

Optimizing the Potential of Research Data Through an Integrated Data Management Approach: Considering Research Method, Data Life Cycle, Big Data and Linked Data in an eResearch Example in Australian Rock Art

Haubt, R.A.^{1*}, Jalandoni, A.²

¹ Billy Blue College of Design, Torrens University, Brisbane, QLD, Australia

² Place, Evolution and Rock Art Heritage Unit (PERAHU), Griffith University, Southport, QLD, Australia

Commission II, WG II/8

KEY WORDS: rock art; cultural heritage management; data management; information management; e-research

ABSTRACT:

This paper looks at the current state of e-research strategies in rock art on the example of the Global Rock Art Database, global and Australian e-research communities. It examines current practice, attitudes and requirements for discipline specific research methods in an integrated data management cycle approach. Analysing qualitative and quantitative data collected between 2012 and 2018 through conversations, consultations, a cross-sectional questionnaire and a longitudinal study of the Rock Art Database, the paper compares its findings to previous interdisciplinary studies within e-research environments. The resulting data illustrates current practice and trends in rock art within an e-research context and aims to inform future best practice towards integrated data models digitally connecting international research data.

1. INTRODUCTION

Data plays a central role towards a better understanding of the world. In fact, data shapes the way we understand the world. Researchers look at data to find answers to their research questions, using tools tailored to their own research needs ranging from project planning and data acquisition to analysis and archiving. Digital workflows have become part of our daily routine and we are producing more and more data, some of which we are aware of and some data of which we are not.

1.1 Motivation

With a deluge of data in recent years, the world is becoming a more informed place but at the same time increasingly faces challenges in making sense of the abundance of data. We need to become more aware of data and understand how the data we produce fits within a larger network of information to explore the full potential of its power. Data networks such as Linked Open Data (LOD) or the Registry of Research Data Repositories (<https://www.re3data.org>) already explore such potential and more and more publishers such as Nature (<https://www.nature.com/sdata/policies/repositories>) ask their authors to share their data in field specific repositories such as figshare (<https://figshare.com>) or Dryad (<https://datadryad.org>).

Most researchers are unaware of the disconnect between what they would consider 'important data' for their research and how they actually treat this 'important data'. Due to a number of factors ranging from non-innovative work practices, culture in the workplace and economic factors, quick results are often favoured over more sustainable data models and data is often discarded after it fulfilled its temporary purpose, instead of storing and sharing it in online repositories, without further exploration of its potential within a larger context. There is a need to explore more sustainable and scalable data models in an ever-growing data driven world towards understanding the potential of data and optimizing its return considering the time and money we invest in producing data.

1.2 Global Research Data

In *The next Web*, Berners-Lee (2009), the inventor of the World Wide Web, talks about these developments and envisioned a model using digital tools and data to connect the world's knowledge for the visible and in-visible (machine readable) Web. His model describes how these new technologies would assist us with designing an interconnected knowledgebase that would store and give the world access to vast amounts of scientific and cultural knowledge. Making large amounts of this information freely accessible would allow the world to use these resources for learning, teaching and research and allow the world to contribute to existing data models through international and interdisciplinary collaboration.

1.3 Australian Government and Research Data

The Australian Government has long realised the potential of such digital tools and is supporting growth of the national e-research capacity through funding the National Collaborative Research Infrastructure Strategy (NCRIS). The NCRIS takes a national approach to invest and financially support Australian research collaboration on a global scale. NCRIS-funded services include the Australian National Data Service (ANDS) and the Australian Access Federation (AAF) that provide world-leading approaches in data management and data access (Cochran, 2015). These services are further linked to the National Computing Infrastructure (NCI) (<https://nci.org.au>), the Paisey Supercomputing Centre (<https://www.pawsey.org.au>) and the National eResearch Collaboration Tools and Resources (NeCTAR) (<https://nectar.org.au>).

