Preserving Qualitative Data: A Data Model to Prepare Computer Assisted Qualitative Data Analysis Software Data for Long-term Preservation

Umar Qasim
University of Alberta
Alberta, Canada
umar.qasim@ualberta.ca

Kendall Roark
Purdue University
Indiana, USA
roark.kendall@gmail.com

ABSTRACT

A rapid change in technology has a great impact on the long-term access to digital content. This makes preservation of a digital content a challenging task due to the content's inherent dependency on a specific hardware/software platform. Changes in the technology without backward compatibility can lead to a content that can't be viewed and qualitative data is no exception. Normalization is a commonly used strategy to keep content prepared for long-term preservation. However, tools are not always available to perform normalization on complex file formats such as qualitative data analysis software files. In this work, we are proposing a data model to normalize computer assisted qualitative data analysis software files to support long term access.

INTRODUCTION

Technological obsolescence is a known phenomenon and a number of strategies have been proposed to reduce the impact of software and hardware obsolescence including normalization, emulation and migration. File format normalization is one of the preservation strategies that is being widely discussed and used in the digital preservation community. In this strategy, digital objects of a specific type are converted into a single selected format, which is thought to have a higher chance of being accessible in the future. This strategy has been used successfully with simpler file formats like text, pdf, images etc. mainly due to the availability of software libraries for normalizing these types of files. One major limitation of this strategy is that there are a large number of file formats in use and not every file type has supporting libraries available for conversion purposes. An alternate way is to do this conversion using the original application by exporting or saving the desired content into an industry standard format. Unfortunately, this process is dependent on commercial vendors to provide such a support, which is not always provided. Data driven applications such as Computer Assisted Qualitative Data Analysis Software (CAQDAS) are one example of applications which store data in complex file formats and currently no libraries are available to do the normalization process. Some of these applications are proprietary, further complicating the situation because these vendors do not always provide support for converting files into a standard file format.

PROPOSED DATA MODEL

Under these circumstances, having a deeper understanding of data models in these complex data files helps in identifying essential pieces of information needed for future access. In this poster, authors are proposing a data model approach for CAQDAS applications, which can help to extract important pieces of information whereas any gaps are covered with necessary documentation. The proposed data model is intended to provide an approach to extract and preserve all the information, which is part of a CAQDAS application file, in a way that this information can be later assembled and viewed in any other current or future CAQDAS application. Currently, some of the major CAQDAS applications lack support for interoperability amongst various CAQDAS platforms. The proposed data model provides an alternate approach to make these CAQDAS applications interoperable.

To get a deeper understanding of the whole process, Roark (2015 forthcoming) conducted one on one interviews with researchers, and Qasim and Roark (2015) conducted both a pilot and a formal workshop on documenting and preserving CAQDAS projects at the University of Alberta. During the pilot and the formal workshop, the authors demonstrated how to take a CAQDAS project apart and capture all the important study documentation embedded in the project file. Participant feedback was solicited to improve the transformation process. In addition, current preservation strategies such as normalization, migration and emulation and the contexts in which each might be used were discussed. Preservation strategies for both proprietary and non-propriety software were discussed. Furthermore, current best practices and workflows for quality assurance and documentation (metadata, provenance, codebooks, scripts) were reviewed and as well as how to operationalize ethical and contractual commitments around data access and ownership into a data management plan and preservation practices.

CONCLUSION

In this poster, we are sharing the findings of our work on preserving qualitative research data and analysis documentation. We have proposed a data model driven approach for CAQDAS file preservation and provide guidance on how to extract data model from both proprietary and nonproprietary file formats.