



openZIM Report

Year 2009

2010-03-25

by Manuel Schneider

Events and Activities

January: Project Start

The base for the openZIM project was established during 2008 when Tommi Mäkitalo, maintainer of tntnet, created an open source version of the ZenoReader published by DirectMedia on their yearly Wikipedia DVD. This new tntzenoreader became part of the Wikipedia DVD 2008/2009 in October 2008. As the idea of having a common, free file format and tools that are open source to solve the Wikipedia offline problem spread to other projects a need of extending the works on the tntzenoreader arose.

In December 2008 the decision to found a project was made by formerly involved and new interested parties. Wikimedia CH was asked to provide funding to make the project possible: A server for development was needed as well as resources for having regular developer meetings to keep the project alive. The swiss Wikimedia chapter agreed in sponsoring the project by giving 4000 CHF for 2009.

DirectMedia was asked if we are allowed to use their name "Zeno" for the new project, as the file format is based on Zeno. As they used Zeno as a trademark and were about to sell their portal "zeno.org" they requested the project to use another name, so we came up with ZIM as the abbreviation for "Zeno IMproved".

Manuel Schneider rented a rootserver in Frankfurt and set up a server with Wiki, Subversion and virtual machines for development on Windows and Mac OS X. Project team members can get a SSH account on the server to write access to SVN and use the server for long-running tasks such as ZIM file creation.

February: 1st Developers Meeting, Official Launch of openZIM

The first Developers Meeting of openZIM with five participants took place in Schopfheim, Germany. The project team officially formed and the first documentation was put into the wiki.

Some changes to the new ZIM format have been discussed and a plan for 2009 has been decided, including making a [Wikipedia DVD for LinuxTag 2009](#).

For a full report of our meeting see http://openzim.org/Developer_Meetings/2009-1.

April: Participation at the Wikimedia Conference, Starting Cooperation with Wikimedia Foundation

The Wikimedia Conference was taking place for the first time and put together the yearly Wikimedia Chapters Meeting and the first MediaWiki Developer Conference, both happening at the same time in Berlin.

Manuel has planned to attend the MediaWiki Conference for his own interest but it was already full, then he was invited to join the Wikimedia Chapters Meeting as a board member of Wikimedia CH had to cancel his participation.

While the Chapters Meeting is an internal meeting of all Wikimedia Chapters around the world and the topics discussed there are Wikimedia-internal and of political / strategic nature there were still some opportunities to discuss about the openZIM projects and its ideas.

In the meeting we got contact to members of Wikimedia Israel, Wikimedia Italia and Wikimedia Polska working which were interested in the Wikipedia offline topic. After the Chapters Meeting Manuel still had some time to meet up with the MediaWiki developers and discuss with them.

A few weeks later we were contacted by the Wikimedia Foundation. Erik Möller (deputy CEO) wanted to learn more about the project. A phone conference took place with Erik Möller, Brion Vibber (WMF CTO), Tomasz Finc (WMF Developer) and Manuel Schneider (openZIM). During this conference the decision was made that the Wikimedia Foundation wants to support openZIM and

look into integrating the ZIM format into the dumping process on <http://download.wikimedia.org/>. Furthermore a one-click functionality to create ZIM files from MediaWiki should be evaluated.

June: Participation at LinuxTag, 1st Wikipedia DVD

LinuxTag is Europe's biggest exhibition and conference for free and open source software. openZIM applied for a booth in cooperation with tntnet, so we got one booth for both projects.

By this opportunity we created nice graphs to show how the ZIM file format is used and how it works internally. These were printed as A2 posters and showed at LinuxTag.

Sponsored by Wikimedia CH with 1000 CHF we published a set of 500 DVDs holding the German Wikipedia and the zimreader application as binary and source code. This brought us coverage in the major german IT news portal, Heise. They published an article about the new Wikipedia DVD and the openZIM project which brought our server's capabilities to the limits for several weeks because of the high web traffic.

On a lightning talk Manuel presented the idea of openZIM and gave away many DVDs to the interested audience. On discussions with Linux4Africa and Skolelinux we found new adopters for the ZIM format and software for the usage in schools.

For details see http://openzim.org/LinuxTag_2009.

August: Participation at Wikimania

Wikimania is the yearly conference of the Wikimedia community. Every year it takes place in another area of the world - in 2009 it was Buenos Aires in Argentina.

