important for distance communication but also for co-located teams. Here the tools service as an open archive or a shared to do list (Bertram et al. 2010). The issue tracker is an example of a social navigation system and links to the observations by Dieberger et al. (2000) that:

> …*many, if not most, digital information systems would be improved if their designers considered how one user within the system could help another. Such thoughts could turn the lonely, void socially information spaces we have into more humane environments, and maybe into real places (p. 45)*.

On the other hand the help emails cover general enquiries about the project. They are received by the whole Scratchpad development team and are answered by the team member best suited to the task. A possible advantage of email over the issue tracker is privacy of communication. Brake et al. (2011) emphasize that the issue tracker and the more traditional help-email serve both specific needs for users and developers. Understanding and facilitating what tool to use for what purpose and when will help increase user uptake (p.190).

More information:
http://dev.scratchpads.eu/project/issues

3. Scratchpad training courses and the ambassador programme

The Scratchpad platform has several mechanisms at place that aim to increase user uptake on the short, mid and long-term. These activities are run by the partners RBINS and NHM within the scope of WP3.

On the short term the basic training courses aim to help people start a Scratchpad site. The basic courses make it easy, also for people with little or no experience as a web master or

with content management systems (CMS), to build their own site and invite others to participate and manage the content and activities of an online community in their field of expertise.

The advanced training courses are a mechanism to increase the usage of the site of people that already signed up and so aim to have an effect on the usage and user participation on the mid-long term. The advanced training courses encourage those users that have been using the system for a certain time to start using it to its full potential. For users who would like to work on very specific improvements of their site or their own skills, tailor-made courses are offered.

To foster long-term sustainability of the Scratchpad community, the ViBRANT project launched the Scratchpad ambassador programme. A select group of Scratchpad users is recruited to be the official local representative of the Scratchpad community, linking the Scratchpad team with Scratchpads' growing user base. Ambassadors spread the word about Scratchpads, promote the use of Scratchpads and arrange or give training in their local Scratchpad community. Some ambassadors are the point of contact person for Scratchpad users in their taxonomic community, and in that way they help the Scratchpad team to better understand the needs of users, so that the Scratchpad developers can keep improving Scratchpads (M3.10).

More information:
Training: http://scratchpads.eu/scratchpad-training-courses
http://scratchpads.eu/ambassadors-programme

4. Marine inventories with citizen scientists

As put forward by Arvanitidis et al. (2011) the acknowledgement of the taxonomic impediment contributed to the formulation of two priorities in the organization of biodiversity sciences: a) find ways to increase taxonomic efficiency and b) establish data collection programmes and networks. The taxonomic impediment relates to the 86% of the known existing species on earth and, possibly, as much as 91% of the species in the ocean still awaiting description, while at the same time the professional taxonomic workforce is quickly declining. COMBER, a pilot project, of ViBRANT, contributes to both priorities by engaging citizen scientists in a coastal marine biodiversity observation network and channeling the data collected to large data aggregators like GBiF. Arvanitidis et al. (2011) report how amateur divers are motivated to participate in COMBER and so to help the research community in tackling the taxonomic impediment.

According to their findings the following mechanisms motivate amateur divers to participate in the citizen science project: 1) the feeling of contributing and being useful for science; 2) being part of an international network; 3) gaining new knowledge about nature (p. 221). An evaluation of COMBER by the amateur divers indicated that user participation could be further increased if the organizers change the following: (a) offer more detailed underwater introductory seminars about marine biodiversity to make identification underwater easier; (b) provide online material (presentations, photos, videos, quizzes); (c) include more fish species and other taxa (e.g. sponges, mollusks) and (d) to better promote the website (through higher ranking in search engines, Facebook and Twitter) (p.222).

More information:
http://www.comber.hcmr.gr/?q=node/5

Outreach of ViBRANT activities

Above we discussed the activities that provide potential mechanisms to increase user uptake and why we think they do. The activities mentioned have something to offer to users but it is not enough to develop a smart and appealing tool if no one knows about it. Users have to be actively informed where to find the tools and services such as the issue tracker, the training courses, the ambassadors programme and the COMBER project.

