

# WEB ARCHIVING OF MUSICOLOGICAL INTERNET RESOURCES

---

Speaker: Barbara Lenk  
Head of library at University of Music Nuremberg

24.06.2015

# Agenda

## 1. **Web archiving**

- General information
- Method
- Example: Internet Archive

## 2. **Web archiving in music libraries**

- Web archiving at Virtual Library of Musicology (Bavarian State Library Munich)
- Research project: „Requirements from musicology for indexing and archiving Internet resources“

# Web archiving – general information I

- History of Internet begins 06.08.1991: first website published by Tim Berners-Lee at CERN
- **DEFINITION of WEBSITE:** A website is a virtual place at World Wide Web where several webpages or documents and other resources can be found. They are reachable via http-addresses. (BSB, 2014, paragraph 3)
- **Website** = whole construct of web presence
- **Webpage** = part of a website

# Web archiving – general information II

- Content of websites has changed since the beginning of the Internet → from pure text to multimedia
- Study 2001: life period of a website 75 – 100 days (Alexa Internet) → websites can be changed, be deleted or be moved to other positions



Source: <https://ijschinger.files.wordpress.com/2012/11/error404.jpg>



Source: <http://www.drweb.de/magazin/404-fehlerseiten-error/>

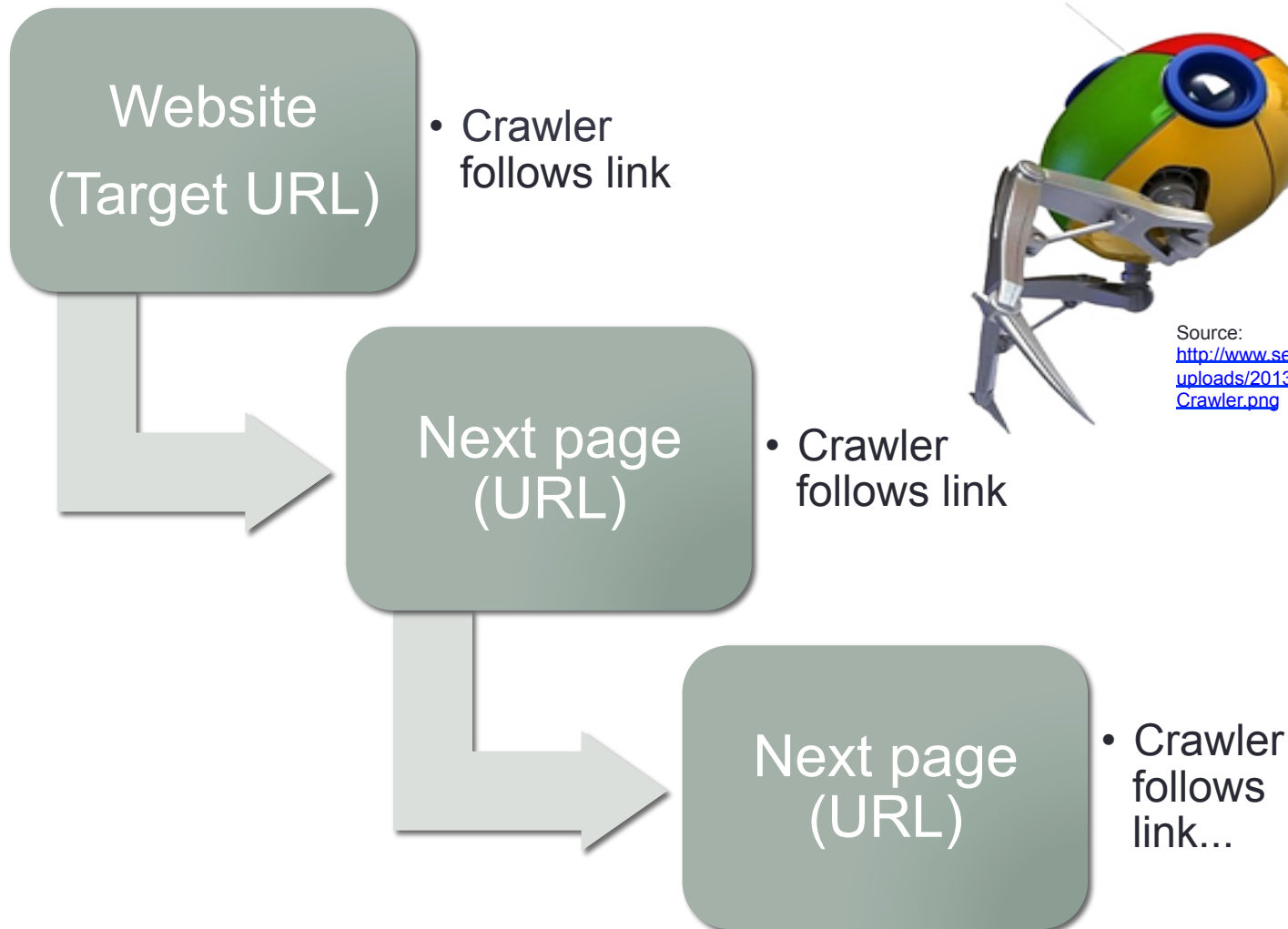
# Web archiving – general information III

