2017

DLM FORUM 8th TRIENNIAL CONFERENCE

Wednesday 13th September

Database Preservation Workshop

Luis Faria
Bruno Ferreira

This workshop focuses on presenting the current state-of-the-art relational database preservation standards and tools used by major national archives and other institutions. It presents SIARD 2, a new preservation format for relational databases. It also presents the Database Preservation Toolkit, a new tool for harvesting information from live database management systems (e.g. Oracle, MySQL, Microsoft SQL Server) into SIARD format and back.
Finally, it presents the E-ARK database viewer, a tool to access and view the information preserved in SIARD-files. The workshop includes live demonstration of the tools and prompts the participants to use them on their own laptops using the demonstration databases provided.

Archives and Data Lakes – can two become one?

Nina Bryant
Clare Sadler

We explore the differences, and similarities, between archives and data lakes and in the context of developing an archiving approach within the Financial Services environment, consider the question 'can two become one'. We examine the definitions, requirements and use cases associated with each, and illustrate this at the conceptual architecture level. We identify that whilst both draw on much of the same data, the requirements of archives and data lakes are often at odds and given current technology, bringing the two together right now would probably result in unacceptable cost and performance implications. We consider governance and find that information and data governance are converging and discuss other organisational aspects that will influence an archiving approach. We conclude that with rapidly evolving technology and the convergence of governance across data and information, two could indeed become one at some point in the future. Whether an organisation chooses that as an approach or not, there are beneficial steps that could be taken regardless of the approach chosen.

Chain Informatisation and the challenges for governance

Mies Langelaar
Ronald Rommelse
Jacqueline Schuurman Hess

The Municipal archives of Rotterdam have been working on a trusted repository for digital archiving during the past 10 years.
In 2016 for the first time a fully digital archive was ingested in the repository: the archives of the City Region Rotterdam Rijnmond. The city region of Rotterdam was a district in South Holland. The city region included fifteen municipalities. The area formed about the agglomeration of Rotterdam, but stretched further eastwards. In 2015 the City Region was dissolved and was absorbed by the

Rotterdam-The Hague metropolitan area. The archives should be deposited at the Municipal Archives of Rotterdam, but had to be accessible to the former partners.
This paper will focus on the problems we met in taking in the digital born archives (metadata and the mapping of metadata, formats, exporting from the original application) and in ensuring availability (how can the data be accessed from the applications of the former participants in the City Region). Furthermore the focus will be on the requirements this type of archives imply for the design of systems and the need for (European) standards.
Building on the experience with the City Region, which was a very basic form of chain informatisation, we are now looking at the governance of chain informatisation within the social sector. Main focus with this part of the paper will be the governance of informatisation, the requirements this poses to systems and workflows (privacy by design) and the way the inspection and the auditing of archives and workflows can be provided within chain informatisation.

Information management: Using Standards to promote regulatory compliance and effective exploitation

Alan Shipman

Alan will talk about how National and International Standards are developed according to business needs, particularly in the area of information management and data protection. This will include details of the new BS 10012 which deals with data protection issues related to the GDPS and Brexit. He will also describe a new International Project which aims to create a 'privacy add-on' to the Information Security International Standard ISO/IEC 27001. Volunteers are sought to ensure that this new project needs business needs, which include the possibility of formal certification in the 'processing of personal information' field.

The awakening of information governance in Belgium: from eIDAS to the Digital Act

Fiona Aranguren Celorrio
Laurence Maroye
Sébastien Soyez

In 2014, the European Union adopted a new regulation concerning electronic identifications and trusted service providers with the aim of increasing trust in e-commerce and e-government across Europe. Surprisingly, the new eIDAS framework had a major positive awakening effect in many private and public organisations. The professional sectors, which up to that moment did not seem concerned by issues related to Information Governance, Records Management or archival principles, are now realizing how vital these fields are if we are to build a trustful and interoperable digital environment.
Following the examples of some European peers, Belgium is working hard to progress in these matters, its Digital Act of July 2016 being the best example. Aiming at analysing this fast-evolving context, the HECTOR research project was launched in 2014 and was presented in the DLM-Forum 7th Triennial Conference the same year. Since then, its members have analysed the issues and opportunities related to the digital transition in Belgian federal administrations in order to propose guidelines in different forms. Furthermore, this work allowed the research team to promote and