In order to make users aware of the ViBRANT tools and services a promotional strategy has been developed (RBINS, M310). One of the core strategies is to have partners continuously contribute to the outreach activities and add their contribution (presentation, poster, etc.) to the ViBRANT website. Events of interest for the project and partner participation to events are also tracked on the ViBRANT website. This is an efficient way of keeping track of which partner participates in what event and to coordinate the outreach of different tools and services.

Finally, all ViBRANT tools when ready for use, have a public website or wiki.

More information:
M310:http://vbrant.eu/sites/vbrant.eu/files/M.3.10_Delivery%20of%20a%20promotional%20strategy%20for%20ViBRANT%20services_0.pdf
Event list and calendar: http://vbrant.eu/content/events.
Participation and type of outreach activity: http://vbrant.eu/talks.

Summary

In this report we highlighted several activities developed under the ViBRANT umbrella, during Year 1, that explicitly addressed the engagement of user. As emphasized, the activities have in common that they contribute in making academic efforts, relations, participation and data, visible. Making work visible can mean that the task gains legitimacy and is rescued from obscurity (see also Star, 1999 p. 9)

Concepts and tools that extend user's autonomy, like access to data and people (not limited by organizational, or national boundaries) are also thought to be central ideas of ViBRANT, offering advantages over the traditional offline organization of science. Highlighting these advantages will help increase user uptake. Below we summarize the lessons learnt from these four activities and in doing so we give a list of mechanisms that we think have the potential to further increase user uptake in ViBRANT.

1) The user studies of the VU aim to provide insights on the social context of ViBRANT tools and services. In doing so the VU wants to contribute to the development of mechanisms to increase user uptake.

   -The 'measure the scope of the audience' method which is currently under development, is based on the principle that meaningful impact reports on online-work provide a tool for researchers to make their invisible work visible and so enhance their autonomy. In doing so an attractive condition for user participation is created. Furthermore, this study does allow us to understand ViBRANT users better, which gives us valuable pieces of information. For instance, if we need to target our outreach and training activities - either to reinforce areas where we are already strong, tailoring the materials to that specific interest group, or rethink why we are missing some potential users altogether. This will inform our choice of materials to develop, at which conferences to offer workshops, and how to best use the available funds.

-The method and theory of multiplex networks suggest that Scratchpad membership has an added value in network terms to traditional academic network relations, such as co-author relations. Stressing this point in promotional activities should help increase user uptake. Hence, a more extended empirical study is needed to test the hypothesis.

Furthermore, the pilot study carried out by the VU (Duin & Van den Besselaar 2011) suggests there is a second mechanism at work that might help increase user uptake, which is social navigation. Collecting the data on multiplex, academic network relations of Scratchpad/ViBRANT users (co-authors; citations; bibliographies; editorial boards networks) and making them visible to others, will help users more easily to link to people and ideas and so contribute to the creation of new knowledge. If social navigation tools will actually contribute to the increase of user uptake of ViBRANT tools is something to be studied. However, previous work on social navigation systems show their benefits to the overall usability of collaborative spaces and to the facilitation of learning (cf. Farzan & Brusilovsky 2006; Wu & Bowles 2010).

2) Organizing co-learning between developers and users by way of online tools such as the Scratchpad issue tracker is a clear example of a mechanism that encourages and facilitates the involvement of users in the design of Scratchpads. Being included in the improvement of the system, by identifying bugs, and by making feature requests gives users a sense of responsibility and autonomy. If something is not working users can actually do something about it.

3) The Scratchpad training courses lower the barrier for new users to join Scratchpads and for experienced users to help them bring their site use to a higher level. The courses give users information to help them organize their research community globally if they wish so, not limited by institutional or national boundaries nor by expensive travel costs. In addition, the ambassadors programme make power-users visible to other users and so offers a structure for help support between users, independent from the training and support given by the Scratchpad support team. Again an example of giving autonomy to users, which should be an asset for user uptake and long term sustainability of Scratchpads.

4) The COMBER project and the work of Arvanitidis et al. (2011) show that citizens are keen to be involved in science, especially if their contributions are made visible and are linked to a larger data aggregator and if they receive training and the right training material. Key here is that they want to make a difference but on the condition that they and their expertise are taken seriously and treated with respect, e.g. providing them with professional training and material.