1.4 Australian State Government and Research Data

In 2008 all states in Australia established bodies in support of the National Broadband Network and information infrastructure plans. New South Wales, for example, established Intersect (www.intersect.org.au), a not-for profit company owned by its members including universities and research institutions that is further financially supported by the Commonwealth and State government. Intersect works closely with the ANDS, NeCTAR and is a member of NCI, AAF and the Australian eResearch

* Corresponding author

Organisations (AeRO). Intersect currently provides eResearch services to its subscribed members.

1.5 Australian Research Institutions and Research Data

The University of Sydney in collaboration with Intersect (2009), conducted an online survey for Developing eResearch Infrastructure: Technology-enhanced research practice, attitudes and requirements. The survey included 658 participants at four major Universities in New South Wales and identified a need to provide better support, training and services for the use of digital technologies for research practice.

Findings of the report included that up to 84% of researchers did not use specialised tools, which included (1) data mining; (2) GIS; (3) digital voice recognition and transcription tools; (4) special audio visual software. The reports further revealed that while specialised tools are not utilised 21% of participants expressed a need for statistical, mathematical and financial data management, 21% expressed a need for modeling and simulation, 19% expressed a need for visualization and visual data analysis tools, 18% expressed a need for voice recognition and transcription tools, 18% expressed a need for data mining tools and 16% expressed a need for qualitative textual and linguistic analysis tools. Looking at research collaboration the report further showed that while over 70% of the participants fully or partially collaborate with other researchers, 91% rely on email, telephone and face-to-face meetings while 88% do not use digital applications for task or project management.

While the majority of participants publish their findings in journals and books, only 45% use their institutional repositories or personal websites and less than 20% use discipline specific repositories to disseminate their data. The majority of participants stated that instead of institutional repositories they use less sustainable platforms for data storage such as USB, DVD and other portable storage solutions. 43% of participant therefore had data management and preservation issues and 54% indicated that no explicit data management plans existed.

Considering the fast development of digital technologies and new mobile and cloud services emerging on a daily basis, these results might be surprising to a tech-savvy researcher but clearly illustrate a need for better understanding of the usefulness for the less technical familiar researcher. Some of these issues are currently being addressed by e-research initiatives within state governments, institutions and global initiatives such as library carpentry (<https://librarycarpentry.org/>), data carpentry (<https://datacarpentry.org/>), software carpentry (<https://softwarecarpentry.org/>) or Open Science Training (<http://opensciencetraining.com>) that train researchers in the use of digital research tools from data collection to dissemination. Software solutions are also being developed through projects such as the Field Acquired Information Management Systems (FAIMS) (<https://www.fedarch.org>) project that develops customized mobile Apps to streamline integrated workflows for field based research.

1.6 Idea and Concept for Improving Research Data

Services and guides to better support researchers in the use of new technologies are clearly needed. These services should not only be facilitated by government agencies but by support services on the ground within universities, faculties and collaborative research communities to respond more specifically to individual researchers data training and data management needs within specific disciplines.

1.7 Problems in Rock Art Research Data Management

Similar issues with regards to research data training and management were observed in a variety of rock art research projects during the development and implementation of the Global Rock Art Database, which ran from 2012 to 2016 (Haut & Taçon, 2016). This paper investigates and summarizes the lessons learned from this international rock art project.

2. METHOD

2.1 Aims & Outcomes

This paper aims to provide a guide for rock art researchers based on lessons learned from the Intersect study (2009) and the Global Rock Art Database project (Haut & Taçon, 2016). The guide aims to address three major issues to maximize the use of data in rock art research considering time and money spent on producing valuable data through understanding the impact of data towards a digitally connected world:

Promote training and use of:

- Specialized digital tools for rock art research
- Digital collaboration tools that allow seamless integration into every day workflows for the duration of the project life cycle
- Discipline specific online repositories for data storage and sharing

The project takes a Digital Humanities approach by placing rock art in an e-research context. Therefore the paper is written from two interconnected perspectives: a) the information and communication designer with a view towards exploring data potential in an international and interdisciplinary data management context and b) the rock art researcher's view to provide insight into the subject's specific issues and needs.