support many initiatives which should reinforce the role of Information Governance in general, and Records Management in particular: 1) participation to the working group in order to elaborate an implementation act for the Digital Act regarding electronic archiving; 2) creation of a Belgian mirror committee for ISO TC46/SC11; 3) publication of an academic book on Records Management; and 4) publication of online guidelines on Records Management and digital preservation.
In this article, the authors intend to explain the Belgian approach to building the foundations of an Information Governance legal and technical framework and compare it with more mature examples to show the current European landscape on the matter and to draw possible trends for the near future.

Thursday 14th September

Workshop on Lifecycle Management & Digital Preservation using Blockchain Technology

Kuldar Aas
Victoria Lemieux
Raivo Ruusalepp
Lars Hansen
Ross King
Hrvoje Stancic

Blockchains and distributed ledger technology promises trusted and immutable records in a wide variety of use cases involving recordkeeping, including real estate and healthcare. This workshop explores ongoing research relating to digital records life cycle management and preservation in the context of blockchain record keeping ecosystems. The workshop will encourage discussion of the capability of innovative blockchain-based systems to deliver trustworthy record keeping and long-term preservation and access to authentic records based on theories, principles and practices of the archives and records research and practitioner communities. The objective of the workshop will be to bring these communities together to discuss open issues, the current state of research, and to build a community of shared research and practice relating to blockchain record keeping and digital preservation.

IT governance of Dutch municipalities and digital information management

Jeanine de Gier

Digitalisation of information management changes the archives sector quickly and has strong implications for information processes within governmental organisations. In this paper, the governance of such organisations is looked at from a business-IT alignment driven perspective, based on a case study conducted within two Dutch municipalities. The study shows that mature IT governance can have positive effects on digital documental information management, as it leads to “desirable behaviour in the use of IT” [1]. Comparing the two municipalities, several findings surface. These concern effective management of input, for example through successful cooperation, and effective management of output of IT, for example by providing the needed digital services to

citizens. Findings also regard the conjoint development of strategies by business and IT, mainly concerning the involvement of IT in decision making. Finally, findings surface concerning the positioning of decision making authority, and whether that is clear to the personnel.

Information Value as a Dynamic Concept: Moving Towards a More Agile Records Management

Elisabeth Bühlmann Herzog
Yves Marleau

Within organizations, appraising records is a complex process in order to adequately preserve critical information, meet compliance requirements while still providing timely access to this information for the employees. Today, the exponential growth of digital information, the added responsibility of records selection on end-users and the fragmentation of the traditional "archival file" lead to rethink the records appraisal process.

This presentation will report on a research project aiming to better understand the concept of 'value' for the appraisal of digital information. It will introduce a model of information value criteria based on factors such as the context of creation and use of the information resources, the initial value perceived by the organization, the depreciation rate over time, the different compliance requirement, like EU GDPR implementation and the frequency and access by users types.

The presentation will demonstrate why information value is dynamic and how the criteria derived from the model can further be used to automatically calculate the value of different information resources through time. This approach would lead to more agile records management practices where information can be effectively re-assess at specific intervals to better answer legal, financial and administrative requirements.

The record system loophole

James Lappin

Society has a way of ensuring that organisations with very strong accountability requirements (because of their political, economic, environmental or safety importance) keep records for as long as society needs them to. Society expects such organisations:
- To maintain a records retention schedule, that lists the organisation’s activities and identifies how long records are kept of those activities
- To ensure that no records are deleted from their record system(s) except in accordance with those rules.

This compliance expectation has only one weakness, but it is a big weakness.

The weakness is that there is no generally accepted definition, in law or in records management literature, of what constitutes a records system. This gives organisations a degree of latitude in deciding which of its applications will be treated as record systems and which of its applications will

not be treated as record systems. The vehicle by which they make, and justify, this distinction is their records management policy.

As individual recordkeeping practitioners, we work to benefit our organisations. This loophole has proved very beneficial for many of the organisations that we work for. It has enabled them to deal economically with their highest volume application - their email system - through the simple means of exempting their email system from the retention rules for their business activities and instead applying a blanket routine deletion policy to all correspondence.

