



# Rosetta Advisory Group 8<sup>th</sup> Annual Meeting

June 13-14, 2017  
Sheffield, UK

## Tuesday, June 13<sup>th</sup>

- 08.30 – 09.00      **Registration & Coffee**
- 09.00 – 09.40      **Welcome Greetings on Behalf of the University of Sheffield**, Anne Horn, Director of Library Services, The University of Sheffield
- Welcome Greetings on Behalf of the Rosetta User Group**, Steve Knight, Programme Director Preservation Research & Consultancy, National Library of New Zealand and Chair of the Rosetta User Group
- Welcome Greetings on Behalf of Ex Libris**, Ofer Mosseri, Corporate VP and General Manager Europe, Ex Libris
- Ex Libris Corporate Update**, Bar Veinstein, President, Ex Libris
- 09.40 – 10:30      **Rosetta Product Update and Roadmap**, Dvir Hoffman, Corporate VP Resource Management Solutions and Adi Alter, Director of Product Management, Digital Solutions, Ex Libris
- 10.30 – 10.50      Coffee Break
- 10.50 – 11.40      **Fake News and Dark Ads: Why Digital Preservation Has Never Mattered More**, Richard Ovenden, Bodley's Librarian, University of Oxford, and President, The Digital Preservation Coalition (DPC)
- 11.40 – 12.00      **Preserving Data Across the Research Lifecycle: Integrating Rosetta with figshare**, Angus Taggart, Library Systems Manager – Research and Jez Cope, Research Data Manager, University of Sheffield
- 12.00 – 12.20      **Getting your Rosetta Digital Archive Certified – Extended Certification, Self-Audit on the Basis of ISO 16919:2014 / DIN 31644**, Yvonne Tunnat, Preservation Manager, German National Library of Economics (ZBW)

- 12.20 – 12.40     **Updated Workflows for Ingesting and Exporting Rosetta Content**, Chris Erickson, Digital Preservation Manager, Brigham Young University
- 12.40 – 13.40     **Lunch**
- 13.40 – 14.10     **How Valid is your Validation? JHOVE as the Go-to-Validator within Rosetta**, Michelle Lindlar, Technical Analyst, German National Library of Science and Technology (TIB)
- 14.10 – 14.40     **Handling File Format Issues in Rosetta – and Beyond**, Franziska Geisser, Data Archivist, ETH Zurich
- 14.40 – 15.10     **Web Archiving and Rosetta - Achievements and Challenges**, Tobias Beinert, Deputy Head of Long-Term Preservation and Data Curation, Bavarian State Library and Steve Knight, Programme Director Preservation Research & Consultancy, National Library of New Zealand
- 15.10 – 15.35     **Workflow Automation with Rosetta – Experience and Challenges**, Dr. Henrike Berthold, Head of the IT Department, and Dr. Gerald Hübsch, Head of the Digital Infrastructure and Long-Term Preservation Division, Saxon State and University Library Dresden (SLUB)
- 15.35 – 15.55     **Coffee Break**
- 15.55 – 16.40     **Digital Building Blocks: Rosetta's Plug-and-Play Ability**, Euwe Ermita, Manager, Digital Strategy & Innovation, State Library of New South Wales and Teresa Soleau, Head, Library Systems & Digital Collection Management, Getty Research Institute
- 16.40 – 17.00     **IIF and Rosetta**, Opher Kutner, Rosetta Business Analyst, Ex Libris
- 17.00 – 17.30     **Let's Talk Shop – a Lightning Talk Session Covering Digital Preservation Practice In and Outside of Rosetta**, Michelle Lindlar, Technical Analyst, German National Library of Science and Technology (TIB)
- 18.00             **Drinks Reception and Dinner**