2.2 The Integrated Research Method and Research Management Plan Approach

To better understand the needs and potential of research data management we will look at three interconnected approaches:

- Research Method: describes the chosen Informed or Formal Method to provide answers to a specific rock art research question
- Research Technologies: describes the tools used in support of the research method
- Project Management Plan: describes the research project life cycle and associated data management life cycle

The three approaches will be mapped against each other to explore the full potential of research data at all stages of the project life cycle.

2.2.1 Rock Art Research Method: It is difficult to find well-defined rock art research methods for tangible and intangible heritage. On the one hand each project has different aims and tries to answer different questions on the other hand the majority of rock art research predominantly seems to apply general archaeological approaches. Over the last two decades these approaches have been examined and improved to better suit specific rock-art needs. Whitely (2011, pp. 71-107) summarizes different methods and describes Scientific Method, developing a hypothesis and comparison of competing ideas, as the most common approach. Chippindale and Taçon (1998, pp. 1-10) break this down further into Informed Methods and Formal Methods. While the Informed Methods relate to the cautious use of ethnological and ethnographic data, the Formal Methods use quantitative, scientific and location data that might be more suitable for tangible data concerned with, for example, social functions rather than symbolic meanings of rock-art. Informed and Formal Methods can both be broken down further to suit more specific research question and needs.

2.2.2 Research Technologies: The table below provides a list of common data gathering methods and associated use of technologies in rock art research.

Research Topics	Method	Tools	Data Formats	Uses
Literature Review	Systematic and Quantitative (Jalandoni et al, 2019)	Excel	XML	Categorising and Quantitating
Recording: 2D images	Photography	Camera	JPG NEF CRW, etc	2D visualization, apply algorithms
Recording: 3D models	Photogrammetry	Camera	STL, OBJ, FBX, COLLADA 3DS, IGES STEP VRML X3D, etc	2D and 3D visualization, measurements, apply algorithms
	Laser Scanning	Laser scanner		
Interviews	Audio recording	Audio recorder	PCM, WAV, AIFF, MP3, AAC, OGG, WMA, etc	Podcasts, historical records
	Video recording	Video recorder	WMV, ASF, RM, MOV, MPEG, etc	Documentaries, historical records

Table 1. Examples for rock art data gathering methods

The aim of the list is not to discuss the preference of one method over another but instead aims to provide an example for methods to start a discourse on how the rock art community could, as a collective, work collaboratively on sharing and improving these methods to develop a more standardized approach given specific research questions (see Table 1).

2.2.3 Project and Data Management Life Cycle: The project was guided by the Australian National Data Service's 23 Things (<https://www.and.s.org.au/working-with-data/skills/23-research-data-things>) and Research Data Management Practice's Guide (<https://www.and.s.org.au/guides/rdm-in-practice>). The data management approach can be broken down into seven steps, the PPADSRI (see Figure 1) and assist with a better understanding of the potential of research data and related research methodologies within each step of the data management cycle. The cycle further contributes towards a more scalable and sustainable data management model as outlined by the 2017 Australian Research Council's 'Data Management' plan requirements. All collected qualitative and quantitative data is mapped against individual steps within the cycle to identify issues that arise within project and data management.

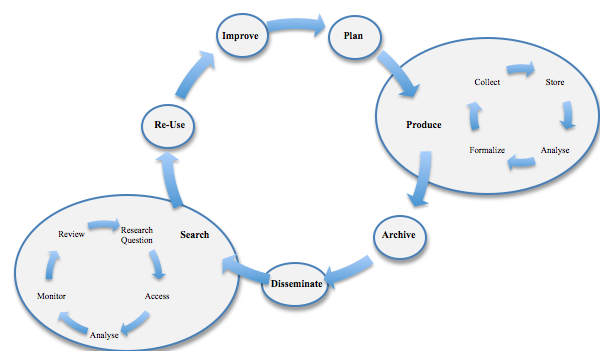


Figure 1. Data Management Cycle using and Agile PPADSRI approach

2.3 Data Collection Process

To inform the development and implementation of the Global Rock Art Database project, qualitative and quantitative data was collected between 2012 and 2017. The collected data assisted with information structure of the Rock Art Database platform to support rock art research methods. The data further assisted with identifying and addressing data management problems in rock art research.