References

Arvanitidis C, Faulwetter S, Chatzigeorgiou G, Penev L, Bánki O, Dailianis T, Pafilis E, Kouratoras M, Chatzinikolaou E, Fanini L, Vasileiadou A, Pavloudi C, Vavilis P, Koulouri P, Dounas C (2011) Engaging the broader community in biodiversity research: the concept of the COMBER pilot project for divers in ViBRANT. In: Smith V, Penev L (Eds) e-Infrastructures for data publishing in biodiversity science. ZooKeys 150: 211–229. doi: 10.3897/zookeys.150.2149

Bertram D, Voida, A, Greenberg S, Walker R (2010) Communication, collaboration, and bugs: the social nature of issue tracking in small, collocated teams. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work* (CSCW '10). ACM, New York, NY, USA, 291-300. http://doi.acm.org/10.1145/1718918.1718972

Brake I, Duin D, Van de Velde I, Smith VS, Rycroft SD (2011) Who learns from whom? Supporting users and developers of a major biodiversity e-infrastructure. In: Smith V, Penev L (Eds) e-Infrastructures for data publishing in biodiversity science. ZooKeys 150: 177–192. doi: 10.3897/zookeys.150.2191

Breu S, Premraj R, Sillito J, Zimmermann T (2010) Information needs in bug reports: improving cooperation between developers and users. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work* (CSCW '10). ACM, New York, NY, USA, 301-310. http://doi.acm.org/10.1145/1718918.1718973

Dieberger A, Dourish P, Höök K, Resnick P, Wexelblat A (2000) Social Navigation: Techniques for Building More Usable Systems. Interactions, Nov-Dec : 36-45. http://portal.acm.org/ft_gateway.cfm?id=352587&type=pdf&dl=portal&dl=ACM

Duin D, Van den Besselaar P (2011) Studying the effects of virtual biodiversity research infrastructures. In: Smith V, Penev L (Eds) e-Infrastructures for data publishing in biodiversity science. ZooKeys 150: 193–210. doi: 10.3897/zookeys.150.2164

Engel GA (1970) Professional autonomy and bureaucratic organization. Administrative Science Quarterly 15 (1): 12-21. URL: http://www.jstor.org/stable/2391182

Farzan R, Brusilovsky P (2006) Social navigation support in a course recommendation system. In: Wade V, Ashman H, Smyth B (Eds) Adaptive Hypermedia and Adaptive Web-Based Systems. Lecture Notes in Computer Science, 4018. Springer Berlin / Heidelberg, 91- 100. doi: 10.1007/11768012_11

Star SL, Strauss A (1999) Layers of Silence, Arenas of Voice: The Ecology of Visible and Invisible Work. Computer Supported Cooperative Work, 8: 9–30.

Suchman L (1995) Making work visible. Communications of ACM, 38, 9: 56-64.

Van den Besselaar P, Heimeriks G (2011) New media and communication networks in knowledge production. Cybermetrics, 15 (forthcoming).

Wu M & Bowles CT (2010) Principles for applying social navigation to collaborative systems. Proceedings of the 4th Symposium on Computer Human Interaction for the Management of Information Technology CHIMIT'10, November 12–13, 2010, San Jose, CA, U.S.A, 2-2.

Appendix

Scope of the audience

The analysis is based on the ISPs visiting the total set of Scratchpad websites (Myspecies.info), covering >200 community sites - over the period October 1, 2010 - March 31, 2011. The web statics are tracked and achieved with help of Google Analytics. Over this period 9212 unique ISPs were identified (the full retrieval had 16484 unique ISPs' or visitors). This is after we used a threshold based on "Average time on site" > 4 seconds. In order to have a relevant set of ISPs for identification of audiences, the commercial ISPs, mostly telecom and commercial webhosting companies, were removed.