## Wednesday, June 14<sup>th</sup>

- 09.00 – 10.00     **Keynote Presentation: Change is Here to Stay: What Does It Mean and What Do We Need to Do About It?**, William Kilbride, Executive Director, The Digital Preservation Coalition (DPC)
- 10.00 – 10.40     **Rosetta Operations Highlights**, Idan Soffer, Director of Support, Ex Libris
- 10.40 – 11.00     Coffee Break
- 11.00 – 11.45     **Bringing It Together: How Rosetta Is Developed and Deployed**, Josh Weisman, VP Development, Resource Management, Ex Libris
- 11.45 – 12.00     **Rosetta Advisory Group Updates**, Adi Alter, Director of Product Management, Digital Solutions, Ex Libris
- 12.00 – 12.45     **Rosetta User Group Annual Update**, Steve Knight, Programme Director Preservation Research & Consultancy, National Library of New Zealand and Chair of the Rosetta User Group
- Open Discussion**, Rosetta User Group members and Ex Libris management
- 12.45 – 13.00     **Wrap-up and Concluding Statements**, Adi Alter, Director of Product Management, Digital Solutions, Ex Libris
- 13.00 – 14.00     **Lunch**

## Presentation Abstracts

### **Fake News and Dark Ads: Why Digital Preservation Has Never Mattered More**

Richard Ovenden, Bodley's Librarian, University of Oxford, and President, The Digital Preservation Coalition (DPC)

Richard Ovenden will discuss the importance of digital preservation in an era of VUCA: Volatility, Uncertainty, Complexity and Ambiguity. He will explore the current reach of digital preservation, and argue that the institutions and organisations that care about digital preservation must extend beyond the current range of content and technology to ensure the health of an open society.

### **About Richard Ovenden**

Richard Ovenden is Bodley's Librarian, the 25th person to hold the title, which is the senior executive position of the Bodleian Libraries. Richard was educated at Durham University and University College London, and has worked as a professional librarian since 1985. He has served on the staff of Durham University Library, the House of Lords Library, the National Library of Scotland (as Deputy Head of the Rare Books Section), the University of Edinburgh (as Director of Collections), and since 2003 at the Bodleian Libraries (first as Keeper of Special Collections, from 2011-2014 as Deputy Librarian, the Bodleian Libraries, then from 2014 as Bodley's Librarian).

Richard sits on the Board of Research Libraries UK and on the Board of the Consortium of European Research Libraries, is a Trustee of the Krazsna Kraus Foundation, of Chawton House Library, and of the Victoria County History for Oxfordshire, and is currently President of the Digital Preservation Coalition. He has published widely on the history of collecting, on the history of photography, and on professional concerns of the library, archive, and information world, and is a Fellow of the Society of Antiquaries. He was elected to the American Philosophical Society in 2015. Recently Richard headed Oxford's involvement with the Google mass digitization project. He holds a Professorial Fellowship at Balliol College, Oxford.

## **Preserving Data Across the Research Lifecycle: Integrating Rosetta with figshare**

Angus Taggart, Library Systems Manager – Research and Jez Cope, Research Data Manager, University of Sheffield

The University Library at the University of Sheffield is taking the leading role in supporting the active management and curation of research data within the institution. We have recently implemented a research data catalogue and repository, ORDA (Online Research Data, powered by figshare for institutions). We have also begun safeguarding library collections and key administrative assets of the University using Rosetta.

We are now working with figshare and Ex Libris to integrate both services in order to provide seamless preservation of published research data across the research lifecycle. In the longer term, this work will enable us to provide complete lifecycle data management services for the university's research community; to identify, understand and act on risks associated with preserving data sets; to provide better-informed advice and guidance around use of data formats for sharing and preservation purposes; and to encourage researchers to share their data more openly with others by guaranteeing the long-term sustainability of that data.

## **Getting your Rosetta Digital Archive Certified – Extended Certification, Self-Audit on the Basis of ISO 16919:2014 / DIN 31644**

Franziska Schwab, Preservation Analyst, German National Library of Science and Technology (TIB) and Yvonne Tunnat, Preservation Manager, German National Library of Economics (ZBW)

The European Framework for Audit and Certification of Digital Repositories (2010) distinguishes three levels of certification: *Basic certification* ([Data Seal of Approval](#), self-audit), *Extended certification* (Self audit based on ISO 16919:2014 / DIN 31644) and *Formal certification* (external audit on the basis of ISO 16363 or DIN 31644).

This presentation focuses on the second level of certification, with a strong focus on Rosetta as the digital preservation software, using the [Nestor Seal](#) application of the ZBW as a practical example.

## **Updated Workflows for Ingesting and Exporting Rosetta Content**

Chris Erickson, Digital Preservation Manager, Brigham Young University

More library and university departments are choosing to preserve their content in the Rosetta system. In order to streamline the digital workflows and submission processes, we created a shared directory to allow producers to transfer digital content to the Rosetta infrastructure at the end of their workflow, which is then ingested directly into Rosetta.

