



TECHNICAL WHITE PAPER

MaDiH (مديح): Mapping Digital Cultural Heritage in Jordan.

Authors: James Smithies, Fadi Bala'awi, Pascal Flohr, Shatha Mubaideen, Alessandra Esposito, Sahar Idwan, Carol Palmer, Issa Mahasneh, Shaheer Rababeh.

Design: Shatha Mubaideen.

Published: May 2021.

TABLE OF CONTENTS



Executive summary	3
International technical context	7
Jordanian technical context & strategy	10
Methodology & processes	14
Context and overview	14
Stakeholder communications & feedback	15
Dataset identification & publication protocol	17
Data vocabulary	19
Technical analysis	23
The MaDiH (مدىح) ckan catalogue	23
MaDiH (مدىح) CKAN catalogue: technical architecture	27
Dataset characteristics	30
Technical sustainability	37
Digitisation requirements	39
Licensing and intellectual property	41
Multilingualism	43
Career development requirements	45
Future Requirements (conceptual)	46
Recommended Design & Engineering Process	48
Appendix A: MaDiH (مدىح) CKAN catalogue, enhancement recommendations	50
Appendix B: Digitisation suggestions	54
Appendix C: Technical challenges	55
Appendix D: Technical recommendations	57
Appendix E: Bibliography	60
Appendix F: Funding	66
Appendix G: Acknowledgements	67

EXECUTIVE SUMMARY



This white paper reports on the findings of the MaDiH (مديح): Mapping Digital Cultural Heritage in Jordan project.¹ It is complemented by a policy white paper². Its goal is to report on technical matters related to the project and provide advice and recommendations for future technical activity. We hope the paper is of interest to a broad audience, although it is primarily tailored to the Jordanian cultural heritage and technology communities.

MaDiH (مديح) is a collaborative UK-Jordanian project between King's Digital Lab (KDL) at King's College London, the Hashemite University (HU), the Council for British Research in the Levant (CBRL), the Department of Antiquities of Jordan (DoA), the Jordanian Open Source Association (JOSA), and the Endangered Archaeology in the Middle East and North Africa (EAMENA) project. The project ran for just over two years, from February 2019 to April 2021.

The goal of the project was to contribute to the long-term sustainable development of Jordan's digital cultural heritage by identifying key systems, datasets, standards, and policies, and aligning them to government digital infrastructure capabilities and strategies. This technical white paper describes the current state of standards, datasets, data repositories and other assets, and defines requirements for future activity. The MaDiH (مديح) Policy White Paper focuses on alignment to national and international policies and standards and makes recommendations for future action. Each section of the white papers lists associated challenges and recommendations, which are summarised in appendices. Taken together, the intention is for the white papers to define the

¹ "Newton-Khalidi Cultural Heritage and Sustainable Development in Jordan", AHRC, accessed March 16, 2021, https://webarchive.nationalarchives.gov.uk/20200923115407tf_/https://ahrc.ukri.org/funding/apply-for-funding/archived-opportunities/cultural-heritage-and-sustainable-development-in-jordan/.

² James Smithies et al, "MaDiH (مديح): Mapping Digital Cultural Heritage in Jordan. Policy White Paper"(2021).

TECHNICAL WHITE PAPER

MaDiH: Mapping Digital Cultural Heritage in Jordan

current state of digital cultural heritage (DCH) in Jordan, inform and align to Jordanian government policy, and inform planning and technical design processes that can enable the future development of Jordanian DCH.



MaDiH (مديح)'s technical goals focused on the definition of a robust technical and operational architecture for DCH to assist, among others, the Department of Antiquities in their planning processes. Analysis also aimed to identify infrastructural gaps and opportunities for further development including system development, data aggregation, and online learning; help product development teams develop their systems; facilitate the aggregation of valuable datasets held in disparate repositories; and ensure data generated from research activity is properly stored and widely accessible. Practical prototyping was used to ensure analysis and lessons learned are cost-effective and aligned to real-world scenarios.

To ensure alignment between its technical and policy goals and to provide real-world evidence of its findings, MaDiH (مديح) produced a range of digital outputs including a publicly available prototype (meta)data catalogue,³ a website⁴ to communicate project activity, and Twitter⁵ and Facebook⁶ accounts for public communication. Although intended as short-term research assets, used to build a sense of community and inform the production of the white papers, efforts have been made to sustain those assets for future use. The MaDiH (مديح) CKAN (Comprehensive Knowledge Archive Network) catalogue, is of particular importance, as it fills an urgent need for a holistic view of cultural heritage datasets held across and outside Jordan. It also collects data and content that could provide the basis for future system design, data aggregation and integration (including the ability to cross-search existing databases), and product development.

³ "MaDiH CKAN repository", the MaDiH team, 2019, accessed April 19, 2021, <https://madih-data.kdl.kcl.ac.uk/>.

⁴ "MaDiH website", the MaDiH team, 2019, accessed April 19, 2021, <http://madih-jordan.org/>.

⁵ The MaDiH team (@madih_info), 2019, accessed April 19, 2021, https://twitter.com/madih_info.

⁶ The MaDiH team, 2019, Mapping Digital Cultural Heritage in Jordan, MADIH, accessed April 19, 2021, <https://www.facebook.com/MaDiHJO/>.



The key technical findings of the MaDiH (مديح): Mapping Digital Cultural Heritage in Jordan project are:⁷

- There is a wealth of DCH data related to Jordan held locally and internationally but it is uneven in type, quality, and degree of accessibility.
- The Jordanian technology community is well positioned to partner with Jordanian universities to develop Jordan's technical DCH infrastructure.
- Jordanian DCH is well served by government open data policies, and alignment to those policies is key to future growth and sustainability.
- Jordanian institutions are developing successful partnerships with regional and international technical communities to build/develop their capabilities.
- Jordan's geographic location and political stability has supported its involvement in significant technical projects (such as EAMENA) to protect cultural heritage despite being surrounded by neighbouring countries in conflict situations.

⁷ See Appendix C for more detailed findings and recommendations.



**The key technical recommendations of the MaDiH (مديح):
Mapping Digital Cultural Heritage in Jordan project are:⁸**

- Work should be undertaken to scope, cost, and fund, the development of a National Centre for DCH in Jordan.
- National digital strategy should be updated to require the use of common international DCH technical standards related to digitisation and collection management.
- Training should be provided in research software engineering (RSE) for cultural heritage, in partnership with the education and technology sectors.
- Regional and international technical partnerships, with business, research institution, and other relevant DCH initiatives should be encouraged.



Wadi Rum
Photo by Jude Al-Safadi on Unsplash

⁸ See Appendix C for more detailed findings and recommendations.

INTERNATIONAL TECHNICAL CONTEXT



It is important to position the MaDiH (مديح) project and Jordan's wider DCH software and infrastructure in the wider international context. Issues with data management, standardization, and quality exist in all countries and adherence to best practice is always uneven both within and across institutions. In many ways, Jordan's progressive approach to Information Technology (IT) nationally, coupled with its relatively light technical DCH footprint offers significant opportunities for growth, due to a lack of technical debt that might complicate development.

National but also transnational solutions (using tools and infrastructure hosted in the United Kingdom and Europe, for example) are needed.⁹ Improved awareness and adherence to international technical standards (metadata, research data management, etc) will help Jordan achieve that. The basic IT methods needed to develop a mature DCH environment in Jordan have been practiced for decades, especially in the commercial and government sectors, but development of research and data infrastructures and associated Research Software Engineering (RSE) and methods is highly specialised and the subject of significant technical discussion globally.

The global history of DCH should inform the development of DCH strategy in Jordan. The field is connected to the development of large archaeological databases after World War II (with significant growth in the 1980s and 1990s), and associated improvements in information management (developed concurrently with broader developments in library and archival science). This resulted in increasing standardization of archaeological documentation, and

⁹ For example alignment to UKRI, UKRI, *The UK's research and innovation infrastructure: opportunities to grow our capability* (UKRI, 2020) <https://www.ukri.org/files/infrastructure/the-uks-research-and-innovation-infrastructure-opportunities-to-grow-our-capacity-final-low-res/> and European Commission, "European Strategy Forum on Research Infrastructures (ESFRI)", European Commission, accessed January 19, 2021, <https://www.esfri.eu/>, and use of domain-specific services such as <https://ariadne-infrastructure.eu/>.

paper-based registries of archaeological sites.¹⁰ The subsequent history of DCH relates to the increasing use of computers to manage and then analyse collected content, increasing in the 1970s and coming to special prominence as the Internet matured in the 1990s. Map-based and spatial data have always been central to DCH, through its close associations with archaeology, and it has been particularly adept at identifying and experimenting with new technologies, from photogrammetry to virtual reality and the Internet of Things.¹¹

The MaDiH (مديح) project is most closely associated with information management aspects of DCH, however. Relevant technical standards are often derived from initiatives instigated in the United States, United Kingdom, and the rest of Europe and are based on cross-sector standards such as the General International Standard Archival Description (ISAD) managed by the International Council on Archives (ICA)¹² and the Dublin Core Metadata Initiative (DCMI).¹³ Domain-specific standards, tailored to support cultural heritage and cognate areas, are also important. The CIDOC Conceptual Reference Model (CIDOC-CRM) is a leading standard for documentation and integration of information across the cultural heritage sector.¹⁴ The UK Forum for Information Standards in Heritage (FISH) vocabulary is another important initiative, used as one of the sources for the MaDiH (مديح) vocabulary.¹⁵

¹⁰ Peter McKeague, Rein van't Veer, Isto Huvila, Anne Moreau, Philip Verhagen, Loup Bernard, Anwen Cooper, Chris Green, and Niels van Manen., "Mapping Our Heritage: Towards a Sustainable Future for Digital Spatial Information and Technologies in European Archaeological Heritage Management". *Journal of Computer Applications in Archaeology*, 2(1), (2019): 89–104, DOI: <http://doi.org/10.5334/jcaa.23>.

¹¹ Davide Tanasi, "The Digital (within) Archaeology. Analysis of a Phenomenon", *The Historian* 82, no. 1 (2 January 2020): 22–36, DOI: <https://doi.org/10.1080/00182370.2020.1723968>. See the MaDiH (مديح) Policy White Paper for more information.

¹² International Council on Archives, *General International Standard Archival Description* (ICA, 1999) https://www.ica.org/sites/default/files/CBPS_2000_Guidelines_ISAD%28G%29_Second-edition_EN.pdf.

