

# WikiTeam: collaborative preservation of wikis

EMILIO J. RODRÍGUEZ-POSADA, FEDERICO LEVA, LUIZ AUGUSTO  
WikiTeam  
{emijrp, nemowiki, lugusto}@gmail.com

## Abstract

Internet users have an increasingly role in web content creation. There are initiatives and solutions for the digital preservation of the web, including the well-known Internet Archive, but they are inefficient for archiving user-generated content in social networks and wikis. In this article we explore the problems in wiki preservation, the lack of tools to achieve successfully this task and we present and assess WikiTeam, the solution that we have built. WikiTeam is a collaborative effort to develop and run software for digital preservation of wikis. As of January 2014, we have extract the texts, histories, images and metadata for more than 4,500 stand-alone wikis, several wikifarms and 24TB of Wikimedia Commons files. The preserved content represents an huge cumule of datasets of the wikisphere, with an incalculable historical and research value.

**Keywords:** web digital preservation, social web archiving, archiving applications and systems

## 1 Introduction

This is a general introduction on web archiving.  
Ideas (write yours):

- brief description and importance of digital preservation, Internet Archive and Wayback
- user-generated content explosion, social networks, wikis, Archive Team, LoC archiving Twitter
- Wikipedia 2001, people start to use MediaWiki for their wikis, wikifarms,
- wikis not only are text and files, but interesting metadata, histories. IA preserves text/files (and not all/always), but it is inefficient saving histories and metadata
- as most wikis are free-licensed, there are no issues preserving this content

## 2 Digital preservation of wikis

This is a section for specifically wikis (an area inside web archiving).

Ideas (write yours):

- lack of public dumps/mirrors
- lost wikifarms (ScribbleWiki)
- existing software: wikitravel scripts oxygen-pump
- proposals: Urobe
- other: manually export with Special:Export, or ad-hoc scripts

## 3 WikiTeam: digital preservation of the wikisphere

Here we are.

Ideas (write yours):

- presentation of WikiTeam
- achievements (statistics, wikifarms, commons)
- single backups and backups in batches (launcher)
- uploaded to IA (long-time preservation, bittorrent webseed)
- how we generate lists of wikis and reuse others (Pavlo list)
- WikiApiary partnership

## 4 Conclusions and future work

Ideas (write yours):

- summarizing the best
- current issues and possible solutions

## Acknowledgements

## License

This work has a license Creative Commons Attribution-ShareAlike 3.0 Unported.