

## Open Source Software, Training, and Support in Digital Preservation: An Interview with Bram van der Werf

This is the # installment of the [National Digital Stewardship Alliance Innovation Working Group's](#) Insights series. Briefly, Insights is an occasional feature in which members of the working group talk with people doing exciting, innovative work in and around digital preservation and stewardship. In this post, I am excited to have the chance to chat with Bram van der Werf, Executive Director of the [Open Planets Foundation](#) (OPF),

*Could you give us some background on the Open Planet's Foundation? How did it come about? Specifically, what problems was the foundation created to remedy and how are you going about trying to address those problems?*

**Bram:** After the EU funded project "Planets" was closed, OPF was founded as a not-for-profit organization with a main objective to sustain the outcomes of digital preservation projects. Funded projects have a strong tendency to produce lots of outcomes in terms of prototypes and research papers during the course of the project. In most cases these same projects have limited intention, plans, or resources to sustain their results and findings. This is very much in contradiction with preserving digital content providing long term access. Digital preservation requires continuous efforts, not only in terms of researching and developing preservation functionality (practices and tools), but more importantly also to integrate, maintain and manage these same practices and tools.

*Could you tell us a bit more about the particular model the OPF has implemented and what you see as the implications for stakeholders in digital preservation?*

**Bram:**

As indicated by its name, OPF is first an open community with the ambition to support a community of digital preservationists around the globe. As digital preservation becomes increasingly technically challenging, OPF sees bringing practitioners, librarians and archivists (collection owners) together with developers and IT staff as instrumental to its success. This is the reason why we see the OPF community as a blend of digital content owners and practitioners together with their technical supporters. Commercial solutions for digital preservation are not often available or are beyond the financial capacities of memory institutions. Connecting content owners' requirements with the reality of technical supporters, along with taking an iterative approach towards integrating preservation functionality into repositories, is the practical and pragmatic OPF endorsed method. OPF browses the research community for practices and tools that fit its direction and, if required, supports the maturing of prototypes and provides life cycle support for software orphans that have identified digital preservation utility. Many stakeholders identified that they need these tools. These identified use cases can articulate the immediate need of a user community. Many of these tools have not (yet) proven to be commercially viable. Having these tools managed as open source tools will take a community of stakeholders that is committed to maintain and improve the tools over time. Providing stewardship to this community is where OPF can actively provide support, making sure that tools are supportable, well documented, and easy to find on the web. Memory institutions being the main stakeholders and users of digital preservation tools today, they need

to realize that pursuing open source software is not the same as buying off the shelf commercial software. The support model requires their active participation from development through life cycle management. It is consumption versus participation and this is exactly the reason why OPF, from its inception, has stressed the importance of building a community or hub of digital preservation stakeholders.

*I have heard a fair bit about the [OPF hackathons](#). Could you tell us a bit about the idea behind these events? What their goals are, how they work, etc. It might be helpful if you could walk us through one of the recent events, who participated, what they worked on and what the results and implications are for the field.*

**Bram:**

In 2012 OPF organized two hackathons, this year we will continue and have planned four events. OPF hackathons are definitely not purely techy events like many other hackathons. Maybe we should consider another name for the events, any good suggestions are welcome. In general our participants are a mixed group of practitioners and techies. Unlike conferences and seminars we sort of enforce participation and the nature of our event comes much closer to the concept of a three-day workshop. I hope I don't sound like a broken record, because the idea behind our hackathons is to build the community of digital preservationists. We hope it will bond practitioners with techies and support them in thinking about the real problems of today and what we need to do to come to practical solutions. During events we stimulate blogging, record event proceedings, and post discussions into the [OPF wiki](#). The idea is that we would like to see webinar versions of these events enable people from around the globe to get actively involved. Each event has a specific focus and we ask participants to bring practical examples; with these cases of relevant test data, we like to discuss and hack around real problems and like to avoid too much discussions around theoretical and hypothetical issues.