¹³ "Dublin Core™ Metadata Initiative", DCMI, accessed February 24, 2021, <https://dublincore.org/>.

¹⁴ "Conceptual Reference Model", International Council of Museums & International Committee for Documentation (ICOM CIDOC), accessed February 24, 2021, <http://www.cidoc-crm.org/>.

¹⁵ "FISH Vocabularies", Forum on Information Standards in Heritage (FISH), accessed February 24, 2021, <http://www.heritage-standards.org.uk/fish-vocabularies/>.



McKeague *et al.* make a strong case for investment in information management standards to support DCH:

Since the quality of research results and heritage management decisions is highly dependent on the nature of the available data, issues of sustainability of digital data repositories, accessibility and reliability of data, standardization of data formats and management of property rights are currently widely debated.¹⁶

This is especially the case with projects like MaDiH (مدىح) that would ideally not only list individual datasets but also aggregate data across them, allowing data collected in one instance to be cross-searched, and augmented with data from another. Provision of Application Programming Interfaces (API) allows this to be done in a programmatic fashion, adding additional value.¹⁷ This requires strict adherence to international metadata standards to enhance interoperability, and ensure databases can 'talk' to each other. Best practice now aims for data to be findable, accessible, interoperable, and reusable (FAIR), using metadata standards appropriate to the knowledge domain in question.¹⁸ This needs to be balanced with cultural and political factors related to open data but aligns well to Jordanian government policy in the area.¹⁹ DCH tools such as the CIDOC CRM-EH (English Heritage) extension, and semantic tools such as Semantic Technologies

¹⁶ Peter McKeague, Rein van't Veer, Isto Huvila, Anne Moreau, Philip Verhagen, Loup Bernard, Anwen Cooper, Chris Green, and Niels van Manen., "Mapping Our Heritage: Towards a Sustainable Future for Digital Spatial Information and Technologies in European Archaeological Heritage Management". *Journal of Computer Applications in Archaeology*, 2(1), (2019): 89–104, DOI: <http://doi.org/10.5334/jcaa.23>.

¹⁷ Günter Waibel, Ralph LeVan, and Bruce Washburn, "Museum Data Exchange: Learning How to Share", *D-Lib Magazine* 16, no. 3/4 (March 2010), <https://doi.org/10.1045/march2010-waibel>.

¹⁸ Wilkinson et al., "The FAIR Guiding Principles for Scientific Data Management and Stewardship", *Scientific Data* 3, no. 1, 15 March 2016m 160018, <https://doi.org/10.1038/sdata.2016.18>.

¹⁹ Jonathan Gray, "What Do Data Portals Do? Tracing the Politics of Public Information Infrastructures on the Web" (UK: Data Publics, Lancaster University, 2017).

TECHNICAL WHITE PAPER

MaDiH: Mapping Digital Cultural Heritage in Jordan

for Archaeological Resources (STAR)²⁰ and Semantic Technologies Enhancing Links and Linked data for Archaeological Resources (STELLAR)²¹ have been developed to enhance data integration, cross-searching, and machine-readability of DCH content.²²



CHALLENGES

- Improve awareness of and adherence to international metadata standards.

RECOMMENDATIONS

- Align to RI roadmaps and DCH infrastructure in the UK and Europe, to ensure Jordan establishes robust national policies.

JORDANIAN TECHNICAL CONTEXT & STRATEGY

Jordan faces significant challenges building a mature DCH technical environment, despite its regional advantages in IT more generally. The experience of MaDiH (مدىح) suggests that while good understanding of international standards exists in cultural heritage institutions and major database-based research projects, making it feasible for them to scale up their DCH activities relatively rapidly and connect their data with other similarly mature projects, the majority of data described in the MaDiH (مدىح) CKAN data catalogue does not conform to international information management standards. Anecdotal evidence suggests DCH training has occurred in the past but the staff trained through those initiatives have now retired. The next generation of early career professionals need to be trained to provide continuity and the human resources needed for the demanding work of managing existing information, and designing and building standards-based

²⁰ "STAR", Hypermedia Research Group, accessed March 17, 2021
<https://hypermedia.research.southwales.ac.uk/kos/star/>

²¹ "STELLAR", Hypermedia Research Group, accessed March 17, 2021,
<https://hypermedia.research.southwales.ac.uk/kos/stellar/>

²² Douglas Tudhope et al., "Connecting Archaeological Data and Grey Literature via Semantic Cross Search", *Internet Archaeology*, no. 30 (2011),
<https://doi.org/10.11141/ia.30.5>.

TECHNICAL WHITE PAPER

MaDiH: Mapping Digital Cultural Heritage in Jordan

databases. The EAMENA-CPF Cultural Protection Fund 2018-2020 training programme is a good start, in which 40 heritage professionals from the Department of Antiquities (DoA) were trained in DCH; however, the focus was mostly on remote sensing and the end user experience, and only to a limited extent on the overall project development cycle inclusive of database design, development and maintenance.²³ Additional effort needs to be placed on knowledge transfer from established global centres of excellence in DCH, including post-graduate training at universities abroad, and professional placements at cultural heritage institutions. This need for training and technical development presents significant challenges that will require time and funding to resolve. As it stands, the MaDiH (مديح) CKAN catalogue represents a view into the current state of Jordan's DCH data landscape, providing evidence for future planning, rather than presenting a range of datasets amenable to comprehensive analysis and ready for aggregation.²⁴ This would be greatly enhanced if teams developing data related to Jordanian cultural heritage actively engaged with local communities during the design phase, partnering with them to ensure appropriate management which might extend, as it has in the case of MaDiH (مديح), to post-project management by Jordanian stakeholders. In an ideal situation, local stakeholders would have the ability to maintain data over the long term.

Jordan's DCH databases or registers are separated between different heritage types (tangible heritage: 'archaeology' before 1750 CE, 'heritage' after 1750 CE; intangible heritage), due to their legal differences (see the policy white paper). Several are of particular importance, for example the DoA's Middle Eastern Geodatabase for Antiquities or MEGA-Jordan archaeological site database.²⁵ This

²³ "Training in Endangered Archaeology", British Council, accessed February 3, 2021, <https://www.britishcouncil.org/arts/culture-development/cultural-protection-fund/projects/endangered-archaeology>; "The Cultural Protection Fund", EAMENA project, accessed February 3, 2021, <https://eamena.org/cultural-protection-fund>

²⁴ An incomplete view. See the MaDiH (مديح) Policy White Paper (data analysis section) for more information.

²⁵ "MEGA-Jordan", Department of Antiquities, the Getty Conservation Institute and World Monuments Fund, accessed January 19, 2021, <http://megajordan.org/>; David Myers and Alison Dalgity, "The Middle Eastern Geodatabase for Antiquities (MEGA): an open source GIS-based heritage site inventory and management system" *Change Over Time* 2, no. 1 (2012): 32-57.

online, partly open access database was based on the 1990s Jordan Antiquities Database and Information System (JADIS)²⁶ and is a forerunner of Arches.²⁷ It is currently being prepared for being upgraded to Arches version 5 in collaboration with the EAMENA project and the Getty Conservation Institute. For movable tangible heritage (artifacts, or objects), there is currently no central database in Jordan, but the Documentation of Objects in Jordanian Archaeological Museums (DOJAM) project's²⁸ prototype for the Jordan Archaeological Museum will be developed for this in the next couple of years. For post-1750 CE tangible heritage in Jordan there is also no central database, but information in the Greater Amman Municipality's register of heritage houses in Amman could form the basis of this (currently Arabic only).²⁹ For intangible heritage, the Ministry of Culture's National Project for the Documentation of Intangible Heritage currently has an inventory of five of the country's twelve governorates (in Arabic only).³⁰

This white paper recommends that DCH is further developed in Jordan using these databases and registers, and many additional datasets listed in the MaDiH (مدىح) CKAN catalogue, as its foundations. By improving the information management landscape across those existing assets immediate practical benefits will be realised, in alignment to government open data policy³¹ and the operational needs

²⁶ Gaetano Palumbo, "JADIS: the Jordan Antiquities Database and Information System; A Summary of the Data." (1994).

²⁷ "Arches", Getty Conservation Institute and World Monuments Fund, accessed January 19, 2021, <https://www.archesproject.org/>

²⁸ "Documentation of Objects in Jordanian Archaeological Museums (DOJAM)", GPIA Amman, accessed November 20, 2020, <https://www.zitadelle-amman.de/projekt/>. The DoA also requires archaeological missions to submit lists of objects found, which could be incorporated in such a database to expand it beyond museum objects alone.

²⁹ "Greater Amman Municipality", GAM, accessed November 20, 2020, <https://www.ammancity.gov.jo/ar/main/index.asp>; "Amman Architectural Heritage Record", MaDiH project, accessed February 3, 2021, <https://madih-data.kdl.kcl.ac.uk/dataset/amman-heritage-houses-record>.

³⁰ "Intangible heritage of Jordan publications", the Ministry of Culture, Heritage Directorate, accessed January 19, 2021, <http://ich.gov.jo/node/52332>; "The National Project for the Documentation of Intangible Heritage", MaDiH project, accessed February 3, 2021, <https://madih-data.kdl.kcl.ac.uk/dataset/the-national-project-for-the-documentation-of-intangible-heritage>

³¹ Ministry of Planning and International Cooperation. *Fourth National Action Plan 2018 - 2020 under the Open Government Partnership Initiative (OGP)* (2018). https://www.opengovpartnership.org/wp-content/uploads/2019/01/Jordan_Action-Plan_2018-2020.pdf.

TECHNICAL WHITE PAPER

MaDiH: Mapping Digital Cultural Heritage in Jordan

of the DoA, the other support to ministries/departments involved in cultural heritage,³² and the higher education sector. By creating a more standards-based information management landscape, cultural heritage content that has already been digitised will be easier to manage and access, additional investment in new repositories will be possible to support areas of high demand, and high priority content will be able to be prioritised for digitisation. The case for investing in a wide-ranging strategy for information standards improvement across the 'data terrain' identified by MaDiH (مدیح) is particularly strong because of the reliance on cultural heritage across the Jordanian economy.³³ By standardising the country's core DCH data a foundation will be created that can be leveraged by a variety of government and private sector parties.



CHALLENGES

- Improve national information management standards, in support of government open data policies.
- Find funding to fill gaps in the technical landscape identified through the MaDiH (مدیح) CKAN catalogue.
- Find funding for training and study opportunities for the next generation of DCH professionals.
- Involve Jordanian stakeholders in the early design of datasets and user interfaces.

