sciendo

Convergence of digital preservation practices with digital repositories management practices for the Research Output and Cultural heritage

Nicolaie CONSTANTINESCU
Transilvania University of Brașov, Brasov, Romania
nicolaie.ciubotaru@unitbv.ro

Angela REPANOVICI
Transilvania University of Brașov, Brasov, Romania
arepanovici@unitbv.ro

Abstract. *Managing and curation of digital assets are roles attached with libraries in their continuous pursuit of expanding their mission into the digital collections' realm. The growth in numbers and variety urged the librarians to take a deep look into digital curation practices, and techniques in order to preserve their ever-growing digital repositories. An investigation into how this happens from the perspective of European and world practice is the target of this paper. Considerate attention is given to the new roles and the required skills these mutations are demanding from the libraries' specialized personnel. The practices are slowly drifting towards a more integrative perspective where one information specialist sits at the intersection of digital curation and archival practices with daily practices.*

Keywords: digital curation, digital repositories, librarian skills, management practices

The growth of digital collections and implication on practices

The roles and responsibilities of the memory institutions concerning research output and cultural heritage evolve in the context of an uphill battle with the growth of digital collections and declining budgetary allocation. This paper seeks to find the important aspects of how digital preservation practices are actively incorporated in digital repositories management actions and policies. The investigation took into account the major documents issued by the most respectable European and international institutions, with a particular interest in strategies, training frameworks, and dedicated policy frameworks seeking solutions for long-time presentation.

Cultural heritage and research are two parts of the same story of humanity's efforts in preservation of the digital works of today. Offering rich representation of art, history, music, etc. is a direct link with further development of creative industries and tourism. Cultural and creative industries represent 4% of the European GDP according to a 2015 European study (Executive Agency for Small and Medium sized Enterprises, 2016). All these digital inputs and outputs need to be taken care of because they represent the collective of who we are today. The numbers and complexity of digital representations stored in digital repositories are in dire need for long-term preservation growing by the year. At a glance, Europeana.eu provides 50 million digital representations of cultural heritage objects according to the late Europeana DSI-4 Annual Report (European Commission. Directorate General for Communications Networks, Content and Technology, 2020a). Museums choose to expose their best digital photographic representations, partly as a possible bridge (Pekel, J., 2014) with new communities, and finally to exert quality control (Verwayen, Kennisland, & Kaufman, 2011) (See Yellow Milkmaid syndrome -

https://doi.org/10.2478/9788395815065-006

https://yellowmilkmaidsyndrome.tumblr.com/). According to some late findings (McCarthy, D., 2021), there are more than 50 millions other digital entities contributed by the various memory institutions around the globe. And this is only a rough estimate of the cultural heritage exposed. We have to add the vast riches locked in digital repositories, and the open access repositories filled with research output. The composition of the digital objects is increasing in variety lonely if we take into account the new 3D digitising projects or the new addition of datasets to the mix of already existing research results.

From standards to policies looking to the new skills

All of these digital representations, and born digital artefacts are exposed to corruption depending on the media carrier and the level of technical support. Most of the management activities involving long-term preservation are placed in the range of risk management. In close association there is the concept of trust as intrinsic quality of digital repositories. Trust is built upon standards implementation, and one primary source of truth is Open Archival Information System (OAIS) which needs to be understood as a “technical Recommended Practice” although later it became standard. Looking back to the OAIS standard/model, there is a clear separation between the administration of a digital repository and the management component. Only the Administration Functional Entity organizes and orchestrates different specific activities from the engineering driven ones. One of the important functions of the Administration entity is to ensure Preservation Planning. This is the level where most of the functions mentioned is reflected in today requirements for jobs connected to digital repository management.

A naive approach to digital preservation would be to set targets concerning bit preservation. Bit preservation concerns data loss, hence represents a direct managerial responsibilities as well. It is tempting to consider bit preservation as a simple storage issue accompanied by proper managerial decision, but that would be an oversimplification of processes not fully considered or understood (Zierau, E., 2018).