For example, last week we had a hackathon in Copenhagen. The focus for this event was database archiving and we took some of the Planets outcome around the [SIARD solution](#). The SIARD solution was developed and sponsored by the Swiss Federal Archives and grew out of the need to have a manageable archiving solution that would address the multiple dependencies when archiving many different databases. The Danish National Archives (DNA) is a full member of OPF interested in database archiving and offered to host this event. The DNA had started to develop a solution inspired by and similar to SIARD called SIARDDK. We also explored [RODA](#) (RODA is an Open source repository project that also has the possibility to archive databases ), a database archiving project from [DANS](#) (Database Archiving and Networked Services), and a archiving solution from the Norwegian, Swedish and Finnish archives. Finally the University of Freiburg demonstrated preserving databases with emulations. During the event the [agenda](#) is flexible. Almost 50% of the time of the events are break-out sessions where practitioners discuss issues and requirements while in parallel techies hack on existing solutions and explore commonalities and sharing opportunities. Early in the event it became evident that a shared database archiving format is really needed to support multiple scalable solutions and it was agreed to establish a working group that will work towards a common database archiving standard. The attendees also agreed to establish a working group

on the OPF wiki around database archiving requirements. Jose Ramalho of RODA announced that RODA will adopt to the SIARD format and that it would be relatively straightforward to change the RODA software accordingly. The SIARD programmer Hartwig Thomas shared lots of his experience during this event. An example of the immediate effect of this sharing of experience can be found on [some of the blogposts](#).

*Thanks for sharing that. I'm thrilled to hear that the hackathons have this focus on solving practical problems and that they try to bring together practitioners and developers from a range of stakeholders. The way you have described this reminds me a lot of the model that CurateCamp follows. For example, in consort with the recent Digital Library Federation conference they held a special event targeted at reaching out to bring Catalogers together with developers to work through issues around future digital tools for cataloging.*

*I would be curious to know more about your, and the foundation's, underlying approach to thinking about open source software. From your perspective, what do you think are the key factors for evaluating open source software for use in digital preservation workflows and systems?*

**Bram:**

There is no commercial commodity software or service in the market today that fulfills the need of many memory institutions. Both a commercial and an open source solution need many users to make it financially feasible from an operational perspective. The existing commercial solutions do not have a widespread user base. If memory institutes prove that they can work as a community and collaborate on solving problems, open source software can be a good alternative to commercial solutions. There are already several open source repository solutions and in that respect it does make sense that preservation services work as add-ons or plug-ins on these existing repository solutions. Like any other software solution that is a candidate for integration into workflows and systems, sustainability, supportability, and robustness comes first. This makes a preservation tool's development strategy around micro-services and small one-task-dedicated tools a preferred method. Small tools are easier to support and to debug compared to middleware types of tools built with sophisticated frameworks.

Another thing to keep in mind is that even though requirements often look very technical, in essence they really are not that technical. Sustainability and robustness in the world of open source software are the result of active usage. Small open source software projects in general lack extensive testing (functional and system) -- an extended user community can compensate for that lack of real testing. This does require stewardship on the status of prototypes, betas, and released version. Evaluating existing open source solutions starts with evaluating the strength of the user community in absolute numbers, while building the same solutions starts with evangelizing the solution within a wide community of potential users.

*What do you see as some of the biggest problems in digital preservation? Are there specific areas where you think the tools are lacking? Is there a need for more extensive training and*

*research? In other words, I am curious to know what you think are the biggest hurdles to long term preservation and access of digital content.*