RECOMMENDATIONS

- Use the MaDiH (مدیح) CKAN catalogue as the foundation of a national DCH initiative capable of enhancing government, higher education, and private sector activities.
- Leverage Jordan's capabilities in IT to position itself as a world leader in DCH.
- Enhance Jordanian data sovereignty and long-term data sustainability through innovative partnerships.
- Attract a new generation of professionals to DCH, through study, training, and employment opportunities.

³² E.g. the Ministry of Tourism and Antiquities and the Ministry of Culture. See also the MaDiH Policy White Paper.

³³ See the MaDiH (مدیح) Policy White Paper for more information.

METHODOLOGY & PROCESSES



Context and overview

The technical context of the MaDiH (مدیح) project, and in particular the MaDiH (مدیح) CKAN catalogue, is straightforward. The project was conceived as a basic first step in the development of capacity in Jordanian DCH, in a broad sense encompassing technical, operational, and human issues.³⁴ The intention was to make a preliminary analysis of Jordan's cultural heritage (CH) data landscape, with a view to identifying opportunities for growth and understanding the current technical state of its supporting DCH infrastructure. The MaDiH (مدیح) CKAN catalogue was to act as a central place to document datasets identified during dataset collection phase and facilitate their analysis. In developing the MaDiH (مدیح) project community and by identifying datasets through desk research, word of mouth, and stakeholder workshops, the intention was to build awareness about the issues across the DCH community and define minimal technical standards, analytical methods, and project management processes capable of scaling over time. Supporting activities such as providing Research Software Engineering (RSE) training for the project team aimed to contribute to that process.³⁵

The MaDiH (مدیح) CKAN catalogue should be considered the Minimum Viable Product (MVP) for realising those goals. It was originally intended to be kept live for two years after the end of project funding, but has proved so valuable for the Jordanian DCH community that additional effort was made to migrate it to Hashemite University for long-term maintenance in Jordan. It is possible that it will be migrated again, to another host within Jordan,³⁶ if deemed appropriate as DCH capability develops.

³⁴ Refer to the MaDiH (مدیح) Policy White Paper for a broad overview.

³⁵ "MaDiH: Research Software Engineering Training", accessed March 17 2021, <https://campus.dariah.eu/resource/rse2019>

³⁶ For example a national cultural heritage institution.

Stakeholder communications & feedback



The MaDiH (مديح) project team implemented a communications strategy soon after project initiation, to ensure a consistent approach with stakeholders and a consistent approach to data collection. Stakeholders were identified through desk research (websites, reports, journal articles), workshops, lectures and events, and professional networks (word of mouth). Useful sources of information included institutional websites such as the Department of Antiquities list of archaeological museums³⁷ and the Department of Culture list of museums and cultural organisations.³⁸ The 2014 UNESCO report on Jordanian Documentary Heritage was another important resource.³⁹ Contact was made using various channels, including in person meetings,⁴⁰ emails, phone conversation, public lectures or events and websites/ social media pages (Facebook, Twitter).

A bi-lingual data collection template⁴¹ was used to record information about individual datasets, recording a range of information aligned to the MaDiH (مديح) data vocabulary (see below). Information gathered was subject to ethical management in line with UK funding requirements, the European General Data Protection Regulation (GDPR) and, where relevant, King's College London research ethics processes. A basic WordPress website was hosted on Reclaim Hosting⁴² for general project communications.⁴³

³⁷ "Museums", The Department of Antiquities, accessed February 24, 2021, <http://www.doa.gov.jo/All-Museums.aspx>.

³⁸ "Museums", The Ministry of Culture, accessed February 24, 2021, <https://culture.gov.jo/museums>.

³⁹ Ababsa, Mariam, Jordan Documentary Heritage: Part 1, UNESCO Amman Office (2014), http://www.unesco.org/new/fileadmin/MULTIMEDIA/FIELD/Amman/pdf/Jordan_Documentary_Heritage.pdf

⁴⁰ Face to face meetings were often requested due to uncertainty about the nature of the MaDiH (مديح) project and its goals, and local professional expectations. This slowed data collection but greatly improved the development of a supportive project community.

⁴¹ MaDiH Team, "The MaDiH data collection form", Zenodo (2020), <http://doi.org/10.5281/zenodo.4289919>.

⁴² <https://reclaimhosting.com/>.

⁴³ <http://madih-jordan.org/>. At the time of writing the intention is to convert this website to a static site (or archive it by other means) after project funding finishes, or transfer control of it to a Jordanian-based partner for ongoing maintenance.

TECHNICAL WHITE PAPER

MaDiH: Mapping Digital Cultural Heritage in Jordan

A total of 50 key organisations were contacted, across the following categories: government; education; international research institutions; private initiatives, professionals, and associations. Further stakeholders were identified in iterative fashion. Responses were generally positive. In some cases, emails were enough to initiate data collection but in other cases meetings were organised or official letters signed by the Council for British Research in the Levant (CBRL, the lead UK project partner in Jordan) were requested. The COVID-19 pandemic impacted data collection in some cases, due to lockdowns in Jordan halting or slowing work. It is worth noting that many stakeholders were resistant to the idea of providing open access to their metadata due to concerns about security and access control, suggesting a need for technical education and training on this topic.

Stakeholder expectations and concerns included:

Expectations	Concerns
<ul style="list-style-type: none">• Assistance digitizing content.• Financial assistance.• Education and training.• Aggregation of their dataset content into the MaDiH (مديح) CKAN catalogue.• Access to the MaDiH (مديح) CKAN catalogue to update their project information.• Long-term maintenance of the MaDiH (مديح) CKAN catalogue.⁴⁴	<ul style="list-style-type: none">• That the project intended to harvest their datasets rather than only record high-level information (metadata) about their dataset.• The need for regular updates about the progress of the project, to align to internal DCH / digitisation plans.• Sustainability and management of the MaDiH (مديح) CKAN catalogue after the project finishes.• Content translation (bilingual website).• The ability of the project to find and map all the available datasets.⁴⁵• Link rot (breakage of urls over time, as datasets moved or were taken offline).

⁴⁴ The MaDiH (مديح) CKAN catalogue was initially conceived as a proof of concept output with a limited life-time.

⁴⁵ As discussed in the MaDiH (مديح) Policy White Paper, this was not the intention of the project, which aimed instead to undertake a representative 'first pass' of the available datasets to inform preliminary landscape analysis.



CHALLENGES

- Improve awareness of the technical requirements and value of open access strategies.
- Maintain project continuity after funding for Phase 1 ends.
- Maintain the CKAN catalogue after funding ends.
- Ensure Jordanian data sovereignty is enabled.

OPPORTUNITIES

- Build on the awareness gained by the MaDiH (مدیح) project to enhance open access initiatives and standardise national DCH datasets.
- Transition a proof of concept DCH catalogue into a long-term national asset hosted in Jordan.
- Bolster Jordanian data sovereignty by demonstrating best practices in migrating the MaDiH (مدیح) data catalogue to Jordan.

Dataset identification & publication protocol

Before communication outside the project was initiated, a dataset identification and publication protocol was established, to ensure data collection occurred in a consistent and transparent manner.⁴⁶ The intention was also to ensure data quality, sustainability, and national and international interoperability. A subsidiary purpose of the protocol was as a communication tool, to assure stakeholders the project could be trusted to collect and manage information about their data appropriately. The document included a statement of values & principles for this

⁴⁶ Alessandra Esposito, Shatha Mubaideen, Pascal Flohr, James Smithies, Fadi Bala'awi, "MaDiH (مدیح) Mapping the Digital Cultural Heritage in Jordan Project. Datasets Identification and Publication Protocol", Zenodo (2020), p. 4. 10.5281/zenodo.4146756.

TECHNICAL WHITE PAPER

MaDiH: Mapping Digital Cultural Heritage in Jordan

reason, noting adherence to the European GDPR framework, and King's College London Statement on Use of Personal Data in Research.⁴⁷



It also provided information about how data can be removed from the MaDiH (مديح) CKAN catalogue, and information about the purpose of the project, project partners, and team members. Information about internal project workflows related to dataset identification, selection, quality assurance, and publication were included to ensure stakeholders were clear about the mechanisms used to identify and include or exclude datasets from the CKAN catalogue.

Additional information conveyed the priority given to datasets with different characteristics so that stakeholders could better understand issues related to data collection and aggregation:

Type of Dataset	Priority	Complexity
Digital online	High	Low
Digital offline	Low	Medium
Analogue	Low	High (approval needed)

The most important information in the protocol related to the data dictionary and vocabulary (see below), which determined and controlled the range of data being collected and the format it was to be collected in.

⁴⁷ "Statement on Use of Personal Data in Research.", King's College London, accessed 23 November 2020, <https://www.kcl.ac.uk/research/support/rgei/research-ethics/kings-college-london-statement-on-use-of-personal-data-in-research#>.



Data vocabulary

The MaDiH (مدىح) vocabulary⁴⁸ implemented in the MaDiH (مدىح) CKAN catalogue (Fig. 1) is based on the expected types/features of the dataset landscape and inspired by glossaries used by EAMENA⁴⁹ and other heritage institution projects such as the Getty Conservation Institute's Arches⁵⁰ and FISH.⁵¹ Data was initially collected in a datasets identification template spreadsheet. The template was based on a pilot CKAN MaDiH (مدىح) data template developed by the King's Digital Lab research software engineering team,⁵² with the addition of two custom fields: Dataset Time Period and Data Type. Both the custom fields were associated with a closed list of values. The Dataset Time Period list is based on the period list used in the MEGA-Jordan project,⁵³ which was very slightly adapted. The Data Type list was based on the Metadata Schema for the Description of Research Repositories⁵⁴ and the DataCite Metadata Schema Documentation for the Publication and Citation of Research Data.⁵⁵ To ensure consistent data quality, the standardized vocabularies were used wherever possible and enforced through drop-down menus in the MaDiH (مدىح) CKAN catalogue. However, due to the limited potential to customise CKAN in this phase of the MaDiH (مدىح) project, only the two mentioned fields could be added. Other information deemed important to the project therefore had to be added as 'tags' (see Figure 1). A standardized list for each of the tags is part of the MaDiH (مدىح) controlled vocabulary, but, as it is not

⁴⁸ "The MaDiH Vocabulary", The MaDiH Team, accessed February 24, 2021, <https://doi.org/10.6084/m9.figshare.12950879.v1>.