Resources are allocated to certifications (*Audit and Certification of Trustworthy Digital Repositories,* 2011) of the repository, hence applying financial solutions in order to meet the goals of fulfilling curatorial duties, organizational and operational ones.

From the managerial point of view, most of the efforts aim at getting the repository to a level of compliance close to the ISO standards universally accepted. For this purpose, some of the memory institutions have designed and adopted frameworks dedicated to digital preservation. Some resorted to dedicated policies matching their own specificities. The vast majority seek to follow ISO 16363: Trusted Digital Repository Checklist when taking decisions concerning the future developments of their repositories, and most of the policies incorporate the Lifecycle Model issued by the Digital Curation Centre in 2010.

The accompanying document to Open Archival Information System (OAIS), *Audit and Certification of Trustworthy Digital Repositories* gives general directions to follow in establishing the actions of a repository manager (*Audit and Certification of Trustworthy Digital Repositories*, 2011). Besides setting the policy framework of the digital repository, there is one very important managerial task: find and employ staff with adequate skills and experience. The aforementioned document present important aspects concerning the competencies and the skill sets needed to operate a digital repository, and these are ranging from technical skills to legal expertise.

Another important metric specified is the *commitment to analyze and report on financial risks* because, actually the management tasks of a digital repository are in the sphere of risk management (risk analysis reports, analyses policies, etc.). These criteria are at the bases of every analysis when a new repository takes form or some are being rehashed into a new modern breed.
Concerning the policy level there are concerns towards how many and what initiatives are being developed and aligned in Europe. The Commission Recommendation of 27 October 2011 on the digitisation and online accessibility of cultural material and digital preservation (European Commission, 2011)is the number one document to look up concerning digitisation and most important digital preservation. Its importance yields out of the track record of the Member States based on the general policy dedicated to *Digitisation & Digital Preservation - https://ec.europa.eu/digital-single-market/en/digitisation-digital-preservation*. With regards to scientific output, there is the Commission Recommendation (EU) 2018/790 of 25 April 2018 on access to and preservation of scientific information (European Commission, 2018) that tackles also the issues of preservation.
In the section dedicated to "Long-term preservation strategies and action plans" of the *European Commission report on Cultural Heritage: Digitisation, Online Accessibility and Digital Preservation* issued by the Commission mid-2019, we get a fragmented picture of Europe with some peaks leading the way with regards to the first recommendation of afore mentioned Recommendation in the cultural heritage field (2011/711/EU). German competence network for digital preservation – *nestor* is pointed out as one important leading examples in the field. The conclusions drawn by the Commission pointed out a lack of clear responsibilities combined with lack of suitable legal framework. The other Member States reported participation in digital preservation projects as the only evolutionary steps. One European dedicated project to the analysis of digital preservation costs took into consideration the managerial role in a roadmap titled *Investing in Curation: A Shared Path to Sustainability - The 4C Roadmap* (4C, (2013). It worth mentioning that in the third, and the fifth part concerned with scalable services and infrastructures, the necessity for the managerial level to understand the costs of curation is underlined. More than that, providing financial support for training becomes paramount.
Looking at the history of the policies, one of the most active and vocal community of contributors and reviewers were the librarians, the managers of critical digital services all over Europe and beyond. This is best reflected in the *2019-20 Community Survey Report* (*2019-20 Digital Preservation Community Survey*, 2020) issued by Open Preservation Foundation. The breakdown of the figures reflecting personnel distribution playing a direct part in their organization's digital preservation activities put on the first place the *cataloguer or metadata analyst* (68% of the respondent institutions), on the second place the *director, manager or administrator* and on third, the *digital archivist or curator*. According to the same study, the most common activities are *digitization, metadata creation/extraction, format validation, storage or bit preservation and format identification*. These activities indicate a need for continuous training for the librarians/archivists, and a broader understanding on the management levels for including new roles or new sets of skills. Librarian/Archivist is the term encompassing now many different specializations many surging also in the managerial top tier. Providing resources needed for training is part of every preservation strategy. Besides the roles mentioned above, in the case of research results, the *Recommendation on access to and preservation of scientific information* (European