We are professionals. We work for the benefit of our organisations, but we also have a professional voice. Professions have to be careful with loopholes. Accountants have to stay on the right side of the dividing line between tax efficiency and tax avoidance. We have to stay on the right side of the dividing line between timely disposal of trivia and wholesale destruction of records.

This presentation will explore the following questions:
- If we continue to make a distinction between 'record systems' and 'non record systems', to what extent do we have an obligation to ensure that the applications we define as being 'record systems' consistently and routinely capture records from any applications that we define as not being a record system?
- To what extent does our profession have an obligation to help society take an informed view on how organisations are applying (or not applying) their retention schedules?
- The record system loophole depends for its justification upon a fudged definition of what constitutes a record system. To what extent is this lack of clarity hindering the development of records management thought and practice?

Risk Management for managing Art Archives, Authenticity and Cultural Memory

Marja van der Made

Many organizations and institutions have art collections. As these items mostly increase in financial value, accountability to the public, cultural memory and corporate social responsibility to manage these items responsibly must be ensured. Yet what are the responsibilities? Which regulations play a role? Which regulations regarding authenticity, artist's rights and archiving apply to managing artworks? Who has the right to authenticate and what are the market rules? Where does Digital Asset Management fit in? How can it be ensured that the items are authentic? Organizing corporate cultural memory may require considerable investments in archiving.

The focus of the paper is on risk management, management, protection and safeguarding of the cultural and governance aspects of archives for art and special collections. The archives and library are important sources of information for research, evidence and authenticity certification. What is the role of the archives in authentication? Which documents should special collections keep? This paper looks at the risk management process and its implementation so as to ensure compliance with regulations and to promote a positive corporate image.

Beyond the boundary of custody

Ruud Yap
Wim Westland

What the National Archives of the Netherlands would like to introduce is a reflection on how the National Archives of the Netherlands envisages its new and broadened role in a digital world and what lessons were learned. The National Archives of the Netherlands has not only built a new and customized infrastructure for recordkeeping, it has taken the opportunity to refocus, to rethink its business, and to create new services beyond the boundary of custody. As such, the digital repository is a driver for change, and innovation, both in procedures, in skills and attitudes of professionals, in relationships with government agencies, and in technological solutions for traditional archival issues.

Democratisation of appraisal - How civil society can be involved in the archival appraisal process

Krystyna Wanda Ohnesorge
Franziska Brunner

The digital revolution is not only altering the way we work, but also transforming the requirements and expectations of our clients. These will affect our tools and our digital collection, as well as the core functions of archival responsibility, for example the appraisal. The appraisal is a process usually conducted by an archivist to select records for permanent preservation. Since public archives are traditionally part of a state's or a regional's administration the responsibility for choosing the relevant information could then be considered as a demonstration of state power, which would be dangerous regarding the neutrality, authenticity and completeness of archived records. Moreover the claim of "building a society's memory" is often a mismatch regarding to the general lack of involvement in appraisal decisions of different parties of a civil society. Finally, archivists are constantly challenged by the increasingly complex political, economic and social processes, which require highly specific knowledge in a broad field of topics to fulfil the decision-making in appraisal. That is why we can no longer solely perform the appraisal process.
Nowadays the Swiss Federal Archives (SFA) and the records producer decide together which documents of the Swiss federal state will be archived or not. This involvement of records producers in decision-making process is a legal requirement, which was formalised with the inauguration of the Federal Act on Archiving in 1998. Since then this approach has been implemented continuously at the SFA and was finally institutionalised in an integrated overall concept of appraisal in 2010.
We enhanced our concept of appraisal by 2013 during the Project Ellipse in cooperation with the Federal Office of Topography (swisstopo). The reason for the modification is directly related to the provenance of the geodata. As geodata is often created by one administrative office, it is then re-used by a variety of governmental and non-governmental stakeholders as a basic layer for their new products. Consequently, the appraisal process has to be done in collaboration not only with the administrative unit that created the data – as we did so far – but also with a larger circle of offices using geodata. The open dialogue with the geodata producers and the early involvement of additional knowledgeable stakeholders has proven to be a valuable addition to our fundamental concept of appraisal.