⁴⁹ "EAMENA", The EAMENA project, accessed January 19, 2021, <https://eamena.org/>.

⁵⁰ "Arches", Getty Conservation Institute and World Monuments Fund, accessed January 19, 2021, <https://www.archesproject.org/>.

⁵¹ "FISH Vocabularies", Forum on Information Standards in Heritage (FISH), accessed February 24, 2021, <http://www.heritage-standards.org.uk/fish-vocabularies/>.

⁵² Arianna Ciula "Exposing Legacy Datasets" (2020), accessed March 17 2021, <https://kdl.kcl.ac.uk/blog/legacy-project-datasets/>.

⁵³ "MEGA-Jordan", Department of Antiquities, the Getty Conservation Institute and World Monuments Fund, accessed January 19, 2021, <http://megajordan.org/>.

⁵⁴ Jessika Rücknagel, Paul Vierkant, Robert Ulrich, Gabriele Kloska, Edeltraud Schnepf, David Fichtmüller, Evelyn Reuter et al. "Metadata schema for the description of research data repositories." (2015) https://gfzpublic.gfz-potsdam.de/rest/items/item_1397899_6/component/file_1398549/content.

⁵⁵ DataCite Metadata Working Group. "DataCite Metadata Schema Documentation for the Publication and Citation of Research Data". Version 4.0. DataCite e.V. (2016) <http://doi.org/10.5438/0012>.

TECHNICAL WHITE PAPER

MaDiH: Mapping Digital Cultural Heritage in Jordan

enforced during data entry, typos and inconsistencies occurred.⁵⁶ Free text was allowed in all other fields, but, as for the tags, the standardized vocabulary and detailed guidelines were adhered to when entering data in these fields.

The MaDiH (مدیح) vocabulary implemented the bi-lingual values of the project, with particular care being taken to define preferred spellings for geographic locations (governorates, city districts), and cultural / archaeological periods. Limitations in language localisation of the CKAN open source catalogue tool⁵⁷ meant that the vocabulary implemented in the MaDiH (مدیح) CKAN catalogue could not be made fully bi-lingual, although contributions were made to the upstream Arabic CKAN project, and relationships established that could improve the system over time.

CKAN metadata	Description
Title	Name of the dataset as indicated by its creators. If an acronym, also indicate what it stands for.
URL (ckan)	URL identifying the dataset in the MaDiH catalogue, e.g. data.kdl.kcl.ac.uk/dataset/eamena
Project URL	Main URL of the dataset, i.e. its Home page.
Project Principal and Co-Investigators	Names of the dataset PIs and Co-Is.
Project Team	Names and titles of the team members accredited in the project.
Project Start Date	Start date of the project expressed in the format yyyy/mm/dd .
Project End Date	End date of the project expressed in the format yyyy/mm/dd .

⁵⁶ Considerable effort was made to avoid these and to correct these during analysis.

⁵⁷ <https://ckan.org/>.

TECHNICAL WHITE PAPER

MaDiH: Mapping Digital Cultural Heritage in Jordan



CKAN metadata	Description
Project Funder(s)	Names of the individuals, institutions, and/or organisations that funded the project.
Description	One paragraph providing an overview of the content of the dataset.
Tags	<ol style="list-style-type: none">1. Digital or Analogue2. Online or Offline3. Geographical provenance of the records: Region / Governorate / Site name in Arabic and English4. pre-1750 and/or post-1750 CE5. Artefact type6. inside_Jordan OR outside_Jordan7. dataset_location_country (e.g. dataset_location_France)8. Dataset language (e.g. dataset_language_Arabic)9. Heritage Type_Tangible-Movable_Third level (e.g. Archaeological Site, Object, etc)
Licence	Licence associated with the dataset.
Organisations	Organisation or institution that owns or holds the dataset.
Visibility	Set as ' Private ' or ' Public ' depending on whether the dataset is published on CKAN.
Project Status	Development status of the project to be marked as: ' Completed ' if the project has ended, or ' Ongoing ' if the project is still under development.
Source	Source of the dataset (usually the project main website URL).
Author	Name of the person/people or the organisation that produced the dataset.
Author email	Work/Business Email contact to enquire about the data recorded in the dataset.

TECHNICAL WHITE PAPER

MaDiH: Mapping Digital Cultural Heritage in Jordan



CKAN metadata	Description
Maintainer	Name of the person/people or the organisation responsible for maintaining the dataset, indicating the postal address (if known). It could be the same as the author.
Maintainer Email	Email contact to enquire about the digital output of the dataset.
Dataset Time period	Expressed in cultural period, as found on MEGA-Jordan
Data Type	Type of dataset expressed according to the MaDiH (مديح) authority list.

Figure 1. MaDiH (مديح) Dataset Identification Template.⁵⁸

CHALLENGES

- Improve awareness of the technical requirements and value of open access strategies.
- Maintain project continuity after funding for Phase 1 ends.
- Maintain the CKAN catalogue after funding ends.
- Ensure Jordanian data sovereignty is enabled.

OPPORTUNITIES

- Use the processes and vocabulary developed by the MaDiH (مديح) project for future data identification and collection activities in Jordan.
- Contribute to the ckan.org open source project, to improve Arabic language support.

⁵⁸ See also the CKAN User Guide, <https://docs.ckan.org/en/latest/user-guide.html>.



TECHNICAL ANALYSIS

The MaDiH (مديح) CKAN catalogue

CKAN⁵⁹ is an open-source data management system designed to facilitate the storage and distribution of open data. Its code is available on GitHub,⁶⁰ and consists of Python and Javascript. Data is stored in a PostgreSQL database, accessible via Web Application Programming Interfaces (API). The project is inspired by the Debian Linux operating system, with a plug-in architecture to ensure additional functionality can be added by its open-source community.⁶¹ The project is managed by the Open Knowledge Foundation⁶² and is used by a wide variety of international governmental and cultural heritage organisations. CKAN is notable for being used as the system of choice for multiple national open data portals, including the United States,⁶³ Europe,⁶⁴ Canada,⁶⁵ and Australia.⁶⁶

The MaDiH (مديح) CKAN (meta)data catalogue⁶⁷ was designed as a proof of concept, to be maintained for two years following the end of project funding in 2021. The intention was to use it as a:

- tool to collect metadata about identified datasets (but not data contained in these) for analysis;
- focal point for communication of project findings, education, and training;
- focus for the development of the data vocabulary and identification of appropriate technical standards;
- model and potential basis for the development of a future national cultural heritage or open data portal.

⁵⁹ <https://ckan.org/>.

⁶⁰ <https://github.com/ckan/ckan>.

⁶¹ <https://en.wikipedia.org/wiki/CKAN>.

⁶² <https://okfn.org/>.

⁶³ <https://www.data.gov/>.

⁶⁴ <https://www.europeandataportal.eu/en>.

⁶⁵ <https://open.canada.ca/en>.

⁶⁶ <https://data.gov.au/>.

⁶⁷ <https://madih-data.kdl.kcl.ac.uk/>.

TECHNICAL WHITE PAPER

MaDiH: Mapping Digital Cultural Heritage in Jordan



After 12 months and two stakeholder workshops it became clear that the catalogue was viewed as a significant national asset that would ideally be maintained over the long term. Issues of data sovereignty and cultural ownership suggested that more permanent hosting should move to Jordan, from its original host King's Digital Lab (UK).⁶⁸

The catalogue is hosted on a virtual server (VMWare) using a configuration similar to the recommended production deployment configuration described in the CKAN technical documentation:⁶⁹

CPU Cores	2
RAM	4GB
Disk	500M (Boot) + 16GB
OS	Ubuntu 16.04
Web Server	Apache
Database	1 x Postgres DB
Additional	Solr + Jetty

As a proof of concept output it was expected that the catalogue would only support basic project requirements, with minimal funding for customisation. CKAN was chosen because it was assumed - correctly - that it was likely to support most requirements 'out of the box' and could be enhanced through plugins or further development if funding allowed. A list of recommendations for MaDiH (مدىح) CKAN enhancements, should future funding allow further development, is available in Appendix C.

⁶⁸ Further migration to another Jordanian institution might occur in the future, pending community consultation and technical capacity.

⁶⁹ <https://ckan.org/documentation-and-api/>.

TECHNICAL WHITE PAPER

MaDiH: Mapping Digital Cultural Heritage in Jordan

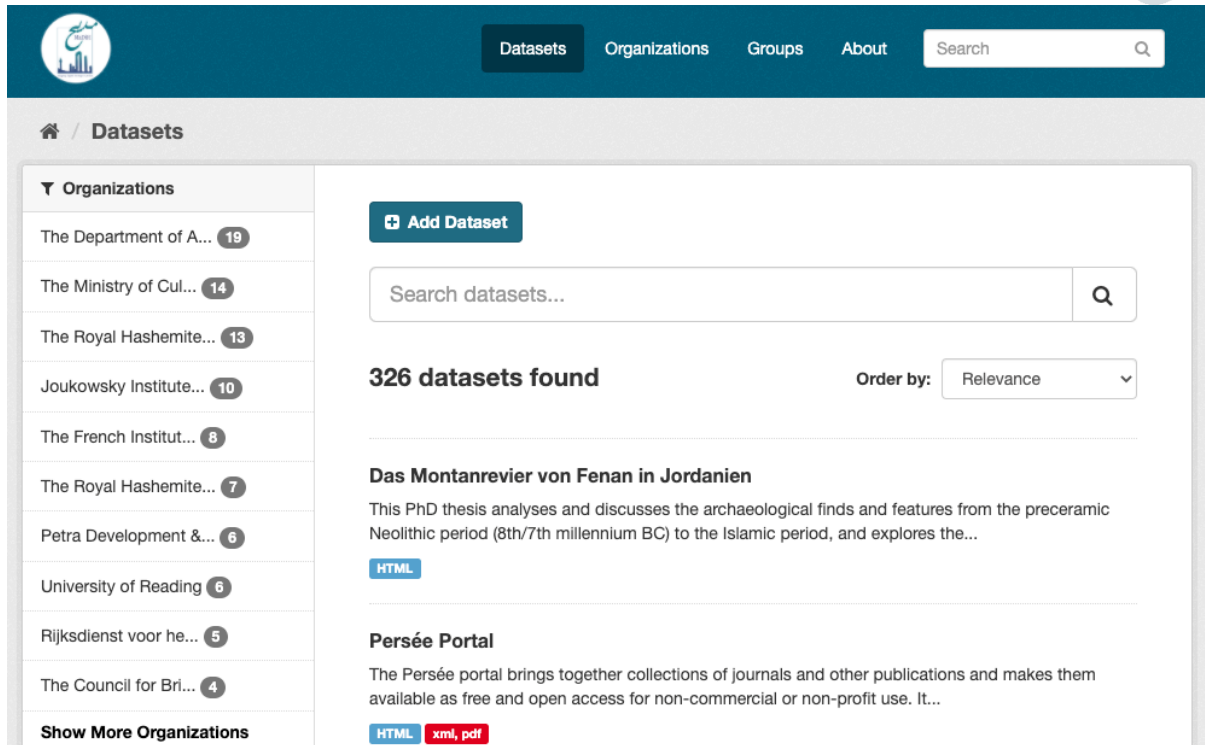


Figure 2. MaDiH (مديح) CKAN User Interface.