Commission, 2018) adds some interesting new traits in the mix: data research management, *data stewardship*, *data preservation*, and *data curation*.
All of these do not constitute the norm in Europe at least, because of unequal development due to historical and resource distribution. According to ENUMERATE 4th Core Survey (Jan Nauta, Van den Heuvel &Teunisse, 2017)*the majority of memory institutions in Europe do not have a formal strategy for digital preservation* and *45% of the respondents do not have a solution yet for long term preservation based on international standards for digital preservation*.
The Rescue of the Danish Bits by Eld Zierau revealed another aspect important for management in the terms of misunderstanding important aspects of digital preservation. Digital preservation on the level of bit preservation is beyond hardware and IT, and every decision upon copies movement and management should be insulated from potential disastrous political decisions. This should be realized through addition to decision making process of highly trained persons with rich and comprehensive background on the whole process. In the wake of the Danish experience mentioned above it should be considered a good starting point a careful planning ahead for sensible copies management in the context of risk analysis and expertise of highly trained personnel in the library.
Demands for long-term preservation of research results extends in the realm of reproducibility in the context of Open Science. A rich example which employs new knowledge and refinement of skills is the European Centre for Nuclear Research Analysis Preservation (CAP) *an integrated platform for scientific analysis, preservation and reuse* - https://analysispreservation.cern.ch. Put in the perspective, CERN manages over 330 petabytes (330 million gigabytes) of data (https://home.cern/science/computing/data-preservation), and Large Hadron Collider (LHC) only generates 90 petabytes of data per anum (https://home.cern/science/computing/storage). Not all the data is persevered, but the data needed for future analysis as bases for reproducibility is stored in CAP repository. Due to the nature of data, CAP operates end-to-end solution including code, datasets and metadata. CAP (Fokianos et al., 2020) is also an interconnected service with the world biggest code repositories: Github and GitLab. This adds a new knowledge layer on the librarians that slowly become more engaged in understanding and sustaining infrastructures dedicated to preserving live context for interpretation and data visualization. And this new layer needs investment in order to get the maximum preparedness possible. The argument to this need would be the growing need to understand and support digital curation.
Digital Curation is broader in scope than digital preservation (Ross, S., 2006). Today practices involved in the management of a digital repository take into consideration skills encompassing documentation, providing access and unique identifiers, storage, redundant backup strategies, etc. Digital Curation Centre gives a concise definition for digital curation as *the management and preservation of digital data/information over the long-term* (Digital Curation Centre, n.d.). This information relates to the requirements of Data Management Plans required by the European Commission as it is clearly stated as part of the document itself: *how data will be curated & preserved (including after the end of the project)* - https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm .
Concerning research output, Plan S which was launched in 2018 came into force beginning with January 2021. Plan S is backed by the cOAlition S, a global alliance of organisations that align their Open Access policies with common set principles (Plan

S Principles. Part I: The Plan S Principles, 2021). In the centre of view there are the open repositories for research as facilitators of Open Access to research and time vehicles for the content. According to the practical advice, repositories must be compliant to a basic set of criteria (Plan S practical advice. Requirements for Open Access Repositories, 2020). To all of these, there must be a match in managerial plan, and in personnel qualifications.

Discussion

According to Europeana Strategy 2020 – 2025, *digital means «usable»*, and this entails opening up the collections, but it lays out the prospect and the necessity of maintaining the digital entities created. In response to the concept of *digital transformation* of cultural heritage sector, Europeana developed the ambition to become a knowledge centre able to boost capacity though training on *digitisation, metadata enrichment, semantic interoperability, content creation, licensing, reuse, business models and innovation* (Objective 3A).

Digital preservation became part of digital repository practices out of the need to develop digital curation skills. Some skills were included in the workflows of the applications used to manage the digital entities. The management level should invest on three fronts once digital preservation plan/strategy kicks in:

- skilled members of the staff,
- multiple copies of the digital assets;
- continuous funding, and expanding opportunities set by consortia or projects.