- Institutions as libraries or archives start to deal with the topic web archiving
- **DEFINITION of WEB ARCHIVING:** All processes and procedures to select, to index, to store permanently, to archive and to supply websites to customers. (Brown, 2006, p. 5)
- **Goal:** Longtime preservation of a website and its content  
→ result: keep it usable as a scientific resource

# Web archiving – method I

- Process of collecting a website to archive it = **HARVESTING**
- Result: **CAPTURE** of a website (like a snap-shot)
- Tool to harvest websites: **CRAWLER** → follows link structure of a website → done several times for one website
  - **Domain Harvesting**: all websites of a domain, e.g. .de
  - **Selective Harvesting**: selected websites, e.g. musicological topics
  - **Event Harvesting**: websites belonging to an event, e.g. elections, FIFA World Cup 2014
- Captures are saved in a web archive, e.g. a repository
- Standard format: ARC or WARC
- Access to web archive with **wayback-machine** or other tools

# Web archiving – method II



# Web archiving – method III

- **LIMITS of web archiving**

- Robot Exclusion Protocols → operator of a website can set restrictions to crawler
- Deep web → e.g. content from data bases
- Authentication → content that is created after authentication
- Dynamic content → e.g. JavaScript or Flash

- **LEGAL aspects**

- Websites and their content are protected by copyright → ask an operator before archiving a website, content of third parties
- Privacy



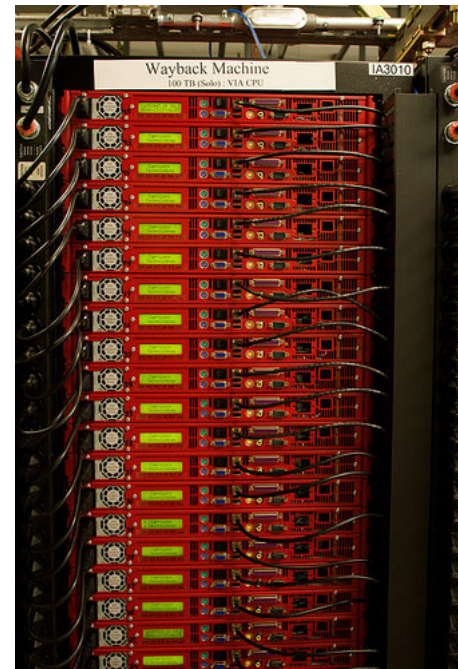
# Web archiving – Internet Archive

- founded 1996: Brewster Kahle and Bruce Gilliat
- <http://www.archive.org/web>



→ Goal: offer access to whole knowledge of mankind

Source: <http://archive.org/web/petabox.php>



Source: <http://www.newyorker.com/tech/elements/what-its-like-to-get-a-national-security-letter>

- **Biggest web archive worldwide**  
= about 479 billion sites all kind of content, multimedia
- Harvesting: Crawler Heritrix
- Wayback-Machine, free and public access
- Browse through many collections

# Web archiving in music libraries

- **Many websites with musicological content:**
    - websites of composers
    - websites of opera houses etc. and musicological societies
    - websites of universities of music and other schools
    - Blogs, Wikis with musical or musicological background...
  - **Heterogeneity:**
    - different types of content and design: pure text to multimedia
    - different levels of quality: musicologist to laymen, interested in music
- **All websites can be used as a scientific resource**