Last year, we started with the development of a new "participatory appraisal". We will open the appraisal for other topics, such as environmental issues or national infrastructure. In doing so, we wish to include not just the different records producers and experts, but also the participating and affected parties of the civil society in the moderated appraisal process. Our goal is to consider the varied interests and reach a consensual development. We believe that the broader participation of various stakeholders in the appraisal process will further increase the confidence and trust between the Swiss federal authorities and civil society.
During our presentation, we will report on our new approach for the appraisal and our initial experiences.

Friday 15th September

RODA 2.0 and the Publications Office of the European Union case study

Luis Faria
Fulgencio Sanmartin

RODA (Repository of Authentic Digital Records) is a long-term digital repository solution that delivers functionality for all the main functional units of the
OAIS reference model. RODA is capable of ingesting, managing and providing access to the various types of digital content being produced from small to large corporations and public bodies. RODA 2.0 release was two years in the making and marks a complete overhaul of the system focusing on performance and flexibility:
– Performance greatly improved with a more monolithic design;
– New design is cleaner and easier to use;
– Customizable descriptive metadata, supporting by default EAD 2002, EAD 3, Dublin Core, and others;
– New storage system that keeps files directly on storage using standards, specifically the E-ARK Archival Information Package (AIP) 1 ;
– New ingest formats, that include BagIt, E-ARK SIP, direct files and folders, and ability to add others;
– New RODA-in tool that allows creation of thousands of SIPs with just a few clicks;
– More ingest action by default, and easier ways to make your own ingest process;
– Apache Solr for indexing all information and keep access fast and scalable;
– Akka.io for orchestrating all ingest, preservation and internal actions;
– Easy deployment using docker containers;
– And much more.

RODA 2.0 is now in production in the Publications Office of the European Union (EUPO) as the European Digital Object Repository (EUDORv3), which aims to cover all information in digital format, produced by EU institutions, bodies or agencies, either directly or on their behalf by third parties, and that is made available to the public. With already near 10 million AIPs, this implementation is a case study on performance and functionality.

Furthermore, the EUPO is undertaking steps towards the certification on ISO 16363:2012 "Space data and information transfer systems - Audit and certification of trustworthy digital repositories", for a formal recognition of trustworthiness. This sets a very high standard for the technical and organizational requirements of this implementation. The sharing of this experience would be of value for any archive that aims to achieve certification in future.

Re-use of Data as Incentive for Data Management

Anne Sofie Fink

Open access, open science, open government… As archives we are met with the concept of openness. We know better than anyone that we in our collections have data of great value for users of all kinds whether it is private citizens, companies, public administrators, researchers, founders etc. Obviously we want to provide data as open as possible in order to encourage re-use of all kinds as much as possible.

In the Danish National Archive we hold a digital collection of research data from the social sciences and health sciences dating back to the 1960's and a collection of digital public records dating back to the early 1970's. User service for the different collections has been developed aimed at different designated user communities. Likewise collection and preservation of these data has evolved in different streams.

The presentation will make an outline of similarities and differences in the collections with regard to data formats, demands for metadata, access restrictions etc. in order to argue for synergies and specialization in development of user services and – support respectively in order to optimize search-ability and re-use.

The best incentive for good data management practices by data producers whether it is public administrators or researchers is the credible expectation of re-use of data in the future. The presentation will provide information on two present cases on re-use: One case from public administration and one case from research.

From the two cases we will identify areas for synergies and mutual development of user services and support for the collection as a whole as well as areas for user services and support focused on specified user communities. Additionally we will from this analysis take implication for data management practices externally by public administration and research as well as internally by the archive.

By this two-sided approach of providing both common and specialized user services and user support we aim to provide optimal openness to data for both present and future users.

Using technology to gain insight, not to solve problems

Mette van Essen

Technology and new trends follow each other in quick succession. Possibilities seem endless. Speak recognition on our smart phones. Search results customized by our preferences. Cars which are no longer mechanical devices with some computers inside, but become computers with four wheels and an engine. This does not only has great influences on our own experiences. It also influences our work and our expectations of what those new technologies mean for the ways we work. But why is it still so hard to get those new and innovative ideas and technologies used within the field of information management?