One possible path to solving some or most of the administrative tasks concerning the management or a digital repository could be outsourcing the services to third party commercial providers. One possible road would be employing the services of companies which are not for profit like archive.org - https://webservices.archive.org/pages/preservation.

Conclusion

A digital repository is the sum of the people actively engaged in development and maintenance. Digital preservation with all the rigours aimed to achieve the attribute of “trusted” may happen or it could become a future target depending of how far sighted is the managerial tier. Memory institutions should become real competence centres for digital preservation and the role it carries for future of cultural heritage and scientific data.

Digital preservation practices are reflected in today’s diverse roles existing in the memory institutions which exert digital preservation attributes. One of the constant preoccupation of the policies and such of the financial decisions is how and where the copies are kept, and under what contractual terms. Soon a new class of solutions will emerge out of the use of Distributed Ledger Technologies – block chain. One compelling research is project ARCHANGEL (https://www.archangel.ac.uk/about/) aimed to sustain long-term sustainability of digital archives.

The public expects to be part of the curatorial processes when it comes to the digital representations accessible via frontends. This particular aspect creates a new layer of what is considered valuable for a continuous endowment, but little do they know the curatorial processes of the data though all its stages of lifecycle. Preserving for long time stands in the remit of memory institutions, and long-time preservation cannot be left to future actions.

References

4C. (2013). Investing in Curation: A Shared Path to Sustainability—The 4C Roadmap [Blog]. Collaboration to Clarify the Cost of Curation. https://www.4cproject.eu/roadmap/

2019-20 Digital Preservation Community Survey. Findings report. (2020). Open Preservation Foundation. https://openpreservation.org/wp-content/uploads/public/resources/surveys/2020/OPFCommunitySurvey_Report_v04_200907.pdf

Audit and Certification of Trustworthy Digital Repositories. (2011). DATA ARCHIVE INTEROPERABILITY WORKING GROUP (MOIMS-DAI); Magenta Books: Recommended Practices. https://public.ccsds.org/Pubs/652x0m1.pdf

Collomosse, J., Bui, T., Brown, A., Sheridan, J., Green, A., Bell, M., Fawcett, J., Higgins, J., &Thereaux, O. (2018). ARCHANGEL: Trusted Archives of Digital Public Documents. Proceedings of the ACM Symposium on Document Engineering 2018, 1–4. https://doi.org/10.1145/3209280.3229120

Digital Curation Centre. (n.d.). What is digital curation? Digital Curation Centre. https://www.dcc.ac.uk/about/digital-curation

Donaldson, D. R. (2020). Certification information on trustworthy digital repository websites: A content analysis. PLOS ONE, 15(12), e0242525. https://doi.org/10.1371/journal.pone.0242525

Edmunds, R., L'Hours, H., Rickards, L., Trilsbeek, P., Vardigan, M., & Mokrane, M. (2016). Core Trustworthy Data Repositories Requirements. https://doi.org/10.5281/ZENODO.168411

European Commission. (2011). Commission Recommendation of 27 October 2011 on the digitisation and online accessibility of cultural material and digital preservation (OJ L 283). Official Journal of the European Union. https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32011H0711

European Commission. (2018). Commission Recommendation (EU) 2018/790 of 25 April 2018 on access to and preservation of scientific information (OJ L 134). Official Journal of the European Union. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32018H0790

European Commission. Directorate General for Communications Networks, Content and Technology. (2020a). Europeana DSI-4 annual report.August 2020 . Publications Office. https://data.europa.eu/doi/10.2759/163536

European Commission. Directorate General for Communications Networks, Content and Technology. (2020b). Europeana strategy 2020-2025: Empowering digital change. Publications Office. https://data.europa.eu/doi/10.2759/524581

Executive Agency for Small and Medium sized Enterprises., VVA., & KMU Forschung Austria. (2016). Boosting the competitiveness of cultural and creative industries for

growth and jobs: Final report. Publications Office. https://data.europa.eu/doi/10.2826/526239