Manuel Schneider presented the openZIM project in a 30 minutes talk in the technical track of Wikimania. It was well received, videos can be viewed online:

http://commons.wikimedia.org/wiki/File:200908271434-Manuel_Schneider-openZIM_serving_the_Wikipedia_Offline_Projects.ogg

November: 2nd Developers Meeting, Major Changes to the ZIM Format

The second Developer Meeting again in Schopfheim, Germany, attracted further participants. Beside the project team itself the Wikimedia Foundation was represented by Tomasz Finc, the CEO and a developer of the french company Linterweb which works on Okawix was present as well as two persons from Qi Hardware, a company based in Berlin selling small embedded device like the Ben NanoNote.

In several discussion rounds the ZIM file format was further enhanced and clarified. Quite oppositional use cases had to be united into a new standard: Small devices such as the Ben NanoNote only can deal with a limited set of markup and need to strip everything but the article data from the format to save memory, while as Kiwix wants to add additional metadata for each ZIM file and article.

With a special meta data namespace for ZIM meta data and a secondary article namespace for per-article meta data and templates these problems could be solved. A unified integer compression, a dynamic MIME type lookup table and a second article index based on URL cleaned up the format and solved some other long-standing issues.

In the end even the long-wanted lzma compression was tested and considered being finally usable. All these changes broke the compatibility with the old format but introduced many new future opportunities for the ZIM format.

One major success was that Linterweb decided after the meeting that they will drop their development of Zeno to switch to ZIM as well with Okawix. This brings another leading publisher of Wikipedia and other offline content to participate in the ZIM community.

For more details see http://openzim.org/Developer_Meetings/2009-2

Media Coverage

While Wikimedia Deutschland was not willing to publish press releases from the openZIM team (despite promises by its president) we were still able to reach several news sites and blogs.

The most notable press coverage was the article on Heise online about the Wikipedia DVD and openZIM at LinuxTag, but several other blog posts from interested parties kept openZIM present in the web.

Started from scratch the project website has now about 3'000 unique visitors per month, generating nearly 5'000 visits, 28'000 page views and more than 50'000 hits per month, in average 0,5 Terabyte of traffic each month.

See detailed statistics: <https://intern.openzim.org/awstats/awstats.pl?config=www>

Budget

Wikimedia CH generously sponsored the first year of openZIM with 4'000 CHF. This money was spent to run a server for the project providing a technical infrastructure (CPU time for ZIM file creation, download area, version management software, bugtracker, wiki, website, virtual machines with other operating systems etc.) and for Developer Meetings.

As openZIM is exclusively run by volunteers openZIM tried to cover most costs for meetings including accommodation and catering for all participants. Travel costs, though, have been paid by the participants.

Date	Subject	Amount EUR
2009-02-18	OpenZim Server 2009-01-01 - 12-31 rgs4487	-600,00 EUR
2009-01-01	openzim.org Domain 2009	-12,00 EUR
2009-02-20	Developer Meeting 2009/1 Accommodation	-346,00 EUR
2009-02-20	Developer Meeting 2009/1 Catering	-45,58 EUR
2009-02-20	Developer Meeting 2009/1 Dinner	-165,40 EUR
2009-06-30	Copying Wikipedia DVD 2009	-803,25 EUR
2009-06-27	Accommodation LinuxTag 2009	-740,00 EUR
2009-06-27	Donations LinuxTag 2009	5,00 EUR
2009-11-20	Developer Meeting 2009/2 Accommodation	-406,00 EUR
2009-11-20	Developer Meeting 2009/2 Catering	-35,91 EUR
2009-11-20	Developer Meeting 2009/2 Dinner	-180,00 EUR
2009-11-20	Wikimedia Foundation Dinner Sponsorship	+180,00 EUR
2009-11-22	Developer Meeting 2009/2 Dinner	-167,80 EUR
Sum in EUR:		3'316,94 EUR
Sum EUR in CHF: (as of 2010-03-23)		4'760,13 CHF
2009-01-01	Wikimedia CH Sponsorship	+4'000,00 CHF
2009-06-06	Sponsorship Wikipedia DVD 2009	+1'000,00 CHF
Sum CHF:		5'000,00 CHF
Rest CHF:		239,87 CHF

See the full budget at http://openzim.org/Wikimedia_CH_Sponsoring.

Outlook for 2010

Roadmap

- ZIM export in MediaWiki
 - MediaWiki extension
 - dumping process on download.wikimedia.org
- stable applications
 - Kiwix
 - Ben NanoNote
 - Okawix / Wikiwix
- Documentation
 - ZIM v.5
 - provide a description for publishers of the whole dumping process

Events and Marketing

- Developer Meetings (spring and autumn)
- MediaWiki Developer Meet-Up (April)
- LinuxTag (June)
- Wikimania (August)