iPres 2015 - Roles & Responsibilities for Sustaining Open Source Platforms & Tools

Overview

This workshop took place on November 2, 2015 as part of the iPres 2015 conference at the Friday Center in Chapel Hill North Carolina.

There were approximately 35 participants.

It was presided over by Trevor Owens, Senior Program Office, Institute of Museum and Library Sciences (IMLS). IMLS provides grants of up to $2 million per grant to Libraries and museums.

The participants were mostly users of free open source software (FOSS) rather than producers, with the notable exception of Evelyn McLellan, President of Artefactual Systems who are the lead developers of Archivematica and AtoM.

Trevor spoke about providing the right funding at the right stage of a development project. This is influenced by the sophistication and maturity of the project. It was mentioned several times that grant funding is not a sustainable model for funding a project.

Working documents as a result of this workshop:

- Schedule
- Break out working groups
- Sustainability Models
- Steps/Issues to consider before pushing out an Open Source Project
- Strategies for community building
- Define and Characterize Organizational/Institutional Roles

Evelyn McLellan, President, Artefactual Systems

Artefactual Systems is a private company that produces two Open Source Software products - Archivematica and AtoM. There were a good number of participants at the workshop who use AtoM and Archivematica.

Archivematica is an "integrated suite of open-source software tools that allows users to process digital objects from ingest to access in compliance with the ISO-OAIS functional model". These tools include FITS and JHove. The pipeline for the tools (which they describe as micro-services) is written in python.

AtoM (Access to Memory) is a "fully web-based, multi-lingual, archival description software that was originally commissioned by the International Council on Archives to make it easy for archival institutions worldwide to put their archival holdings online using the ICA’s descriptive standards."

Artefactual now use the Bounty Sustainability model - where users request features and pay Artefactual to develop them. They have accepted features developed by others in the past, but that is not unusual. The new features become part of the OSS product. Artefactual survives on paid services around the tool - consulting, training, hosting, development etc. fees.

Archivematica is released under the AGPL v3 license (a free, copyleft license). They chose this license in order to facilitate the integration of third party tools with restrictive licenses. They also have a CLA (contributor license agreement).

Pros of the Bounty Sustainability Model:

Institutions can pool funds to request features which Artefactual develops e.g. Ontario Council of University Libraries is partnering with Artefactual to develop the ability to ingest content from Dataverse to Archivematica!

Cons of the Bounty Sustainability Model:

Very difficult to get "unsexy" features funded (upgrades, development of scheduling system etc). Try to add these into the development of new features.

There is no pool of money from s/w licenses. Tends to be a drought or flood. With flood have too many requests for new features and they only have a fixed number of developers.

Where to spend money - integration vs core development.

She discussed what goes into an OSS project:

- Development
- Community discussion list
- Maintenance of User and Technical Documentation
- Governance model
- She would like a technical advisory group to guide s/w development - to make it bigger than the company.

She was very concerned about the sustainability of the projects should Artefactual step aside - who would step up? Participants seems confident that the software was so popular that someone would.

Questions and Answers:

Q. Wendy Gogel asked about the idea of maintaining 2 versions of a project - internal and external.
A. Evelyn said that she did not like that model. There should be generic features. Forks and branches will exist in the wild with institution specific features, but you should decide on what you want as the "Reference" implementation.

Q. Could she comment on a model for sustaining software

A. This is influenced by the Open Source nature of the code. They use the AGPL v3 licence which she considers to be the most "Open Source" licence. It is also important to use open standards. System independence of output is also very important.

Q. What about Community code input?

A. Artifactual has rejected code where did not meet the direction that they wanted the project to go. Artifactual is usually aware of what outside people are working on - they tend not to get contributions without first knowing about them.

Edward Corrado, Associate Dean Library Technology Planning & Policy, University of Alabama

Indefinite futures - we don't know what s/w will be around in the future.
Open standards are important.
Processes are important in addition to OSS.
Important to speak with the right people at the start of a project to be sure are fitting their needs.
Need to partner with people on campus.
Need to speak with Upper University Admin. to get support to fund it.
Need a sustainable model internally.
Issue - if tailor the project too much for your institution it won't be useful outside.
Funding - a grant may not be sustainable.
Need to build a community from the beginning for sustainability of the project.
Need to document the libraries required (dependencies).
Need to document the installation requirements.
Build the software so that it meets the basic level of maturity. There are 3 levels - basic, intermediate and advanced. To meet the basic level there are 10 things required.