Fokianos, P., Feger, S., Koutsakis, I., Lavasa, A., Maciulaitis, R., Naim, K., Okraska, J., Papadopoulos, A., Rodríguez, D., Šimko, T., Trzcinska, A., Tsanaktsidis, I., & van de Sandt, S. (2020). CERN Analysis Preservation and Reuse Framework: FAIR research data services for LHC experiments. EPJ Web of Conferences, 245, 06011. https://doi.org/10.1051/epjconf/202024506011

Jamie, S., Frank Olaf, B., Jakob, B., Gerardo, G., Dallmeier-Tiessen, S., Simko, T., & Cancio Melia, G. (2016). CERN Services for Long Term Data Preservation. https://cds.cern.ch/record/2195937

Jan Nauta, G., Van den Heuvel, W., & Teunisse, S. (2017). Europeana DSI 2– Access to Digital Resources of European Heritage: D4.4. Report on ENUMERATE Core Survey 4. Europeana. https://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/ENUMERATE/deliverables/DSI-2_Deliverable%20D4.4_Europeana_Report%20on%20ENUMERATE%20Core%20Survey%204.pdf

Jetten, M., Grootveld, M., Mordant, A., Jansen, M., Bloemers, M., Miedema, M., & Gelder, C. W. G. V. (2021). Professionalising data stewardship in the Netherlands. Competences, training and education. Dutch roadmap towards national implementation of FAIR data stewardship. Zenodo. https://doi.org/10.5281/ZENODO.4486423

Lin, D., Crabtree, J., Dillo, I., Downs, R. R., Edmunds, R., Giaretta, D., De Giusti, M., L'Hours, H., Hugo, W., Jenkyns, R., Khodiyar, V., Martone, M. E., Mokrane, M., Navale, V., Petters, J., Sierman, B., Sokolova, D. V., Stockhause, M., & Westbrook, J. (2020). The TRUST Principles for digital repositories. Scientific Data, 7(1), 144. https://doi.org/10.1038/s41597-020-0486-7

McCarthy, D. (2021). Open Access and Art History in the 21st Century: The Case for Open GLAM [Website]. CODART. https://www.codart.nl/feature/museum-affairs/open-access-art-history-and-the-21st-century-museum/

Pekel, J. (2014). Democratising the Rijksmuseum. Why did the Rijksmuseum make available their highest quality material without restrictions, and what are the results? [Blog]. Europeana Pro. https://pro.europeana.eu/post/democratising-the-rijksmuseum Plan S practical advice. Requirements for Open Access Repositories. (2020). European Science Foundation. https://www.coalition-s.org/plan-s-practical-advice/ Plan S Principles. Part I: The Plan S Principles. (2021). European Science Foundation. https://www.coalition-s.org/plan_s_principles/

Potterbusch, M., McGovern, N., Whiteside, A., Mumma, C., Verbruggen, E., Meyerson, J., Lyons, B., Pang, N., Schaefer, S., Weatherburn, J., Pennock, M., Lindlar, M., & Patterson, T. (2019). IPres2018 Conference. https://doi.org/10.17605/OSF.IO/U5W3Q

Ross, S. (2006). Approaching Digital Preservation Holistically. In Information Management and Preservation. Chandos Press. https://www.researchgate.net/publication/272482321_Approaching_Digital_Preservation_Holistically

Tough, A., & Moss, M. (Eds.). (2006). Record keeping in a hybrid environment: Managing the creation, use, preservation and disposal of unique information objects in context. Chandos.

Verwayen, H., Kennisland, M. A., & Kaufman, P. B. (2011). The Problem of the Yellow Milkmaid. A Business Model Perspective on Open Metadata. https://pro.europeana.eu/post/the-problem-of-the-yellow-milkmaid

Zierau, E. (2018). 208.2 The Rescue of the Danish Bits - A case study of the rescue of bits and how the digital preservation community supported. -- Award winner: Best Long Paper. https://doi.org/10.17605/OSF.IO/6T85D