2.3.1 Participants: All participants in this research were selected based on their current roles within University and Government based cultural heritage research institutions and museums including project directors, project leads, managers, researchers, surveyors and administrators. The samples include archaeologists, anthropologists, historians, information scientists, museum and art curators, librarians and communication and information designers. Due to privacy policies, participant's personal details and affiliations were omitted.

2.3.2 Qualitative Approach: Qualitative data was collected through conversations and consultations with international rock art researchers at cultural heritage conferences, phone interviews, meetings and through engagement with subscribed members via the Global Rock Art Database. The data was stored and analysed in a semantic PM Wiki using a Systematic Quantitative Review based on the the PIMRI Data Management Life Cycle (Plan, Implement, Manage, Review, Improve) (Haupt, 2015)

2.3.3 Quantitative Approach: To provide a snapshot and insight into the state and development of data management strategies and research methods commonly used in rock art, data was collected in a cross-sectional study and longitudinal study based on question of the Intersect study in 2009.

A cross-sectional online questionnaire was distributed in 2015 to international rock art research and cultural heritage institutions. The multiple choice and short answer survey was based on questions relating to the Rock Art Stability Index (RASI), Google Structured Data, Linked Data, the CIDOC CRM and Informed and Formal Methods in rock art (Haubt, 2016).

A longitudinal study was conducted for the duration of the Global Rock Art Database life cycle between 2012 and 2016. Monthly samples were taken and analysed using WebVOWL and ReFinder for semantic content and structural analysis, while Google Analytics was used for statistical evaluation of user contributions and interactions (Haubt, 2016).

3. RESULTS

A total of fifty rock art researchers and heritage professionals contributed to the qualitative data collection between 2012 and 2016. Information was provided following discussions with thirty conference participants after ten international heritage conference presentations on the subject, including talks at the Society of American Archaeology (SAA) in 2015, the Australian Archaeological Association (AAA) in 2013 and 2016, ICOMOS body for Cultural Heritage Documentation (CIPA) in 2013, 2015 and 2016. A further twenty conversations were recorded through consultation sessions with rock art researchers and heritage professionals and inquiries by members of the Global Rock Art Database.

A total of fifty rock art researchers and heritage professionals contributed to the quantitative data collection. Thirty people participated in the cross-sectional survey, while twenty members of the Global Rock Art Database contributed a total of 250 rock art projects to the database for analysis.

3.1 The Integrated Research Method and Project Management Approach

Qualitative and quantitative data was collected, stored and cross-analysed in the Global Rock Art Database. The resulting data was exported into Excel and semantic Wiki for further analysis. The results and resulting discussion of the qualitative and quantitative data analysis have been mapped against each step of the data management life cycle to provide examples and potential solutions for rock art data and research management approaches to enhance research methods.

3.1.1 Data Planning: While international and national research bodies like the Australian Research Council (ARC) provide guidelines for writing research proposals considering e-research tools and data management approaches, researchers addressed a need for more targeted examples of data management guides for cultural heritage or rock art to assist with grant applications.

Participants further addressed the difficulty of finding specific rock art information on the internet or in cultural heritage archives due to a highly decentralized cultural heritage system. As a result missing information effected the planning of new research projects and sometimes resulted in doubling up research efforts.

3.1.2 Data Production: The data production cycle of collecting, storing, analyzing and formalizing was addressed with regards to optimizing workflows. Researchers discussed the difficulty of eg working in the field collecting data and having to wait to return to the lab or office until analysis can begin. A more integrated data collection approach could assist with streamlining the process of collecting and analyzing data against existing data sets while in the field. Examples included updating of recording forms once new data was compared against old data sets, which lead to the introduction of new categories or data types.