MaDiH (مديح) project: operational context

Figure 3 describes the operational context the MaDiH (مديح) project worked within. Grey boxes indicate operational contexts the project focused on. Blank boxes indicate operational contexts the project was aware of but chose not to engage with to control scope. Project assumptions:

- I. Although focused on Jordanian content it was necessary to design the project using international technology and metadata standards to enable scalability.

TECHNICAL WHITE PAPER

MaDiH: Mapping Digital Cultural Heritage in Jordan

- II. Although using United Kingdom funding it was necessary to implement technical solutions that could allow for maintenance or extension using continued UK or alternative international funding.
- III. Immediate focus needed to be placed on cultural heritage sector content, but the project could potentially be scaled in the future to also gather content related to natural heritage, digital social science (for example data related to refugees), Science, Technology, Engineering and Mathematics (STEM), and the business & technology sector. If scaled to include such content its scope would be widened significantly, however, requiring major funding from government and/or external sources. In such a scenario MaDiH (مدىح) would function as a national data portal similar to <https://open.canada.ca/en/open-data> or <https://data.gov.uk/>.

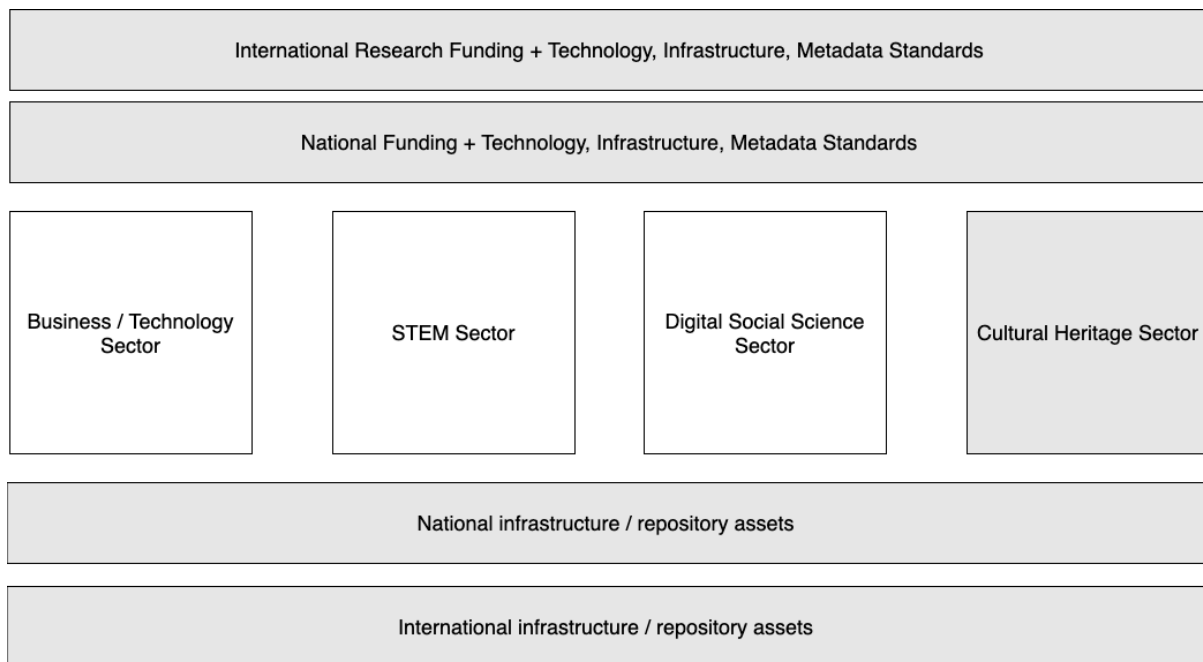


Figure 3. MaDiH (مدىح) Project: Operational Context



CHALLENGES

- The project is well positioned to scale up but requires additional funding to do so.
- The project could be scaled to support other sectors.
- The project could be used as a model for similar repositories supporting other sectors.

OPPORTUNITIES

- If scaled up, part of the project should remain focused on the cultural heritage sector.

MaDiH (مدیح) CKAN catalogue: technical architecture

It became clear over the course of the project that stakeholders did not always understand the nature and purpose of the MaDiH (مدیح) CKAN catalogue. The diagrams below provide conceptual portrayals of the catalogue's current (Fig. 3) and possible future (Fig. 4) states.

Current state

The catalogue is currently in its most rudimentary form. Metadata describing datasets held in various locations inside and outside Jordan (in databases and websites, but also sometimes offline) have been manually entered into the MaDiH (مدیح) CKAN catalogue. Users can search the catalogue from <https://madih-data.kdl.kcl.ac.uk/>⁷⁰ but need to visit external sites (mostly digital, sometimes physical) to view the content.

⁷⁰ This URL may have changed after the publication of this paper, following migration of the catalogue to Hashemite University, but a redirect should be in place.

TECHNICAL WHITE PAPER

MaDiH: Mapping Digital Cultural Heritage in Jordan

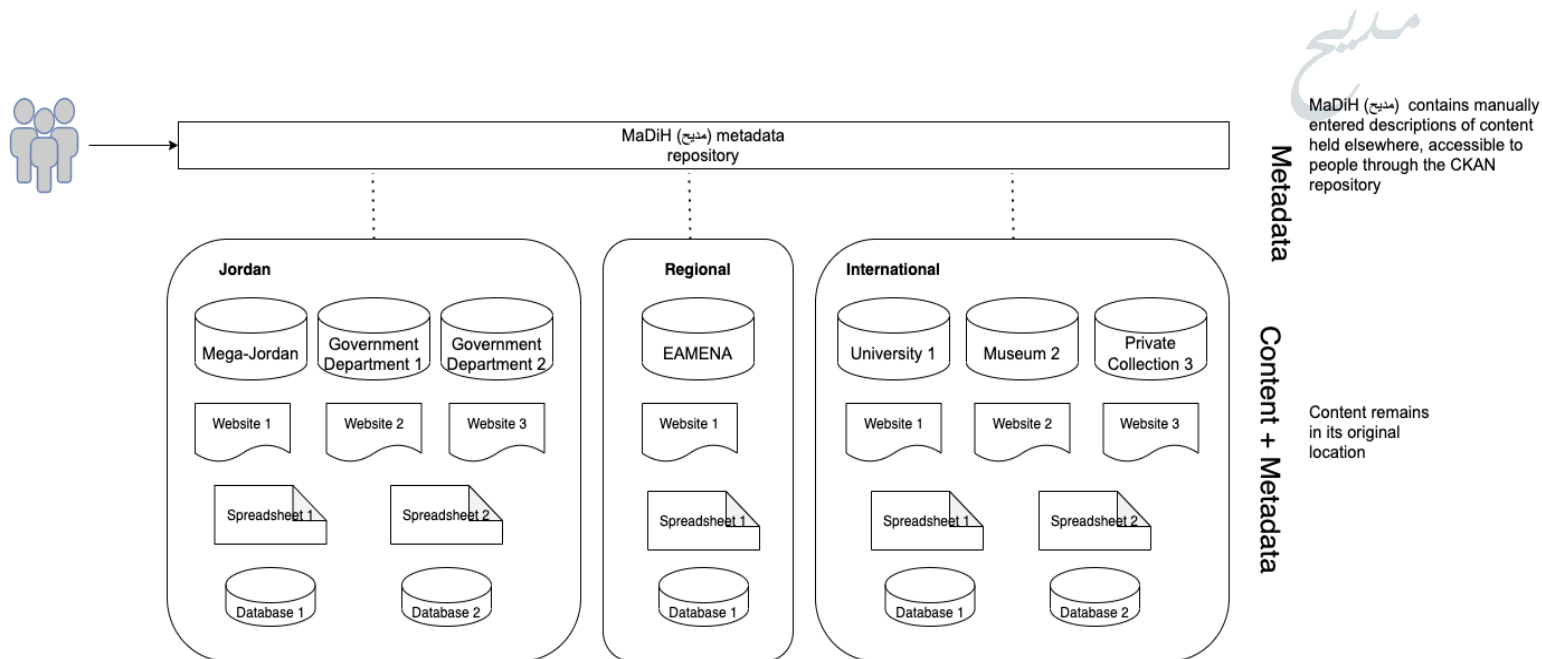


Figure 4. MaDiH (مدیح) CKAN catalogue: technical architecture (current state), using the archaeological site databases MEGA-Jordan and EAMENA as key examples.

Possible future state

With additional funding, it would be feasible to start to automate the metadata collection process to some degree, turning <https://madih-data.kdl.kcl.ac.uk/> into a federated catalogue (Fig. 5).⁷¹ Metadata content would be automatically harvested from web services with Application Programming Interfaces (API) and mapped into the MaDiH (مدیح) CKAN data schema. In many cases, this would not be technically feasible, either through the lack of an API, incompatible or inadequate data standards, or licensing and data ownership issues. In the best-case scenarios, however, it would also be feasible to make the *content* (open access to data) in those locations available through <https://madih-data.kdl.kcl.ac.uk/> as well. MaDiH (مدیح) would then become a national cultural heritage data portal similar in architecture to *Europeana*.⁷² A more focused

⁷¹ R. Rankin, M. Gordon, and R. Potter, 'Implementing a Federated Data Archive with Asynchronous Data Query, Gathering and Analysis Capabilities', *AGU Fall Meeting Abstracts* 33 (1 December 2009): SM33B-1570.

