Sustainability strategies for digital humanities systems

Brigitte Mathiak (GESIS social science institute, Cologne, Germany)

Brigitte Mathiak is a senior scientist at GESIS (German institute for the social sciences). Before, she held a Junior Professorship for Digital Humanities at the University of Cologne and was speaker of the Data Center for the Humanities (DCH). In this position, she studied the sustainability of Digital Humanities projects as part of the DFG (Deutsche Forschungsgemeinschaft, German Research Foundation) project SustainLife. She holds a diploma in informatics.

"Why is sustainability an issue in digital humanities?" Mathiak and colleagues asked humanities scholars this question (Neuefeind 2020) and the most frequent answer was that there was no maintenance of websites at the end of their research projects. The average lifespan of digital scholarly editions is 8.5 years – in contrast to books, which survive for centuries without attention. But what is a digital scholarly edition? It is a book with annotations but online, usually text based. These websites often have a database or an XML repository to store the raw data, but interactive components such as search functions, text statistics and translations, and visualisation on top of the raw data are constructed in many different ways. All these requirements make interactive components extremely hard to preserve.

Figure 3: Results of the Kronenwett/Mathiak 2017 survey (see references)

Life spans of Digital Scholarly Editions

8,5 years on average
6 years half-life

Books survive for centuries without attention

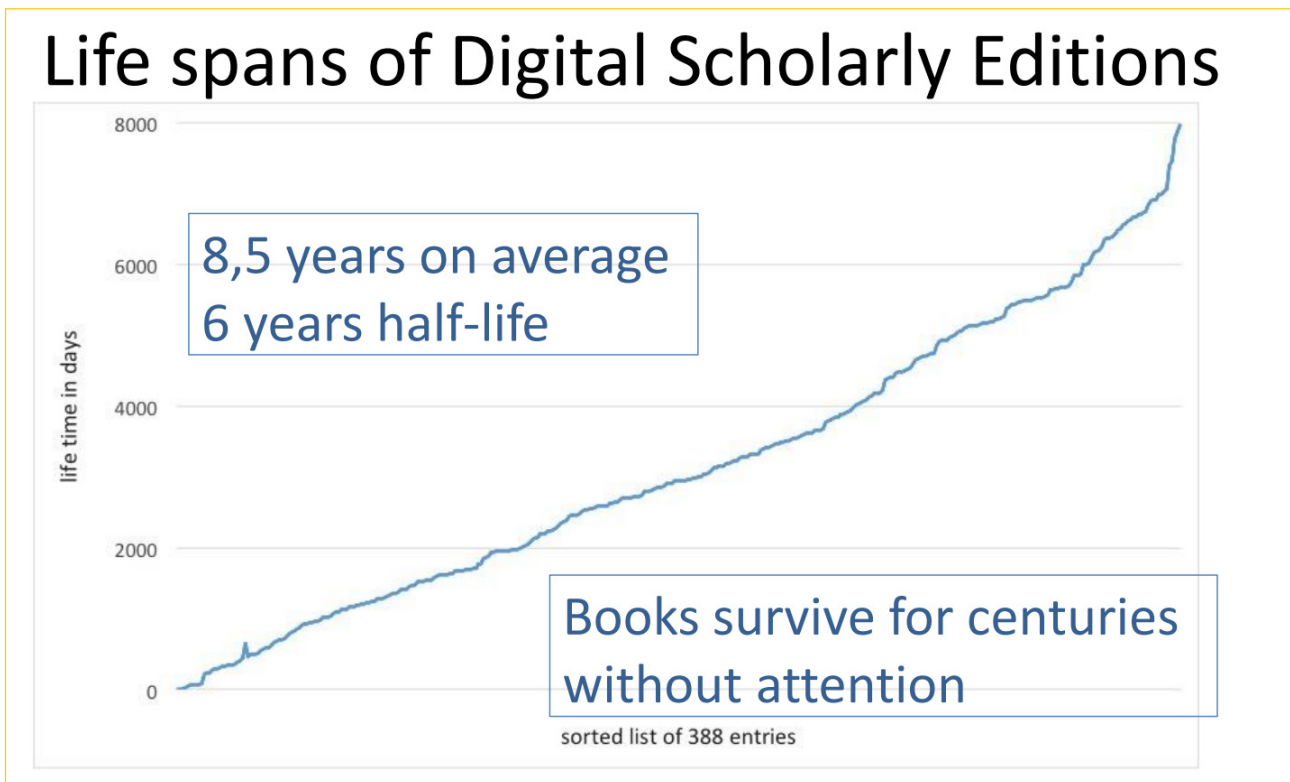sorted list of 388 entries

life time in days

Figure 4: Life spans of Digital Scholarly Editions (Neuefeind 2020)

Mathiak outlined several projects that deal with this problem (see figure 5). Some try to sustain the scholarly editions as living systems, reduce them to a smaller application with only standard access methods or move the old systems to newer platforms. These methods can be used retroactively on legacy projects, they are relatively quick to learn and to apply. Their application does improve maintenance costs significantly but inevitably re-design or shutdown is necessary to reduce security risks.

Other projects try to steer scholars away from standalone custom software and toward more modular, standards-based software frameworks like the Austrian GAMS (see references) to implement their digital assets which then will be maintained through the central system.

An interesting crossover solution is having the Internet Archive (https://archive.org) harvest a scholarly edition with adapted settings. Simultaneously, the database management system's content is preserved, which makes it possible to tentatively reconstruct the whole application based on its data and the application in the future. This approach involves low costs in the beginning.

Sustainability of digital projects is a challenge. It requires adequate funding to keep them fully functional. Cheaper options are web archiving (may only cost 5 minutes of your time) or a static HTML page (with or without Javascript), but not all projects can be transformed without loss and findability of the resource suffering.

„Sustainability Strategies for Digital Humanities Systems". Claes Neuefeind, Brigitte Mathiak, Philip Schildkamp, Unmil Karadkar, Johannes Stigler, Elisabeth Steiner, Gunter Vasold, Fabios Tosques, Arianna Ciula, Brian Maher, Greg Newton, Stewart Arneil, Martin Holmes. Panel at the ADHO Digital Humanities Conference 2020.

- Cheaper maintenance by bundling projects
  – SustainLife, DCH, University of Cologne
  – GAMS, Centre for Information Modelling, University of Graz
- Re-design to preserve
  – Humanities Computing and Media Centre, University of Victory, Canada
  – Lazarus project, CCeH, University of Cologne
- Sandbox
  – King's Digital Lab, King's College London
- Web Archiving
  – The Internet Archive a.k.a. the Wayback Machine

Figure 5: Mathiak's comparison of sustainability strategies. Representatives for each project associated by colour coding.

It is possible to use a combination of measures including tombstone pages[1] and archiving of the data and the code to mitigate this. These methods can be used retroactively on legacy projects. They are relatively quick to learn and apply and thus reduce maintenance costs significantly. But inevitably re-design or shutdown is necessary to reduce security risks.

Mathiak concluded that a multi-layered approach (e.g. King's Digital Lab, see references) and prefabricated environments like GAMS (see references) are good ideas. She also stressed that the people working in the projects are a key factor for preservation efforts and adequate funding is necessary. A good approach would be to think about individuals who might want to access the resource in 100 years' time (at the point of creation/design). It can be more helpful to think about them rather than current users.

Questions and discussion

Audience commented that these approaches are very focused on web archiving, but it should be noted there are other ways of tackling database preservation.

---

1   The term is used metaphorically for URLs signaling a former presence of web content that has been removed or transferred to another URL.