Table 1. Number of Internet Service Providers to myspecies.info. (Oct 2010-March 2011)

Myspecies.info	Number	%
All ISPs†	9212	100
ISPs without commercial ISPs	2316	25
† Average time on site >4 sec		

The first filter generated a list of ISPs including specific words that are part of the ISP name for instance:

> ...
> marin|medical|medisch|microsoft|mineral|mining|ministerie|ministry|monsanto|museo|museum|national park|naval|navy|nerc|news|novartis|observatoire|office …

In order to make a good selection of words to be integrated in the filter we first manually scanned the full list of all ISPs. Words that appeared frequently in the names of the ISPs were noted and associated words (marina|marine) were added later. A second filter was used to exclude ISPs having words in their name that relate to telecom companies, such as:

> ... dsl subscribers |http |O2|telecom|telefonica ...

These words were selected when manually scanning the list of ISPs. After removing most telecom providers, the remaining list of ISPs was significant shorter. With help of the two filters we reduced the list from 9212 to a final set of 2316 ISPs. These were then grouped in similar categories as used by Van den Besselaar et al. (2011). The categories were made up from words that could easily be recognised in the name of the full ISPs such as: "university; academic; research; library, school". These ISPs were grouped under the category Research/Education. For the category Government we searched for "Gov*" or a combination of words "state"+ geographical name" for instance. This approach enabled us to semi-automate an important part of the clustering. A final 100 results or so had to be labelled manually. The "Health" category includes all ISPs who had within their name the terms "health" or "medic*" and therefore encompasses a range of research, educational, governmental and corporate affiliated ISPs. This was thought to be important as particular in the health sector it is hard to distinguish public and private institutions.

The analysis of the ISPs visiting the combined set of web pages resulted in a clustering in eight audience categories (see Table 2). We were able to classify 25% of the total number of ISPs that visited the web domain of the Scratchpads in the set period. The others are the commercial ISPs, and of course we cannot identify the professional or social background of users that enter the Scratchpads through these ISPs.

Table 2. Categories of Internet Service Providers to myspecies. Info. (Oct 2010-March 2011)

Categories of ISPs	Number	%	Unique Visits†
Research/Education/R&D	1933	83%	22680
Government	204	9%	1867
Company	50	2%	107
Non-profit	41	2%	263
Health	39	2%	96
Art/culture/ Media/publishing	25	1%	163
Travel	23	1%	89
Other	1	0,04%	1
Total	2316	100%	25266

† Represent the number of unduplicated (counted only once) visitors to the website over the course of a specified time period. A Unique Visitor is determined using cookies.

As Table 2 shows, the category of Research/Education clearly dominates. Yet one should note the category still covers a wide variety of organisations in the field of research and education such as universities, colleges, research centers, schools, libraries, and R&D institutions. Also it is not surprising that the Scratchpads attract mainly audiences with a research or educational affiliation. The sites are maintained by scientists to collaborate and disseminate knowledge. However, because science today is under pressure to demonstrate its use by a diversity of societal stakeholders and its benefits for society at large, we were in particular interested in the less obvious ISPs representing other than biodiversity research institutions. What we showed here is that with relatively simple tools one can identify types of audiences that are represented by their organizations (the ISPs).

The methodology (line of thought) and the data mentioned here have been presented at the Altmetrics workshop of the ACM Web conference. We are currently working with partner David King (OU) to improve the data mining process and build a more robust methodology for clustering ISPs. There are several caveats that we would like to mention. The main caveat of the current exercise is that this clustering is not yet developed enough to be applied on another set of ISPs. First of all because it has been mainly a manual process of selecting specific filter term and clustering the ISPs in categories. Hence, the filter works well for *this* specific data set but not necessarily for another set. Also we assume that the users represented by the identifiable ISPs are a representative sample, but this is something that needs to be tested in the future.

What's more, further improvement of the search strings could refine the category Research/education and identify the different educational levels (e.g. primary, secondary and higher education) and the specialization of the institutions (Geology,

Agriculture, Marine Biology, Physics, etc.). This may lead to a better insight in the (inter)disciplinary use of the Scratchpads. As this was a first attempt to cluster Scratchpad audiences, we purposefully kept the threshold “average time on page” rather low, at >4 seconds**.** We wanted to keep the data set at large to learn as much as possible about the audiences coming to the sites. However, increasing the threshold may give more robust results, which has to be tested. In any case, increasing the threshold will only keep the more heavy users in the database, and comparing these with all users is something to be tackled in future research as well as if some specific Scratchpads attract a more homogeneous or heterogeneous audience.