One of the reasons is that most of our organisations are driven by automatization. The main reaction to the exponential growth of information is to deal with it as if it is a technical problem, with a technical solution. We make the processes as we know more efficient and scalable. In this quest for efficiency we let us guide by third-party tech companies who offer software solutions for automatic analyses and classification of information. More and more we start to rely on those black boxes but we don't have a clue what's happening inside.

To find out more about this black boxes and algorithms, the National Archives started a series of experiments with machine learning tools and systems in its own organisation. The two main objectives of these experiments are:
- Development of new insights in methods for information management in a digital area;
- Gaining knowledge about machine learning and the impact it has on our processes, our organisations and the people who have to work with those systems.

In this paper we discuss the first results of these experiments and share our methods for setting them up in a real life environment. We share the problems we encountered and how we dealt with them. No surprise here: most of them are not technical at all.

What global standardization in documentation management would entail

Bernadette Bosse

There is a global pandemic in the management of documented and digital information. Historically, the focus on standards has been for the analysis and use of digital data after commissioning, and the focus on document management has been on quality documents and corporate governance documentation.
There is an entire subset of information that has mostly been ignored, possibly because it is so logical to manage that industry decided it did not need standardization. Unfortunately, leaving this element in the hands of the many, has created an unmanageable beast; mismanaged information.
The result is that every company manages their documentation to a different philosophy and standard of quality. With such variations in control, the integration and interoperability between different software systems and different organizations is virtually impossible; which is particularly evident when an organization purchases a facility or asset through the acquisition process.
In addition, because of the differences that a lack of standardization allows, personnel who move between organizations, or even between projects within the same organization, do not receive or are able to find all the pertinent information needed to productively involve themselves, and have to learn a mostly new structure and system each time.

A lack of standardization has also invited poor control over management of modifiable documentation files, creating an entire additional level of concerns and impacts in an organization.
Along with improved software functionality, the antidote to this issue lies in professionally trained personnel available in any region of the world, and also in the global standardization of the core elements that make up documentation and data management – and this does not just refer to the storage and distribution methods, but to the culture of organizations and the infiltration of documentation and data management strategies into each and every department and activity in the organization.
The standards outlined in this paper are designed to provide a 5,000 ft. level of a structure, unlike the existing ISO standards, which while they are well worth the inclusion into any organization, only provide a 20,000 ft. level of how to create processes and procedures. Cost improvements will also be defined, providing insight on how to improve the overall fiscal benefit to projects and construction of facilities and pipelines.
The added benefit of the global standards initiative is to ensure seamless expectations and delivery on information between organizations involved in joint ventures, engineering and construction contracts, and for the purpose of acquisitions. This incorporates the creation and movement of information between divisions within an organization, which have historically been developed into silos; further breaking the management of information.

Citizen Archive as a Case Study in Personal Information Management, MyData and Digital Archiving

Mikko Lampi
Noora Talsi
Miia Kosonen

This paper proposal is about personal information management and governance, and its relation to MyData – human-centric data management approach and personal digital archiving. People are increasingly important actors in today's digital world where increasing amounts of data is born, used and managed. This paper will present a pragmatic case study of Citizen Archive in personal information preservation and management.

Digitalia – Research Center on Digital Information Management – is developing a professional-quality digital archiving solution available for common people. The Citizen Archive relies on an open-source platform allowing users to manage their personal data and ensure access to it on a long-term basis. The solution is on an early pilot phase and based on previously developed and enhanced OSA (Open Source Archive) platform. The motivation for the development work is preserving the valuable digital heritage by ordinary citizens and family archivists – in other words, "keeping their found things found", and capturing the part of the digital footprint they consider worth preserving. In general, such archive strengthens the civil society, digital rights, and information transparency.

MyData refers to personal data that complies with the principles of human-centric management and use. MyData paradigm connects with personal archiving by managing coherent descriptive metadata and access rights, while also ensuring privacy and usefulness. People have the right to obtain their personal data, use it freely, and to share, donate or sell that data to third parties. However, from the

Citizen Archive point of view, MyData is rather community data – containing information that links various users together. A small community, such as a family, produces, manages and collects the data.