3.1.3 Archiving: A seamless integration of information storage to optimize workflows, similar to the Data Production stage, using cloud computing or similar technologies was addressed. Further, information storage and retrieval, especially with regards to using a variety of multimedia file formats eg Laser Scans, Photographs, Videos, Sound, GIS or text files on different storage devices or locations was another major problem. Integrated systems allowing for seamless navigation, search and information retrieval between different file formats and archives are missing.

3.1.4 Dissemination: The biggest issue with regards to the dissemination of data was that researchers struggled with understanding new data management requirements by government authorities towards linked and open data. While the majority of researchers had a keen interest on participating in wider collaborative networks, sharing their data with relevant researchers and authorities, concerns were raised with regards to data integrity and the issue of sensitive heritage data.

3.1.5 Search, Accessibility and Visibility: Researchers addressed the issue of accessibility and visibility of research outputs after publication. Publication of research data in commercial high profile journals was generally favored over Open Access due to institution's research impact factor requirements associated with academic rigor and general assumption of higher visibility in high profile journals. It is interesting to note that most research outputs were published in formalized formats such as book chapters and journal articles, while researchers were reluctant on publishing or sharing their raw data sets.

3.1.6 Re-Usability: In connection with search and visibility researchers often struggled with re-using old data due to accessibility (eg data locked away on external hard drives, filing cabinets, old servers etc.) or incompatibility due to newer multimedia and software versions or updated data recording forms that were inconsistent in use with older forms. Only little old data was re-used rendering most old data sets obsolete posing the question of the sustainability of the original research project.

3.1.7 Improving and Building onto Existing Data: Following up on the re-usability of data and improving new research undertakings, researchers addressed lessons learned considering their research methods but often did not consider how research method and sustainable and scalable project planning would enhance their data management towards a more integrated and interdisciplinary data management model. When asked, the majority of researchers used their hard work collected data "to make ends meet" but mostly failed to explore the potential of their findings within a larger interdisciplinary, sustainable and scalable context.

4. DISCUSSION

Collecting qualitative and quantitative data from international rock art and cultural heritage researchers through conversations, consultations and surveys and analyzing 250 rock art projects listed in the Global Rock Art Database project, this

research identified a need in rock art for promoting training and use for 1) specialized digital tools; 2) integrated collaborative management tools and; 3) research specific data repositories.

Following the 2009 Intersect study on technology-enhanced practice, attitudes and requirements, a similar situation was found in rock-art seven years later. While e-research services have been developed to better understand the importance and usefulness of data, there is a need for a more discipline specific support structure considering all stages of a project and data life cycle from planning, producing to dissemination, re-use and improvement.

Project and Data Life Cycle	Data Guides	Digital Tools	Data Format Outputs	Open Access / and Open Data Repositories
Plan	ANDS 23 Things, library carpentry, Open Science Training	Atlassian, Rock Art Database, GitHub	text	Global Rock Art Database, University of California's Bancroft Library, Web of Science
Produce	Rock Art Stability Index (RASI)	LIDAR, Digital Cameras, FAIMS mobile App, Rock Art CARE App	3D, image, GPS, text	Shared Google Cloud Services
Archive	ANDS 23 Things, library carpentry, Open Science Training	RASI documentation conceptual model	none	tDAR, figshare, Dryad
Disseminate	ANDS 23 Things, library carpentry, Open Science Training			MDPI Open Access Journals, tDAR, figshare, Dryad
Search				Web of Science, tDAR, figshare, Dryad
Re-Use				tDAR, figshare, Dryad
Improve				tDAR, figshare, Dryad

Table 1. Examples for an Integrated Research Method and Project Management Approach

The research identified that researchers need more support with data management considering new data management requirements set out by international and national governing bodies. While support teams and online guides exist in, for example, the ARC data management guide, more discipline specific examples are needed.