⁷² "Europeana Collections", Europeana, accessed January 19, 2021, <https://www.europeana.eu/en/collections>.

TECHNICAL WHITE PAPER

MaDiH: Mapping Digital Cultural Heritage in Jordan

example of a national cultural heritage archive is *CEISMIC*.⁷³ It should be noted that the process of developing a national CH catalogue is valuable in and of itself, however, even if the perfect archive is never realised (an impossible outcome). In identifying and assessing potential content, improving its quality, and updating it to adhere to international standards a great deal can be learned and vibrant communities established. This was the case with New Zealand’s DigitalNZ service, which began as a small project but scaled to become the country’s central DCH asset.⁷⁴

The architecture described in Figure 5 implies significant funding and effort, and robust alignment to the Department of Antiquities and government strategy and policy. The current MaDiH (مدىح) CKAN catalogue provides the seed of such an effort, although it would be advisable to use it as a conceptual model rather than scaling the existing implementation.

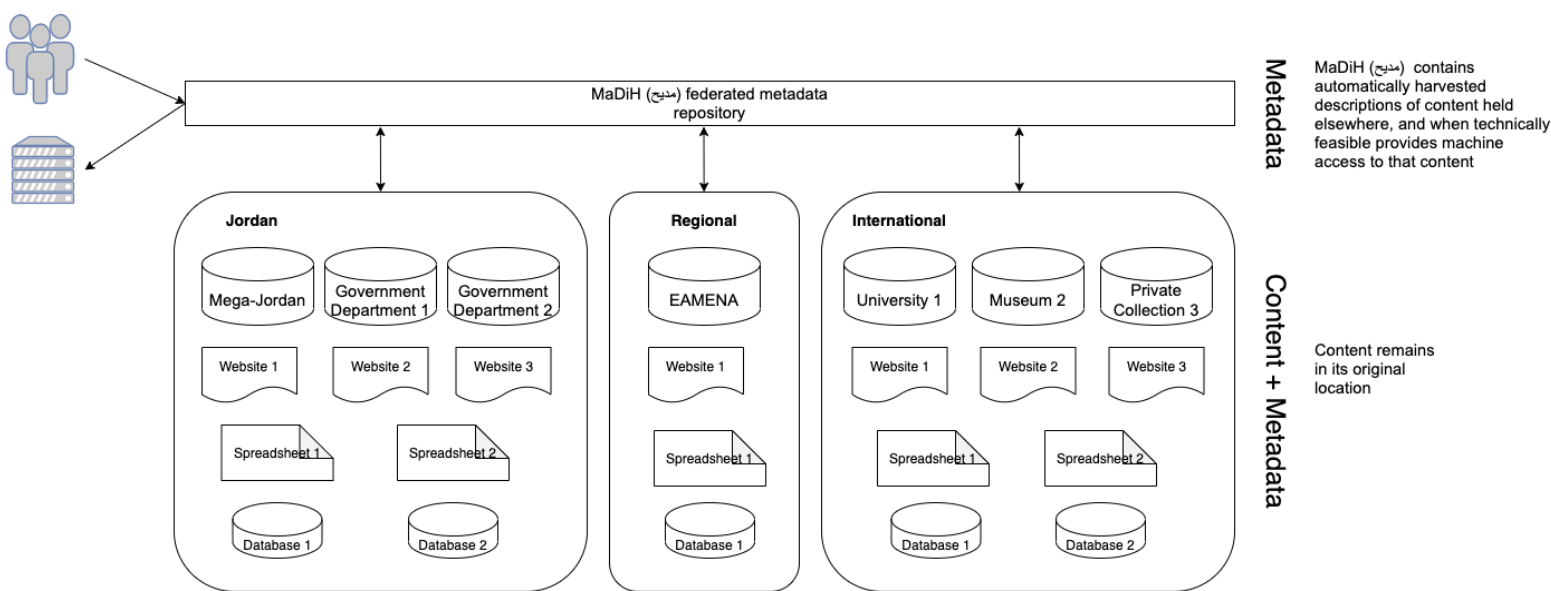


Figure 5. MaDiH (مدىح) CKAN catalogue turned into a federated metadata repository: Technical Architecture (possible future state)

⁷³ Paul Millar, James Smithies, Chris Thomson and Jennifer Middendorf, “UC CEISMIC Canterbury Earthquakes Archive”, accessed February 24, 2021, <http://www.ceismic.org.nz/>, and see James Smithies, Paul Millar, and Christopher Thomson, “Open Principles, Open Data: The Design Principles and Architecture of the UC CEISMIC Canterbury Earthquakes Digital Archive”, *Journal of the Japanese Association for Digital Humanities* 1, no. 1 (2015): 10–36.

⁷⁴ “DigitalNZ”, National Library of New Zealand, accessed February 24, 2021, <https://digitalnz.org/>.



CHALLENGES

- Significant (but not insurmountable) technical barriers, related to data normalisation and mapping, would need to be dealt with to implement federated search functionality.

RECOMMENDATIONS

- Modest future funding should be used to upgrade the existing MaDiH (مديح) data catalogue with the requirements listed in Appendix C.
- Major future funding should use MaDiH (مديح) as a model, but undertake significant preliminary engineering and design processes before embarking on major enhancements.

Dataset characteristics

In total 325 datasets on Jordanian cultural heritage were recorded in the MaDiH (مديح) CKAN catalogue.⁷⁵ The key results are discussed in the MaDiH (مديح) Policy White Paper and a journal paper, where an in-depth discussion of content and technical aspects of the datasets can be found.⁷⁶ The dataset information can also be accessed by searching and exploring the project CKAN instance. In this white paper, we focus on the technical aspect only.

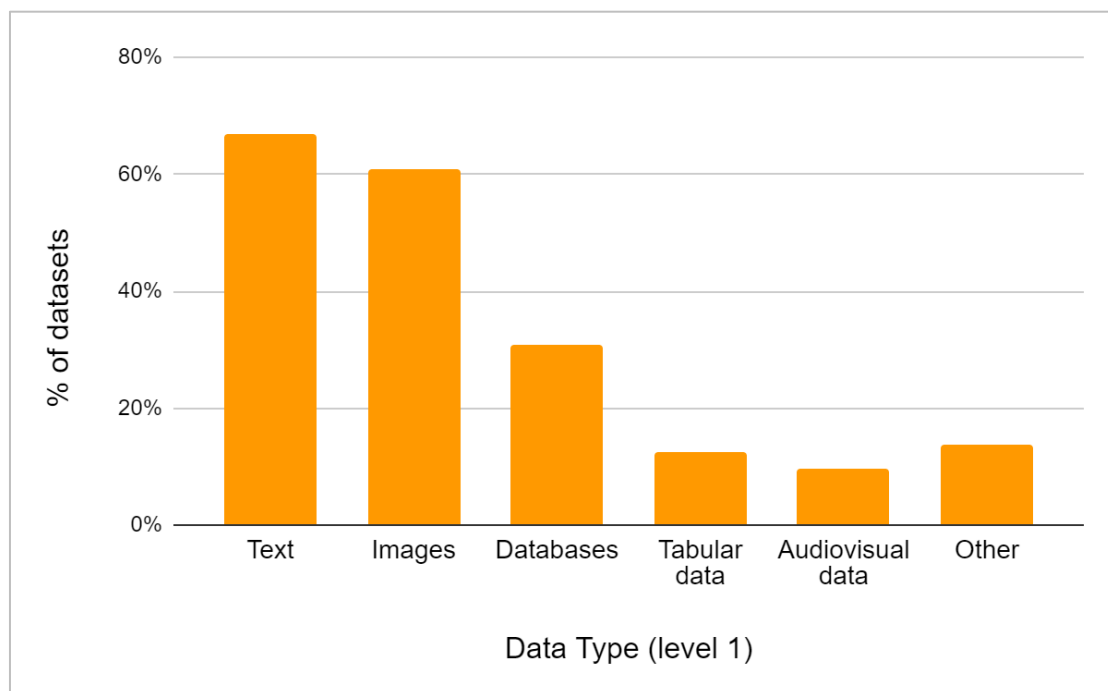
The majority of the datasets recorded were available online (74%): these were prioritised over others partly because they were much easier for us to find and access. The large majority of the datasets were in a digital format (93%), which

⁷⁵ The MaDiH CKAN continues to be updated. For this white paper we use the cut-off point of 1 January 2021.

⁷⁶ Pascal Flohr et al. "Mapping Digital Heritage in Jordan (MaDiH): What are the data telling us?" (in preparation, 2021).

is also partly caused by the number of online (and therefore digital) datasets in the catalogue, but it is likely also a reflection of reality.⁷⁷ Related to this, the majority of the recorded datasets are public (47%) or partially public (30%). English is the most common language (71% of datasets), followed by Arabic (36%),⁷⁸ with datasets that are *only* available in another language a small minority. The majority (65%) of datasets that we recorded are located outside Jordan, while a considerable minority of 35% are curated from within Jordan.

The majority (67%) of datasets are of, or contain, a text data type, followed by images (61%), databases (31%), and tabular data (13%) (Fig. 5).⁷⁹ The tabular data (or 'table') category is similar to the database category in the type of information recorded, but presents it in a single table format. In at least several cases the recorded 'table' is in fact an export from a database, where the database itself is not published or accessible and an Excel table is published instead (i.e. of the 41 datasets with tables, 17 are also classed as containing databases).



⁷⁷ For discussion, see the Policy White Paper and Pascal Flohr et al. "Mapping Digital Heritage in Jordan (MaDiH): What are the data telling us? (in preparation, 2021).

⁷⁸ 11% of datasets are bilingual English and Arabic

⁷⁹ One dataset can contain multiple types, for example both images and text.

Figure 5 Data type (main groupings) for the datasets recorded in the MaDiH CKAN. One dataset can contain multiple data types.



Due to the CKAN data entry template setup, not all datasets had formats recorded: when a dataset was not online and therefore not linked, no format was, normally, recorded. For the online datasets for which a format was recorded (276 in total; Fig. 6), most common are the 'web formats' (54%), which mostly represent information given on web pages and in turn often link to other types of data (note that one dataset can have multiple formats). 'Text formats' (25%) relate to documents such as PDFs, often research articles or reports. Images are very common (26%) and relate partly to large image datasets, but most commonly to images in combination with other formats (e.g. images in a report or on a website). Table or spreadsheet formats are present in 15% of the datasets. It is interesting that database formats are not very common (3%) even though databases form 31% of the data types.