The Citizen Archive is more than just a digital storage such as cloud drives. Most importantly, it relies on the above-listed MyData principles. Users may determine whether they preserve the materials for a set period of time or indefinitely. They can grant access rights to their family members, relatives and researchers. They also have easy-to-use searching and browsing facilities. Finally, users may collectively enrich information by describing the stories behind the archived content.

Citizen Archive: My Precious Information

Anssi Jääskeläinen
Liisa Uosukainen

Personal archiving as well as digital materials possessed by an average citizen are underrated areas. Harsh thing to say, but let us explain by asking a few simple questions Are you able to get your precious photographs, contracts and documents inside the shelter of a national archive? Is any business archive interested in your materials? Are there any other true digital repositories that would accept your material? We, as the developers of a Citizen Archive are truly surprised if even one of the above answers was yes.
The fact is that one needs to be politically or otherwise important person to get materials in the official repositories, most of the citizens aren't. Currently, citizen options for storing precious materials are portable USB devices, optical media or clouds such as Dropbox, Google Drive and OneDrive. These are good for storing backup copies especially if multiple simultaneous methods are used but by all means even with goodwill these cannot be considered as digital archives.
True digital archive needs to consider multiple aspects that plain cloud drives does not. E.g. legal aspects, possibility to share usage rights, metadata and findability by using metadata, guarantee that the data remains safe and inside the country perimeters, data and file format migration, suitable preservation formats, etc. During the Digitalia project that we are representing, also a need to edit and modify the digital material in order to get it more usable, aroused. For example, if a large PDF file with hundreds or even thousands of pages is automatically split into multiple smaller files according to the subject, content or keywords, the search results as well as using will be more user friendly. Furthermore, imagine a situation where you possess a multi gigabyte e-mail container (.pst file) exported from Outlook. You know that the important contract is inside that file but you don't have Outlook anymore. Our solution takes the usability and accessibility of the e-mails to the next level by transforming e-mail containers with their original attachments and metadata into searchable PDF/A-3b files.
Citizen Archives is the binding solution for all these aspects and more. This paper will describe the ongoing development work around Citizen Archive as well as some already implemented, functional and tested solutions considering e-mail preservation workflow and PDF splitting workflow. Finally, the experiences from the pilot users of the Citizen Archive are demonstrated.

Unique e-archives in Spanish public administration: Challenges, Roadmap and Tools

Elena Cortés Ruiz

Since 2007, when citizens' right to accede electronically to Public Administration, Public powers have developed both legal and technical tools oriented to reach unique e-archives in 2018.

First steps were settled in 2010, with National Security and Interoperability Schemas developed in several technical standards, all of them available through the E-Administration Portal.

In 2015 two new laws, one about common administrative procedures and public administration juridical regime become e-procedure as common and preferential system for relationships between citizens and public administrations.

Spanish Ministries are encouraged to develop their e-document management policies (Education, Culture and Sports; Public Administration, Defence…) and different tools are suggested, such as INSIDE (for e-documents and e-files management), PORT AFFIRMA (for e-signature) or ARCHIVE (for long-term e-archiving).

Accelerating records management at CERN

Andrew Short

In 2015 CERN (European Organization for Nuclear Research) implemented a records management system to manage personnel records dating back to the mid 1950's. This paper will look at the situation at CERN before a digital solution was available, what solution was chosen, what customisations were required and the current situation. There is also some insight into the future direction of the project and remaining goals.

A Case Study: Management and Exploitation of the Nuclear Decommissioning Agency Geoscience Data Archive

Jaana Pinnick
Andrew Riddick
Robert McLaverty
Garry Baker

The British Geological Survey (BGS) is responsible for managing a major geoscience data archive on behalf of the UK Nuclear Decommissioning Authority (NDA). Much of this geological data was captured during the 1990s and early 2000s using now obsolete software and data formats. However, this valuable data asset remains an important resource for the NDA and the wider scientific community. The NDA wishes to ensure the data remain accessible and usable for many decades into the future. BGS has been working closely with Radioactive Waste Management (RWM), a wholly owned subsidiary of the NDA, on a programme of data management and digital continuity measures to ensure the long-term usability of the data. This paper describes some of the challenges and outlines the approaches we have taken to address these issues.

