

## **A collaborative digital preservation risk model**

**DAVID UNDERDOWN, ALEX GREEN, and HANNA MERWOOD**

**Abstract:** The National Archives (UK) has been leading a collaborative project with five other English archives, and statisticians from the University of Warwick's Applied Statistics & Risk Unit, to develop DiAGRAM: the Digital Archiving Graphical Risk Assessment Model.

DiAGRAM's foundation is a Bayesian Network - a statistical model that estimates the probability of outcomes by considering conditional events and changes over time (e.g. the lifespan of our archival storage depends on a changing mix of storage media types). Bayesian networks have been used as a foundation for decision support tools in a variety of contexts including aviation, credit scoring, and food security, and the technique is widely used in risk assessment.

All partners were involved in a series of workshops that first mapped out the network of risks and the interactions between them, identified potential data sources, and took part in the expert elicitation process (using the IDEA protocol) facilitated by our statistical partners.

DiAGRAM, will:

- Improve users' understanding of the complex digital archiving risk landscape and of the interplay between risk factors.
- Empower archivists to compare and prioritize very different types of threats to the digital archive: from software obsolescence to natural disaster.
- Aid in quantifying the impact of risk events and risk management strategies on archival outcomes to support decision making, communication with stakeholders and developing business cases for targeted action.

We will explore the pros and cons of the collaborative process and the issues that arose.

Supported by the National Lottery Heritage Fund and EPSRC Impact Acceleration Account.

### **About the authors:**

*David Underdown* (main speaker), Senior Digital Archivist, joined The National Archives in 2005 as a database administrator and soon gained his introduction to digital preservation from supporting our PRONOM registry of file formats and involvement in projects to refine and update our digital repository system. Since his background (a degree in mathematics from Imperial College London and several years working in systems development for a life and pensions company) had not really prepared him for working in archives, he used a general interest in First World War history to develop his experience of archival research and archival theory. David is also involved in defining image and metadata specifications for digitisation projects such as First World War Unit Diaries and 1921 Census. His current research project sees a return to his mathematical roots, applying Dynamic Bayesian Networks to

modelling digital preservation risk through the National Lottery Heritage Fund supported project [‘Safeguarding the Nation’s Digital Memory’](#).

**Alex Green** is the Service Owner for Digital Preservation at The National Archives and is the project lead. She is an experienced Digital Archivist having worked on creating user-centric digital tools and services for the past twenty years.

**Hannah Merwood** is a Research Assistant in Applied Statistics at The National Archives (UK) on secondment from the Department for Digital, Culture, Media and Sport. She holds a bachelor’s degree in mathematics and statistics from the University of Oxford and is a member of the Government Operational Research Service analytical profession. Hannah has experience of developing complex models and data tools to support decision makers within government, including for prisoner escort contracts and broadband delivery programmes.