# Virtual Library of Musicology (ViFaMusik) I



**VIRTUELLE FACHBIBLIOTHEK  
MUSIKWISSENSCHAFT**

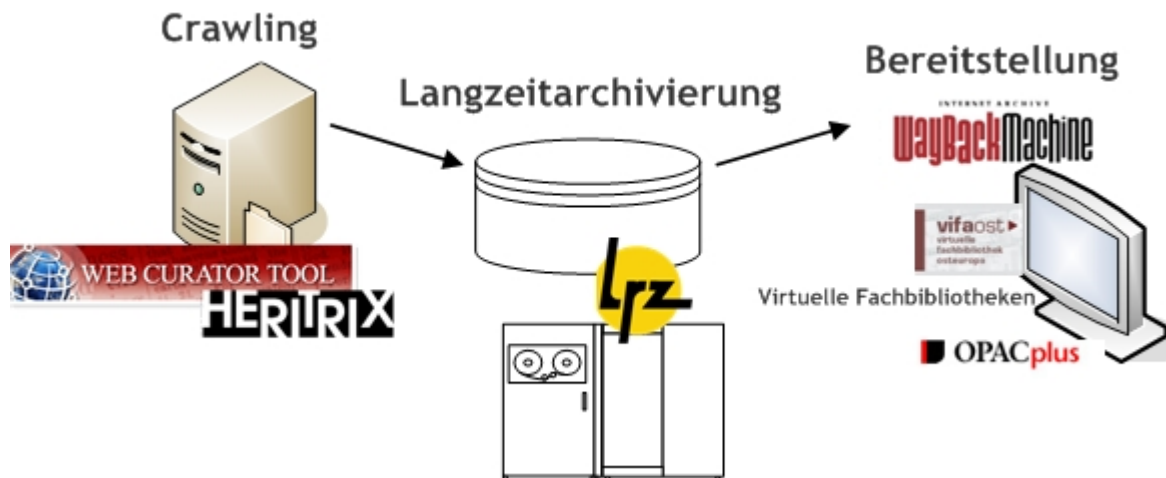


Source: <https://www.vifamusik.de/startseite.html>

- = Central online portal for musicology in Germany
- 2005: project was started at Bavarian State Library (BSB)  
→ funded by DFG (German Research Foundation)
- Partner: „Staatliches Institut für Musikforschung“ (Berlin),  
Society for Music Research (GfM)
- Part of: „Fachinformationsdienst Musikwissenschaft“ at  
BSB

# ViFaMusik II

- Collection of musicological websites
- Web archiving is done by „**Bibliothekarisches Archivierungs- und Bereitstellungssystem**“ (BABS)
- 2005: Project at BSB with Leibniz-Rechenzentrum (DFG-funded)



Source: [http://www.babs-muenchen.de/index.html?c=workflows\\_web&l=de](http://www.babs-muenchen.de/index.html?c=workflows_web&l=de)

# ViFaMusik III

- **Stock of website collection** by 06.06.2015:
  - 2,797 websites indexed → 154 of these are archived
  - <https://www.vifamusik.de/startseite.html>
- **2 types of searching:**
  1. Browsing by subject
  2. Keyword
- All websites with indexing information and live-link
- Archived websites in addition with link to web archive of BSB → wayback-machine is used
- Application to „**Fachinformationsdienst Musikwissenschaft**“ (DFG-funded) → **Discussion:** What are the benefits of web archiving for musicologists?

# Research project I



## „Requirements from musicology for indexing and archiving Internet resources“

- Was carried out during Masterstudies „Library and Information Management“ at Hochschule der Medien Stuttgart: March – August 2014

### Questions:

- Which requirements do musicologists have for archiving and indexing websites?
- Websites are already used as scientific resources?
- What opinions do musicologists have about ViFaMusik?

# Research project II

## Method:

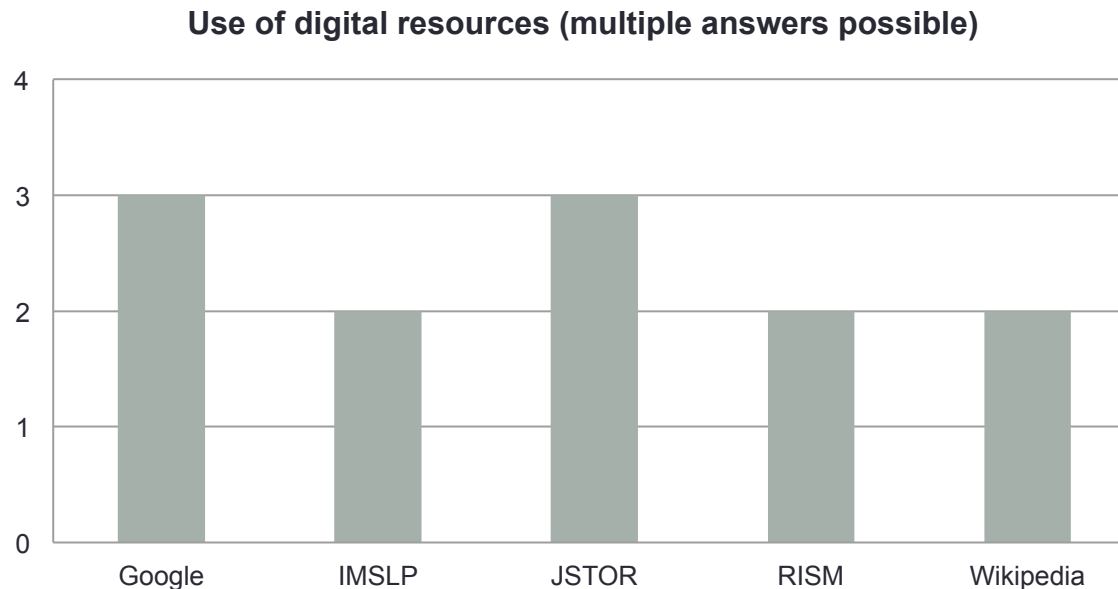
- Interviews with experts → 8 musicologists in Germany (doctor or professor)
- Project is not representative