The archive consists predominantly of digital data in a variety of file formats. They vary from legacy spreadsheet and word processor formats, to ASCII files in various formats and proprietary geological software formats. The data was derived from multiple sources (e.g. from different consultants). These factors present a number of data management and data quality issues, such as many file formats being difficult to read using current software tools. Some files represent outputs from modelling and can be re-built in modern software if the raw input data remains accessible. These factors pose clear challenges ensuring the continued usability of NDA data requiring appropriate digital preservation methodologies.

To address these challenges we have undertaken a programme of analysis of the various file formats to capture metadata about the file formats and their provenance using appropriate file profiling tools and registries (DROID and PRONOM). The data gleaned from this exercise has allowed us to identify those components of the data which are most at risk in terms of onward usability.

A number of data rescue initiatives have been progressed, including migrating a number of important datasets from legacy database formats into a new Oracle relational database structure so that this data is now in a modern format, usable, and easily accessible.

Preservation metadata is being captured and created during these activities using and customising the PREMIS standard, to support the onward usability of the archive and to allow us together with RWM to put in place appropriate data management and preservation measures to support their onward capture of geoscience data.

The Integrated Archival Information System "ZoSIA" and its implications for archival descriptive and reference practices in the Polish State Archives

Bartosz Nowożycki

The Integrated Archival Information System "ZoSIA" was created by the National Digital Archives in Warsaw as a unified platform for archival description and access, based on such international standards as ISAD(G). The system is being developed (along with an IT infrastructure) since 2007 and nowadays is the primary archival data management tool used by the Polish Archvies. The "ZoSIA" provides both a way to record descriptive information about archival holdings (fonds, series, files and items – description levels) and a means to view, search, and browse that information about them. It is also used as a tool for creating finding.
The online interface of the system - "szukajwarchiwach.pl" portal – is being developed since 2009. It contains information on holdings of the Polish State Archives and other institutions which utilize the "ZoSIA". ""szukajwarchiwach.pl" also grants access to digitized copies of archival material stored in the Polish State Archives, including download option.
The resulting change in rules and practices was accelerated by the technology development and appearance of the "ZoSIA", as a primary tool for archival description. Despite significant development and change in descriptive practices over the last two decades, no formal attempt to redefine description was made until last year. Existing rules that had been created by The Head

Office of State Archives from ‘60s to ‘90s have been recently evaluated, as a result on full implementation of the “ZoSIA”, planned this year, in all of the 33 State Archives.
In 2016 the Central Committee for Archival Methodology made a conscious effort to try to adjust the process of establishing intellectual control over holdings through the preparation of finding aids to the usage of Integrated Archival Information System. The Committee sought to identify all elements of information collected or used by archivists in any aspect of their work in order to codify a new set of rules - logical extension of archival information system covering the operations of the entire repository.
This paper discusses the mutual relations between archival IT systems, archival methodology and codes of best practices. Author is going to analyse the impact of implementation of the “ZoSIA” system on the archival rules according to changes recently introduced in Polish archives.

Machine learning and information governance: a happy marriage?

Seth van Hooland
Mathias Coeckelbergs
Simon Hengchen

The current hype on machine learning has spurred a new vague of hope and enthusiasm amongst records managers and archivists to rely on algorithms to reduce the amount of manual intervention for the management and appraisal of large volumes of non-structured content. Commercial players promote out-of-the-box tools for auto-classification, seamlessly integrated into shiny dashboards, but are the integration of machine learning within an information governance context such a happy marriage as it seems? This paper wants to give a pragmatic overview of both the possibilities and the limits of machine learning from an archival and records management perspective. After a global overview of the different types of contexts in which automation can be applied, the paper focuses more in particular on Topic Modelling (TM). This non-supervised machine learning method to automatically extract keywords from large volumes of non-structured text is presented with the help of a case-study in which TM is applied on digitised archival holdings of the European Commission (EC). The paper demonstrates that machine learning holds a lot of potential for the information governance community, but its implementation requires a lot of preparation and manual intervention.