Content exported out of Rosetta is also placed in the shared directory. Access files that are missing or corrupted are replaced based on Rosetta preservation masters. Digital content intended for the Digital Preservation Network (DPN) goes through a new vetting process to determine the most important files to preserve in the national network. We have chosen to add only the most important files to the DPN, and to keep it to 5 TBs per year. Digital content intended for the DPN is exported to a shared folder that is then immediately ingested into the DPN.

## **How Valid is your Validation? JHOVE as the Go-to-Validator within Rosetta**

Michelle Lindlar, Technical Analyst, German National Library of Science and Technology (TIB) and Yvonne Tunnat, Preservation Manager, German National Library of Economics (ZBW)

Validation is a key task of any preservation workflow and JHOVE is the tool of choice within Rosetta for characterizing and validating several file formats. While the tool was first released over 12 years ago, maturity by age and widespread use do not automatically translate into fail-proof results.

While JHOVE can be used as a stand-alone tool, e.g. for pre-ingest analysis, the integration into Rosetta allows for rule-based automatic decision-making. However, adequate rules must be based on a profound understanding of the validity of JHOVE's output. TIB and ZBW have been collaborating on an analysis of JHOVE output, focusing on three widely used modules: PDF, TIFF and JPEG. This presentation gives an overview of the analysis, describing the quality, intelligibility and usability of the three modules. We will conclude by discussing the implications of the outcome for the Rosetta validation stack and by highlighting further work being undertaken to improve the quality of JHOVE.

## **Handling File Format Issues in Rosetta – and Beyond**

Franziska Geisser, Data Archivist, ETH Zurich, ETH Library

While the processes of format identification, validation or technical metadata extraction are reasonably well understood in principle, a number of issues crop up in everyday practice. They require a thorough evaluation followed by a decision on how to proceed, keeping in mind overall efficiency and not compromising future preservation options. The practical implications of these issues differ with every use case, data producer, type of material and format.

Which methods and tools can help us analyse specific cases and make an educated decision? How can we handle such file format issues in Rosetta? Can we always rely on what the validation tools tell us? In which cases is an immediate action required, and when should preservation actions be postponed? In examining some concrete examples, this presentation highlights some of the considerations made and conclusions reached at the ETH Data Archive operated by ETH Library, the main library of ETH Zurich.

## **Web Archiving and Rosetta - Achievements and Challenges**

Tobias Beinert, Deputy Head of Long-Term Preservation and Data Curation, Bavarian State Library and Steve Knight, Programme Director Preservation Research & Consultancy, National Library of New Zealand and Chair of the Rosetta User Group

Web archiving has become a daily routine for many memory and research institutions around the world. This means that a lot of our digital cultural heritage is already preserved for future generations.

However, as the web archiving community is currently primarily focused on increasing and improving access to the archived content, new challenges arise for keeping the existing tools for web harvesting up-to-date. This presentation will give insights into the motivation, workflows and infrastructures for web archiving at the National Library of New Zealand and the Bavarian State Library and how they integrate with Rosetta. Challenges of the daily practice will be covered as well as short-term and long-term perspectives for future developments in web archiving.

## **Workflow Automation with Rosetta – Experience and Challenges**

Dr. Henrike Berthold, Head of the IT Department, and Dr. Gerald Hübsch, Head of the Digital Infrastructure and Long-Term Preservation Division, Saxon State and University Library Dresden (SLUB)

SLUBArchiv, the digital long-term archive of the Saxon State and University Dresden, considers the automation of workflows in digital long-term preservation as a key enabler for large scale digitalization projects that produce millions of digital objects per year. Even multiple institutions with heterogeneous production environments can be supported without sacrificing the quality requirements of the archive. As a result, efficient deployment of personnel and the provisioning of digital long-term preservation as a service have become possible.

Rosetta web services and the ability to customize the internal processing of digital objects by Rosetta are essential for workflow integration and custom quality assurance. This talk introduces the digitalization workflow and tool chain developed and operated by SLUB Dresden. Special focus will be put on the Submission Application, a tool developed by SLUB for processing pre-ingest and access batches that interfaces with Rosetta via web services. The implementation of our integration approach, experiences and challenges, e.g. when a Rosetta instance must integrate with heterogeneous producers of digital objects, will be presented in order to provide useful insights for Ex Libris and the community.