There are different "sustainability" models used in OSS projects:

- Bounty Model - Artefactual uses this for Archivematica and AtoM.
- Membership Model - archivesspace, jHove use this. This model can make it difficult for new members to join because only members can access the latest documentation, latest version of the code etc. It is a pretty significant barrier to participation.
- Benefactor Model (aka Benevolent Dictator Model) : Villanova uses this model. Whoever wrote it keeps control. What happens if the benefactor drop out? Hard to get input from others.
- Standard OSS Model: Koha uses this model. People from all over the world contribute. How do you get to that community?

Education is important in this area and is lacking - Library schools need to be aware of these tools and to provide education on how to use them.

Questions and Answers:

Q. Controlled Releases?

A. Issue is how to get engagement and make sure people can contribute and have some say in the direction the project is taking. Cathedral and Bazaar development models. He referred specifically to Raymond’s Lessons for creating good open source software

Q. Many OSS became sustainable when embedded in commercial software.

A. There is a certain amount of truth in that e.g Apache HTTP server.

Carl Wilson, Technical Lead, Open Preservation Foundation

They are only 2 people! He is a software developer by trade. He specializes in small tools.

Documentation is important. Use of interoperable libraries Testing and testability.

He looks after jHove, a Better PDF (may be veraPDF)
Testability is very important. Code coverage of testing of jHove very low. Was a problem and needed to refactor the code to address it. Make unit tests part of contribution policy.

Release software which doesn't do much to get interest.

Continuous delivery so can do regular releases.

Accessibility ie. installer is very important so can download and run easily.

OSS is only embedded in commercial software when it is already successful.

Questions and Answers:

**Trevor Owens, Senior Program Office, Institute of Museum and Library Sciences (IMLS)**

He would like to add the idea of another sustainability model he has seen.

Freemium Model:

Any scholar could use it. Free OSS, anyone can configure it and use it. Also a paid licensed version where users pay for added features beyond the core free features.

IMLS has $12m grant money. Can submit a 2 page preliminary proposal.

Grant funding is the opposite of sustainability model. He made some reference to ePadd having this issue.

He spoke about current **OSS IMLS Projects** - ePadd, Omeka, Web Archiving, Hydra in a box, eBooks, artStor, BHL.

There is $10 billion per year spent on public libraries.

Library of Congress spends $600m per year.

Smithsonian - $800m per year

e tc. There is a lot of money spent on public libraries. How best to use it?

Collaborative investing.

Archivematica and Bitcurator integrate different products to produce a tool suite.

Questions and Answers:

Q. Why do the same large institutions show up practically all the time on these grants? How can small institutions get a say?

Q. Paul Wheatley (Digital Preservation Coalition) questioned the quality of review done by grant funders. Too much emphasis on innovation where the projects disappear afterwards.

The participants then broke out into 4 groups to address certain topics.

- **Sustainability Models**
- **Steps/Issues to consider before pushing out an Open Source Project**
- **Strategies for community building**
- **Define and Characterize Organizational/Institutional Roles**

I joined the Sustainability Models group while Wendy Gogel joined the Steps/issues group. At 3pm we were permitted to change groups at which point I switched over to the Steps/issues group.

The above documents are works in progress. The idea was floated that they could be made into whitepapers with a home at Code4lib.

Wendy will be working on the **Steps/Issues to consider before pushing out an Open Source Project** whitepaper with others. This group was led by Carl Wilson of the Open Plant Foundation.

The **Sustainability Models** group was led by Aaron Trehub, Assistant Dean for Technology, Auburn University. Evelyn McLellan's idea of a Center of Excellence Model was extensively discussed.

Development models vs sustainability models. Different models at different stages.

**Hydra**

Was a consortium between two universities. Started out as DIY and then moved to a different model.

**Islandora**
University Prince Edward Island created for their own needs. Then they open sourced it. Then:

UPE set up a commercial company - Discovery Garden (to help integrate).

Lyrasys hosting - for running as a service

Islandora Foundation - they decide on what contributions to accept.

This is a Bounty plus model!

Archivespace

Need to be a member to access docs and videos. They have around 200 members.

Trevor warned against setting up Foundations - cost of setting up a non profit is significant and most of the money can end up going to accountants.

Misc:

Do a survey of what tools already used at institution.

Open source for sustainability reasons - it is not cheaper.

Artefactual - User stories are contributed by external users.

Licensing - license, patent, copyright, trademark. Who on team to decide on license.

Frequent releases - so contributing developers get to see their work used, is a motivation.