This reflects the way they are published or accessible (and recorded by the project), often accessed through front end interfaces that present data in html format but do not give access to the data sources or backend and, as such, were classed as a 'web format' when deciding on a data format. It was recognised that a database will often underlie this format, but this was left implicit for simplicity's sake.

Moreover, it is often the case that databases are referred to, but they are not actually published themselves, for example as a result of funding limitations (it is free to add data as an Excel table or csv to a paper, but it costs money and time to publish a database online, which subsequently requires maintenance) and sustainability requirements (for example when funding bodies recommended data is submitted in long-term sustainable formats like csv rather than database dumps in formats like SQL).

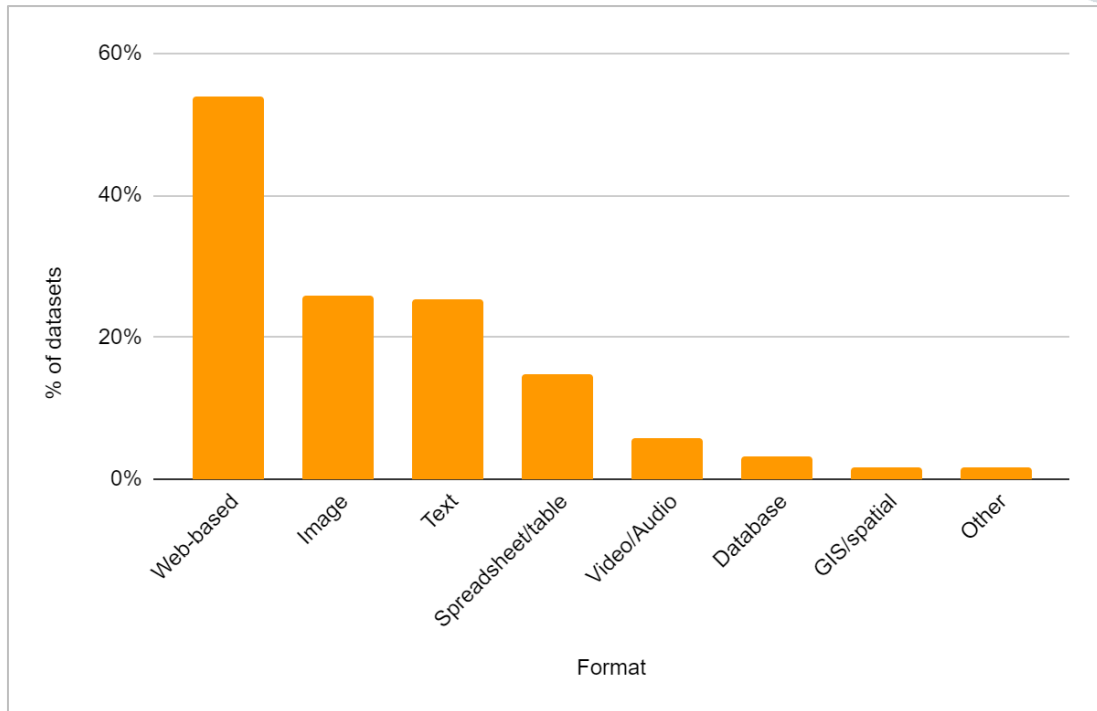


Figure 6 Formats represented in the online datasets recorded in the MaDiH CKAN. One dataset can contain multiple formats (e.g. text and images).

The fact that many of the recorded datasets are ‘summary’ texts (with images) in the form of websites and PDFs, rather than more tractable data sources like databases/tables or photo archives, has implications for collating the data and making it searchable. While a majority of the datasets we recorded were public or partly public, and therefore open or partly open in the form in which they were made accessible, they are often at most ‘1-star’ open data,⁸⁰ or are not even designed to disseminate their underlying data. These datasets still form valuable resources, but as it will be harder to integrate them (except perhaps via a search engine), our recommendation is to focus on datasets from ‘2-star’ level and above as well as to improve key datasets (i.e. data compilations) to reach these levels.

⁸⁰ James M. Kim and Michael Hausenblas. “5-star Open Data”, 2012 (updated 2015), accessed January 18, 2021, <https://5stardata.info/en/>; Tim Berners-Lee, “Linked Data”, 2006, accessed January 18, 2021, <https://www.w3.org/DesignIssues/LinkedData.html>.

Even when datasets are '2-star' or above, challenges remain. These include data quantity: for example in the DOA's archaeological site database MEGA-Jordan 15,000 physical sites are currently recorded,⁸¹ but it is estimated that perhaps as many as 100,000 sites exist in Jordan. Data quality is also an issue: for example archaeological site data is often based on older, pre-GPS era information, resulting in the wrong location in archaeological site databases. In addition, vocabularies are not always comparable between or even within datasets. For example, the term 'site' is used for single walls or findspots, but also for large concentrations of remains or complete landscape features like terrace systems.

The varied nature of the datasets, in terms of data types, data formats, language, and quality, will make a federated catalogue that is easily searchable (and ideally also presented as Linked Open Data) challenging to implement. As in the case of projects such as *Europeana* and *DigitalNZ* this challenge is certainly not unique to Jordan,⁸² and can be addressed to a large degree by choosing only certain datasets to be linked, clever investment of time and resources, support for crowd-sourcing initiatives, and integration into long-term institutional and government planning. Initiatives like PeriodO for cultural periods⁸³, Geonames,⁸⁴ Pleiades,⁸⁵ The Pelagios Network,⁸⁶ and the Heritage Gazetteer of Libya,⁸⁷ for modern and ancient

⁸¹ "MEGA-Jordan", Department of Antiquities, the Getty Conservation Institute and World Monuments Fund, accessed January 19, 2021, <http://megajordan.org/>; S. Hababbeh, Personal Communication 29 October 2019.

⁸² "Europeana Collections", Europeana, accessed January 19, 2021, <https://www.europeana.eu/en/collections>; "DigitalNZ", National Library of New Zealand, accessed February 24, 2021, <https://digitalnz.org/>.

⁸³ Adam Rabinowitz, Shaw, R., Buchanan, S., Golden P., Kansa, E. "Making sense of the ways we make sense of the past". *History and Landscape* 59(2): 42-55.(2016) <https://doi.org/10.1111/j.2041-5370.2016.12037.x>; Adam Rabinowitz, Ryan Shaw and Patrick Golden, "A gazetteer of periods for linking and visualizing data", accessed January 19, 2021, <https://perio.do/>.

⁸⁴ "GeoNames", GeoNames Team, accessed January 19, 2021, <http://www.geonames.org/>,

⁸⁵ "Pleiades", Ancient World Mapping Center, the Stoa Consortium, and the Institute for the Study of the Ancient World, accessed 18/01/2021, <https://pleiades.stoa.org/>.

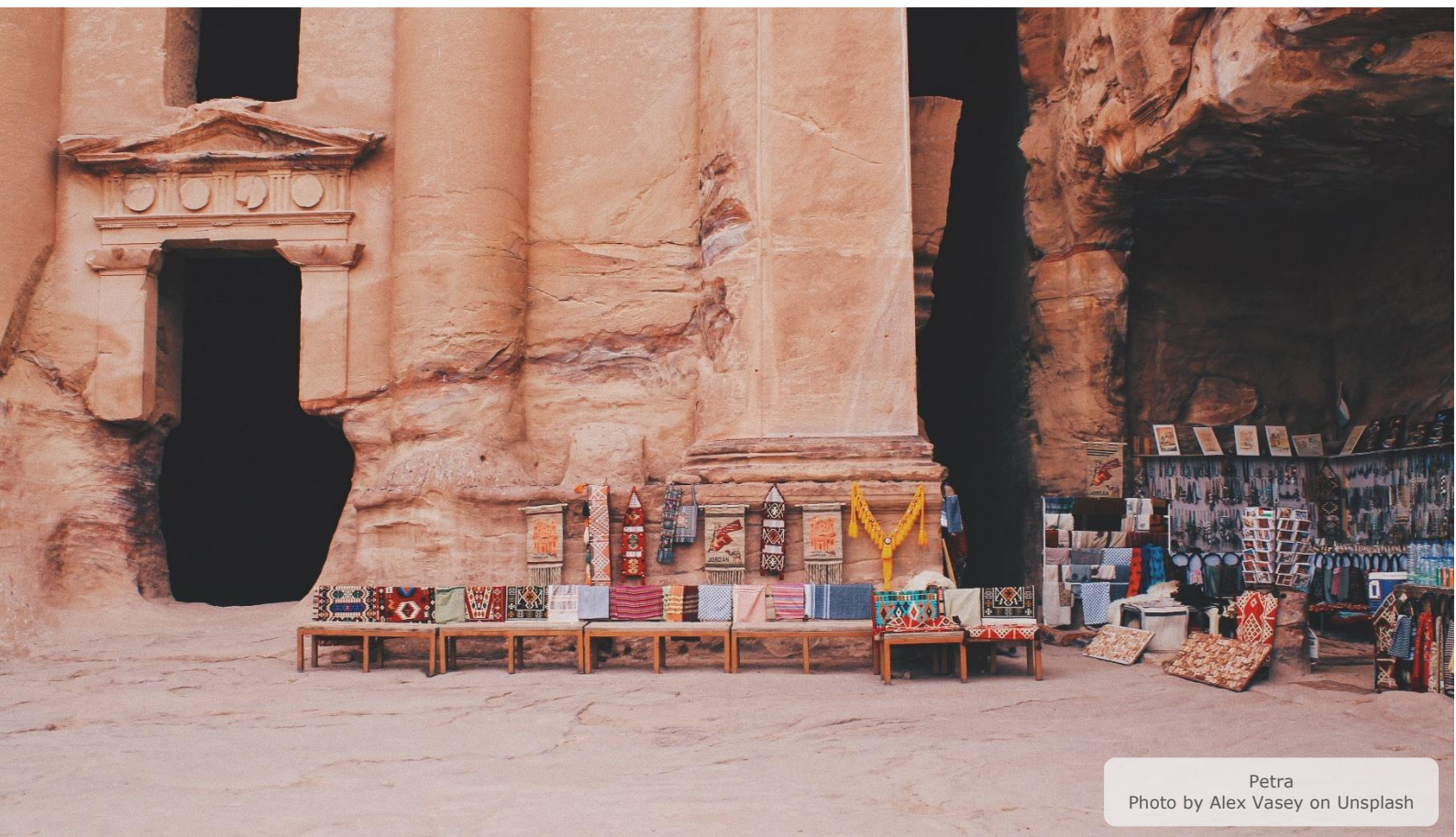
⁸⁶ "The Pelagios Network". The Pelagios Network, accessed 18/01/2021, <https://pelagios.org/>.