## **Digital Building Blocks: Rosetta's Plug-and-Play Ability**

Euwe Ermita, Manager, Digital Strategy & Innovation, State Library of New South Wales and Teresa Soleau, Head, Library Systems & Digital Collection Management, Getty Research Institute

The State Library of New South Wales recently undertook major digital transformation initiatives, replacing core library and corporate systems.

Successfully integrating a system like Rosetta into a complex digital ecosystem is challenging. This presentation will focus on how pluggable Rosetta is out-of-the-box, and the customisations needed to co-exist within the Library's digital environment.

The presentation will be followed by a more general discussion of the use cases and requirements for integrations with archival management systems including workflows, infrastructure and integration points.

## **IIIF and Rosetta**

Opher Kutner, Rosetta Business Analyst, Ex Libris

A preview of forthcoming (version 5.3) IIIF native support in Rosetta and some impressions from the [2017 IIIF Conference](#), which takes place just a week prior to the RAG meeting.

## **Let's Talk Shop – a Lightning Talk Session Covering Digital Preservation Practice In and Outside of Rosetta**

Michelle Lindlar, Technical Analyst, German National Library of Science and Technology (TIB)

Do you feel like you're alone in the world with those pesky PDF files? Are you currently thinking about ways to ingest new materials? Has IT approached you about an infrastructure change and you'd like to hear how other institutions have approached it? Do you have something great to share but it's not quite ready for a full presentation yet? This is the session to present work-in-progress, pressing questions and novel ideas!

Possible topics can include, but are by no means limited to:

- Staffing
- System architecture
- Pre-ingest work
- A specific content workflow
- In-house developments
- Digital preservation community participation
- Current research

All attendees are invited to participate and can do so in the following ways:

- Give a lightning talk (1-5 minutes, 1-2 slides)
- Add a question to the forum
- Participate in the open discussion

Do you have something you would like to share in a lightning talk? Do you have a question you would like to ask the RAG crowd and see discussed during this session? Please contact the session organizer at [michelle.lindlar@tib.eu](mailto:michelle.lindlar@tib.eu).

## **Keynote Presentation: Change Is Here to Stay: What Does It Mean and What Do We Need to Do About It?**

William Kilbride, Executive Director, The Digital Preservation Coalition (DPC)

Digital preservation is a constantly emerging challenge. Just as we think we have solved it, some new technology or tool comes along, sending us back to the drawing board. For digital preservation, technological innovation is not only the cause of our anxieties but also the promise of a solution, because when we return to the drawing board it is due to some new tool or technique that has become available to us, and which improves, in some small way, our ability to tackle the underlying problem of obsolescence and documentation. We well understand the need to renew the technical infrastructure, and we've grown to expect rapid innovation and prototyping. Everything is always in beta. We're also used to the idea of a community: that one does not simply procure a product but increasingly one procures a community as well. But perhaps we are less well adapted to the social and organizational change that this implies. How do we renew and adapt the social infrastructure which surrounds digital preservation, and what does this mean for us as a community?

### **About William Kilbride**

Dr William Kilbride is Executive Director of the Digital Preservation Coalition (DPC), a not-for-profit membership organization providing advocacy, workforce development, capacity and partnership in digital preservation. William started his career in archaeology in the 1990s when the discipline's enthusiasm for new technology outstripped its capacity to manage the resulting data. He joined the DPC from Glasgow Museums where he was Research Manager and before that was Assistant Director of the Archaeology Data Service at the University of York. He was previously a lecturer in archaeology at the University of Glasgow, where he retains an honorary position.

## **Bringing it Together: How Rosetta is Developed and Deployed**

Josh Weisman, Rosetta Development Director, Ex Libris

The world of application development is constantly in flux, and the Rosetta team is not immune from those changes. In this session we will take a look at how we develop Rosetta, from design to development to testing to documentation, and we will learn how the Rosetta team is leveraging the latest in best practices, from continuous integration to developing open source plugins on Github. There is a lot that goes into each Rosetta version and this session will provide a behind-the-scenes glimpse.