**Bram:**

Throughout the history of mankind people have been inventing things and most of these things needed people to maintain them. So in the long term the one and only thing that will make digital objects survive is people. These people need to be supported with tools and learning systems, but it is all about skilled and motivated people. I agree we need continuous training of our IT staff to enable them to integrate preservation tools into repositories; and actually there are already many tools available. Many of these tools are command-line based and can be integrated via scripting and this takes training of IT staff on how to integrate, deploy and maintain. In earlier days much effort was made to develop GUI based tools, unfortunately this was not the way to move forward for integrating tools into workflows. I strongly believe in API or command line tools for digital preservation, since they can enrich existing repositories with long term preservation functionality and that is what we should strive for. But I cannot stress this enough, this takes skilled and trained people. A basic rule of thumb in commercial business is, you do not outsource or subcontract your core business skills. So, is managing digital content over time the core business of memory institutes? If this is the case, it would be highly desirable to retain core competences and reconsider our organization and people strategy.

Research is a slightly different organizational challenge. Where it is needed it is less of a continuous process compared to data management and its technical interventions. A project well connected to the academic world can work effectively as long as we make sure that research stays connected to developing solutions for real problems. Typical research areas with future potential for long term access would be virtualization, emulation, and cloud computing. For me research is a proactive risk management type of activity with a vision to the future, whereas lots of digital preservation actions should be a response to where things go wrong today and how we should react. This is very similar to traditional R&D in relation to maintenance services. We should be very considerate about the fact that these are two separate disciplines that cannot exist without each other. We need the right people in our community who can respond pro-actively (R&D developing tools) or actively (IT implementing tools).

So to summarize my answer to your question, maybe we have a need for tools but we have a much bigger need for skilled people and an active preservation community. Motivated people with the right skills will be able to produce and maintain tools, so training and investing in people is the key for getting the tools. Our hackathons prove that bringing the right people together can generate immediate solution to existing problems. Maybe these solutions have the nature of patches, but with feedback to R&D one can still plan for fundamental solutions. So the biggest hurdles are doing and learning.

*Based on your experience, if there was one key piece of advice that you wanted to give to anyone putting together and establishing workflows and systems for digital preservation what would it be? Is there anything that you think is foundational but you often find is missing in these approaches and plans?*

**Bram:**

Learn from the Swiss knife analogy. The Swiss knife (don't tell the Swiss army) is a bad can-opener, you can't even kill a rabbit with the knife, and worst of all the corkscrew will ruin your fine bottle of French wine. So always keep it simple, solve problems step by step, and stay focused on one problem at a time. Single purpose tools are good in what they are designed for, so if there is no direct need for multiple solutions, stick to single solutions. Many major organizations have a big reputation that actually drives them to exaggerate on functionality and integrations of solutions into one single tool. These solutions will be painful to integrate and an absolute nightmare for maintenance. Develop tools for the actual user and they will need performance and robustness. I know that GUIs are nice for demos and senior management presentations, but inside a workflow they kill performance, productivity, and robustness of the system and in the end that is what really counts. Start with something that really meets the system requirement and iterate the functional requirement.

*Do you think there are any problematic issues with how organizations think about the role of software and software systems in digital preservation practices? If so, how do you think we should be thinking about the role of software in digital preservation?*

**Bram:**

In all honesty this is where most confusion is in many organizations. Software and software systems are of crucial importance in digital preservation practices, in fact it is indispensable in modern society. The big challenge comes when organizations become involved in applying and developing software and software systems. The actual tangible value of software and software systems is minimal – the biggest economic effort for software systems lies in the investment into human resources. When we buy software we shift the cost of human resources to the vendor via the licence and maintenance agreements. In case organizations become producers and maintainers of software, they need to be prepared to compete on a competitive market of human resources. This means that organizations need to think about how they can become attractive employers for technology staff. This is not only a compensation issue, but also a HR and people management challenge.

Software is here to stay in our lives in libraries and archives and with challenges like web-archiving, mobile apps etc, etc, it is hard to deny the importance of it. Organizations need to re-think their organization and position for the future and this future is very digital with lots of software systems managing digital content. The actual time machine and engine for digital preservation is deep inside of our HR system in how can we train, hire, motivate and retain staff to manage and maintain our software systems.