SAFEGUARDING THE NATION'S DIGITAL MEMORY

A bayesian model of digital preservation risk for the disruptive digital archive

Alec Mulinder
The National Archives
United Kingdom
alec.mulinder@
nationalarchives.gov.uk
0000-0002-8900-3798

Sonia Ranade
The National Archives
United Kingdom
sonia.ranade@
nationalarchives.gov.uk

David H Underdown
The National Archives
United Kingdom
david.underdown@
nationalarchives.gov.uk
0000-0002-8123-4655

Abstract – This poster offers the iPRES community an interim update on work at The National Archives (UK) to develop a Dynamic Bayesian Network which describes and explains the complexity of the digital preservation risk environment. The intention of this project is to evaluate the applicability of a Bayesian statistical approach to support a risk-based approach to digital preservation. This would complement our existing standards-based approaches and support evidence led decision making by the archive.

This poster focuses on the technical work undertaken so far during the project, highlighting the potential benefits of the approach and identifying areas for further investigation.

Keywords – risk management, decision support, Bayesian Network, elicitation protocol, evidence base
Conference Topics – Exploring New Horizons.

I.   EXTENDED ABSTRACT

This poster offers the iPRES community an interim update on work at The National Archives (UK) to create a Bayesian statistical model of the digital preservation risk environment [1]. We discuss the requirements and potential benefits of the approach; describe our investigation of techniques for combining hard data with a knowledge base distilled from the experience of our archivists; present and review our initial experimental Bayesian network; identify areas for further development and highlight our plans for further work, in terms of technical development, building partnerships and growing the capability of the digital archive.

A.   Benefits of the Approach

We identify gaps in our current approaches to digital preservation and highlight how these may be addressed through application of the evidence based methods being proposed. The benefits of the method are discussed:

1.   *Complexity and Transparency:* The digital preservation environment is highly complex with multiple interdependent factors at work. It is difficult to understand and harder still to explain to stakeholders. The proposed model is well suited to expression as a graphical map of dependencies, events, actions and impact, which is accessible and relatively easy to communicate, understand and critique.

2.   *Prioritization:* The model quantifies risk exposure. We anticipate that this will allow us to compare and prioritize very different types of threats to the digital archive with potential impact in different areas.

3.   *Evidence Based:* The approach is data-driven, incorporating evidence from our observations and experience in the form of conditional probability tables within the Bayesian network. The model will also operate in areas where we lack hard data, via techniques for the elicitation of structured expert judgement to populate prior values for our probability distributions.

B.   Eliciting Expert Judgement

We offer an overview of our trial of a structured protocol for expert elicitation [2]. This is designed

iPRES 2019

to compensate for bias, be transparent, open to challenge and well documented. We describe our experience of working with the protocol to improve the quality of expert judgements being incorporated into the experimental model.

C. The Experimental Bayesian Network

We present a graphical representation of our experimental network as it currently stands. This highlights areas for further development, including the need to better reflect risk mitigation actions in the model; handling complexity; modelling failure events that arise from the same underlying factors; incorporating factors relating to the structure and information density of the records being preserved.

D. New Horizons

We highlight the potential of the tool for building a more inclusive digital archival practice for the future.

1. *A New Approach to Trust:* This 'bottom up' data driven approach offers an alternative to our current standards-based model of digital preservation. It is flexible enough to accommodate diverse contexts and different priorities, giving greater control to the archivist.
2. *An Inclusive Digital Archive:* This is primarily a structured and evidence based model, which also makes space for expert judgement for a more nuanced approach to decision making. It can help us analyze our experience, decisions and reasoning to reduce bias and improve transparency and accountability. This will start to narrow the gap between the technical work of the digital preservation function and the archive's wider role within our society.
3. *New Partnerships:* The approach is potentially greater than the sum of its parts. It will operate most effectively when digital preservation specialists pool our knowledge, expertise and data to share insight and experience across archival institutions and the wider digital heritage community.
4. *Building Capability:* This project starts to put quantitative modelling techniques into the hands of archivists, helping us to develop the skills and thinking we will need to build the next generation of disruptive digital archives [3].

REFERENCES

[1]   S. Wright, M Barons, R Wilkerson, J Smith, "Report of National Archives workshop for modelling digital preservation using Bayesian Networks", University of Warwick, TNA internal report, November 2018.

[2]   M Burgman, A Hanea, M McBride, B Wintle, "A practical guide to structured expert elicitation using the IDEA protocol", Methods in Ecology and Evolution, 9, 1, 2018.

[3]   The National Archives, "Digital Strategy", March 2017, www.nationalarchives.gov.uk/documents/the-national-archives-digital-strategy-2017-19.pdf

POSTER

16th International Conference on Digital Preservation iPRES 2019, Amsterdam, The Netherlands.

2

iPRES 2019