The support for rock art researchers could be broken down into three main areas of training and resources for 1) planning for data; 2) software that looks at integrating research methods into digital project management tools and; 3) rock art specific data standards and shared repositories.

Training and resources should include tools for better understanding of data management strategies such as library carpentry, data carpentry, Open Science Training or the ANDS's 23 Things.

Further examples for software training such as software carpentry and solutions for integrating research method workflows in collaborative project management tools such as FAIMS or Atlassian's Jira.

The results further demonstrated a demand for training and resources for relevant data repositories and data standards including the formatting of rock art specific information. While platforms such as the Registry of Research Data Repositories (<https://www.re3data.org>) exist, they currently provide no information as to how this can be used in rock art research.

Considering issues identified within the data life cycle in rock art research, the following table is aimed to be a first step for rock art researchers towards improving and exploring the potential of rock art research data by mapping research method and technological tools against each step of the project and data management life cycle.

5. CONCLUSION

A need for an Integrated Research Method and Project Management Approach considering the use of research technologies has been identified.

Data is only ever going to be as good as the contributions. While it may be frustrating for some to be the first to contribute their information when the community is still incipient, these first contributors are the pioneers in an inevitable global trend. The data dark ages of hoarding information are over, or at least should be.

This paper is a call for more specialized training for rock art researchers in integrated data management. Data should not be collected with the intention of single use or to answer a few research questions. Properly stored and catalogued, data can be reused or recycled to address many more research questions that may not have been asked yet or are being asked by others in the academic community. At a time when funding in the Humanities is being threatened, researchers need to make an effort to maximize the data collected. Finally, it is the responsibility of every researcher to ensure that their data is optimized not just for themselves but for the present and future academic community.

ACKNOWLEDGEMENTS

The Global Rock Art Database is an Open Source Cultural Heritage project supported by our online community contributors and technical contributions from enthusiasts on our GitHub repository.

Author 1 conceived and designed the Global Rock Art Database project and performed all data collection and analysis for this paper; Author 2 contributed rock art gathering methods (Table 1) and further to the discussion and conclusion providing the rock art researchers perspective in this paper.

REFERENCES

Berners-Lee, T. 2009. The next web. Ted.com. Retrieved 22 June 2019, from https://www.ted.com/talks/tim_berners_lee_on_the_next_web?language=en

Chippindale, C., and Taçon P.S. 1998. The Archaeology of Rock Art. Cambridge: Cambridge University Press. ISBN10: 0521576199.

Cochran, T. 2015. Status report on the NCRIS eResearch capability. Australian Government Department of Training and Education. Available online: https://docs.education.gov.au/system/files/doc/other/abridged_eresearch_status_report_-_web.pdf (9th November 2018).

Haubt, R.A. 2015. The Global Rock Art Database: developing a rock art reference model for the RADB system using the CIDOC CRM and Australian Heritage Examples. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Science, Volume II-5/W3, 2015, 25th International CIPA Symposium 2015, 31st August – 04th September 2015, 96-98. doi:10.5194/isprsannals-II-5-W3-89-2015

Haubt, R.A., Taçon, P.S.C. 2016. A collaborative, ontological and information visualization model for a centralized rock art heritage platform. Journal of Archaeological Science: Reports. <https://doi.org/10.1016/j.jasrep.2016.10.013>

Intersect, 2009. Developing eResearch Infrastructure: Technology-enhanced research practice, attitudes and requirements. Available online: <http://www.intersect.org.au/docs/eResearchsurveysummaryreportv1.1.pdf> (9th November 2018).

Jalandoni, A., Taçon, P., & Haupt, R. (2019). A Systematic Quantitative Literature Review of Southeast Asian and Micronesian Rock Art. Advances in Archaeological Practice, 1-12. doi:10.1017/aap.2019.10

Whitley, D. Introduction to Rock Art Research. Santa Fe, USA. Left Coast Press, Inc. ISBN: 9781315426006 (e-book).