## Results:

- All experts have heard of web archiving → different grades of knowledge
- 7 of 8 experts know ViFaMusik and its collections

# Research project III

- Experts still prefer printed resources
- Digital resources more and more used → dominance of data bases and search engines
- Experts cite webpages





# Research project IV

- **Experts communicate a need for web archiving:**

Aspects:

- Stability and sustainability of a web archive → longtime preservation of websites
- Indexing and possibility to search for websites → best: as a fulltext-archive
- Quality assurance for websites → content and availability

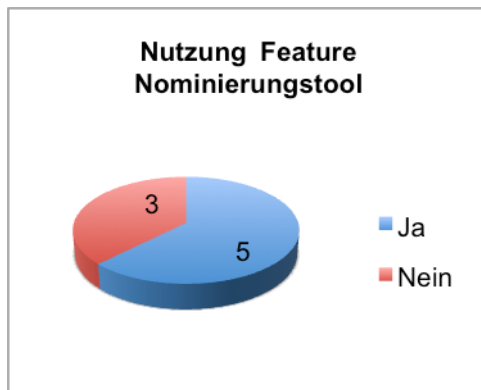
- **Experts criticize points of ViFaMusik**

- Too little content
- Too little thematic width
- Lack of transparency

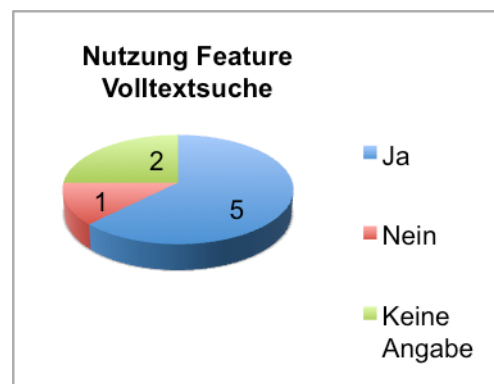
**Solutions?**

# Research project V

- Broader thematic width
- Better search functions
- Better mechanisms to organise content
- RSS-Feeds
- More activities to make ViFaMusik known in the community



Use of feature nomination tool



Use of feature full-text search

# Conclusion

- **A need for web archiving is seen**
- Use of digital resources is increasing → **websites need to be more accepted as a scientific resource**
- Using digital resources is a matter of generations → **Digital Natives**
  
- **Web archiving can help to establish websites as scientific resources** → keep it quotable
- ViFaMusic and its collection of websites are a useful solution → criticism needs to be solved

# References

- **Bayerische Staatsbibliothek (2014):** Website-Archivierung an der BSB. URL: [http://www.babs-muenchen.de/index.html?c=workflows\\_web&l=de](http://www.babs-muenchen.de/index.html?c=workflows_web&l=de), 03.08.2014
- **Bayerische Staatsbibliothek (2014a):** Langzeitarchivierung von Netzpublikationen. URL: [http://www.babs-muenchen.de/index.html?c=projekte\\_babs1&l=de](http://www.babs-muenchen.de/index.html?c=projekte_babs1&l=de), 03.08.2014
- **Brown, Adrian (2006):** Archiving websites. A practical guide for information management professionals. London : Facet Publishing
- **Lenk, Barbara (2015):** Studienprojekt „Anforderungen aus der Musikwissenschaft an die Internetressourcen-Erschließung und -Archivierung“ – überarbeitete Fassung. URN: <http://nbn-resolving.de/urn:nbn:de:bsz:900-opus-45384>
- **Liegmann, Hans, Rauber, Andreas (2010):** Webarchivierung zur Langzeiterhaltung von Internet-Dokumenten, in: Neuroth, Heike, Oßwald, Achim, Scheffel, Regine et al. (Hrsg.): nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung, Version 2.3, Kapitel 17.9., S. 88-103
- URN: **urn:nbn:de:0008-20100305365**
- **Masanès, Julien (Hrsg.) (2006):** Web archiving. Berlin [u.a.], Springer

# Questions?

# Thank you for your attention 😊

Contact: [barbara\\_lenk@web.de](mailto:barbara_lenk@web.de)