⁸⁷ "Heritage Gazetteer of Libya / SLS Archives", The Society for Libyan Studies & King's Digital Lab, accessed March 17, 2021, <https://www.sls gazetteer.org/>.

TECHNICAL WHITE PAPER

MaDiH: Mapping Digital Cultural Heritage in Jordan

(Classical) geographical location would also be useful. In addition, in the paleoclimate sciences it is for example accepted to make available one's results according to specific guidelines of repositories, with specific required fields and only certain allowed formats;⁸⁸ this is also applied in other type/format-specific datasets, e.g. IIF for image interoperability. For archaeological sites the DoA already has such requirements, which correspond to spatial (location) and attribute fields in their MEGA-Jordan database: project teams have to submit their data on 'MEGA-Jordan cards' and until a few years ago the teams could even enter their own data.⁸⁹ However, it will take considerable work input and funding to reach homogeneity for all, or even the most relevant, legacy and current datasets available for Jordanian cultural heritage.



Petra
Photo by Alex Vasey on Unsplash

⁸⁸ E.g, "Contributing Data", NOAA, accessed January 14, 2021, <https://www.ncdc.noaa.gov/data-access/paleoclimatology-data/contributing>.

⁸⁹ Reasons for stopping users from editing data directly were various, including quality control.



CHALLENGES

- The majority of the recorded datasets are of or contain a non-structured or encoded text data type, which will make integration (e.g. through Linked Open Data) complicated.
- Almost a third of the data types recorded were databases, but only a minority were available (to us at the time of recording) in this format.
- While the majority of datasets we recorded were (partly) publicly available, the majority of these were of '1-star' or below open data quality.
- Datasets varied in data types, formats, and quality of data: a challenge for creating one searchable catalogue and Linked Open Data.

RECOMMENDATIONS

- Focus future data quality and integration work on datasets containing structured data, i.e. avoiding more text- and image-only journal papers or websites but include structured datasets (e.g. csv files) and favour aggregated datasets (e.g. archaeological site or object databases).
- Aim to digitise and/or 'structure' key datasets that are not yet available in that format.
- Encourage producers of CH data to follow specific, recognised styles, centralised ontologies, and standardised controlled vocabularies.
- Encourage producers of CH data to use structured formats.
- Aim to increase the level of openness in DCH datasets by adopting Linked Open Data approaches for core entities (minimally locations, e.g. compatibility with LOD gazetteers).
- Obtain resources for curation, cleaning, and maintenance of legacy data in existing databases.
- Resolve data quality and integration issues through long-term planning, aligned to central government policies.



Technical sustainability

It is difficult to accurately assess the sustainability of datasets described in the MaDiH (مدیح) CKAN catalogue (the likelihood they will remain publicly accessible into the future, their technical state, their access to funding etc), partly because of the necessarily high-level assessment that was undertaken (identifying and documenting basic information therefore prioritising breadth of coverage over depth of description), but also because of known issues with data sustainability more generally. Large datasets such as MEGA-Jordan and EAMENA are in some important senses more sustainable than smaller datasets, due to their need of ongoing funding and integration to larger strategic plans, but they are always threatened by the need for constant technical maintenance and have larger budget overheads than smaller initiatives. Medium and small datasets are often less complex and cheaper to maintain, and datasets in the form of simple csv files (especially if held in institutional data repositories) are often the most robust of all. Care needs to be taken, in other words, not to assume that simplistically large-scale initiatives are the answer to sustainability issues. The kind of federated data catalogues and archives prototyped and recommended by the MaDiH (مدیح) project provide a good balance, allowing for robust long-term investment in a handful of major repositories (ensuring there will not be too much competition for ongoing funding) alongside minor funding and community support for many medium and small repositories. This kind of architecture, combined with long-term strategies to analyse and progressively improve and appropriately integrate content, provides the best of both worlds.

As McKeague et al. note, it is also useful to recognise that “[o]bsolescence in both software and storage formats are major risks to the sustainability of digital data”.⁹⁰ Triage of datasets should take this into account, perhaps prioritising datasets that

⁹⁰ Peter McKeague,, Rein van't Veer, Isto Huvila, Anne Moreau, Philip Verhagen, Loup Bernard, Anwen Cooper, Chris Green, and Niels van Manen. "Mapping Our Heritage: Towards a Sustainable Future for Digital Spatial Information and Technologies in European Archaeological Heritage Management." *Journal of Computer Applications in Archaeology* 2, no. 1 (2019): 89-104. <http://doi.org/10.5334/jcaa.23>.

might be lost due to the failure of the systems and software that store them ahead of other data that although valuable is stored using well maintained infrastructure. This speaks to the need for investment in *human* capacity, and permanent technical careers, to ensure this kind of holistic analysis is undertaken and high-quality infrastructure is maintained. Humans are also needed to mitigate 'link rot' or the loss of functioning URLs that occurs naturally over time as resources are removed or moved on the Internet, and 'bit rot' that occurs when digital information degrades over time (such as music on a CD becoming unreadable).⁹¹ Link rot was encountered even over the short time span of the MaDiH (مديح) project, for example a URL to the EAMENA project website (a centrally important data resource) was found to be broken in November 2020 only shortly after it was first entered.

CHALLENGES

- The MaDiH (مديح) project has not been able to accurately assess the sustainability of the datasets described in its CKAN catalogue.
- Permanent staff are needed to keep Jordan's DCH data assets up to date, and attend to issues of system sustainability, link rot, and bit rot.

RECOMMENDATIONS

- Jordan's DCH community should use a triage system, balancing the need for system and data maintenance, to manage its DCH assets.
- Jordan should invest in permanent technical careers, and develop graduate career pathways, to ensure infrastructure is maintained and link and bit rot is minimised.
- Jordan should secure strategic funding to maintain the architecture/ecosystem of RIs, beyond single project funding.

⁹¹ Karol Król and Dariusz Zdonek, 'Peculiarity of the Bit Rot and Link Rot Phenomena', *Global Knowledge, Memory and Communication* 69, no. 1/2 (1 January 2019): 20–37, <https://doi.org/10.1108/GKMC-06-2019-0067>.



Digitisation requirements

Several stakeholders assumed that the MaDiH (مدىح) project would have funds available for digitisation. This was not the case, but significant amounts of analogue content were identified during the course of data identification that would benefit from digitisation.

Different categories were identified at the start of the project: Digital + online; digital + offline; analogue + offline.⁹² Of the 325 identified datasets:

Category	Number of datasets	Percentage
Online + digital	239	74%
Offline + digital	71	19%
Offline + analogue	23	7%

The majority of the documented datasets were online and digital, but this is partly because these are the datasets that were most visible to us during data collection. We have identified a number of important datasets that are not yet digital and/or online, and expect many more of these to exist off the MaDiH (مدىح) 'radar'. In addition, some of the online and digital datasets do not contain all of the data from the project in question (i.e. part of the data is not part of the online and/or digital dataset).

Even taking into account the research bias of finding relatively more online datasets, we surmise that the large percentage of digital datasets partly reflects the fact that almost everything is recorded at least partly in digital form. However, there are older archives (for example archaeological excavation documents from the 20th century) that have not yet been digitised, or of which only summaries exist in a digital form. Refer to Appendix D for an indicative list of future digitisation requirements. Preliminary scoping and technical requirements definition will be required, with reference to digitisation standards such as the IIIF image standard to enhance interoperability, and those published by the U.S.

⁹² These are the datasets that are offline and analogue only, in the numbers reported here we have not reported the datasets that have been partially digitised.

TECHNICAL WHITE PAPER

MaDiH: Mapping Digital Cultural Heritage in Jordan

National Archives and Records Administration (NARA)⁹³ and the UK National Archives.⁹⁴ Associated high quality cataloguing should also be undertaken, following guidelines such as those provided by the UK National Archives.⁹⁵ The 5 Star Open Data approach could inform the process, and facilitate the progressive development of a Linked Open Data ecosystem.⁹⁶ This initiative is extremely useful when developing communities, because it positions datasets and their stewards along a continuum rather than dividing them into 'high-quality' and low-quality' assets.

CHALLENGES

- Significant value could be gained from targeted digitisation of content identified by the MaDiH (مدىح) project, but funding is required for this.

RECOMMENDATIONS

- Future funding bids should include funding for targeted digitisation.
- Cost-benefit analysis, to ensure alignment to national policy goals and identify 'low-hanging fruit', should be used to help select content for digitisation.
- Future digitisation should be conducted in collaboration with well-placed Jordanian organisations and institutions, such as the Royal Hashemite Documentation Centre.
- Digitisation should be undertaken using international technical standards for digitisation and subsequent cataloguing, and initiatives such as 5 Star Open Data and Linked Open Data initiatives and toolsets.

⁹³ Steven Puglia, Jeffrey Reed, and Erin Rhodes. *Technical guidelines for digitizing archival materials for electronic access: Creation of production master files--raster images*. Digital Library Federation, 2005.

<https://www.archives.gov/files/preservation/technical/guidelines.pdf>.

⁹⁴ UK National Archives. *Digitisation at The National Archives*, (UK, 2016), accessed February 24, 2021, <https://nationalarchives.gov.uk/documents/information-management/digitisation-at-the-national-archives.pdf>.

⁹⁵ "Cataloguing archive collections", UK National Archives, accessed February 24, 2021, <https://www.nationalarchives.gov.uk/archives-sector/advice-and-guidance/managing-your-collection/developing-collections/cataloguing-archive-collections/>.

⁹⁶ James M. Kim, and Michael Hausenblas. "5-star Open Data". 2012 (updated 2015), accessed January 18, 2021, <https://5stardata.info/en/>.

Licensing and Intellectual Property

As discussed in the MaDiH (مديح) policy white paper, Jordan is committed to open access and open data.⁹⁷ In light of this, and following international principles,⁹⁸ the *metadata* recorded in the MaDiH (مديح) CKAN catalogue are open access, and available for use under a Creative Commons 4.0 CC-NC-SA license.⁹⁹ However, this has no impact on the availability and licensing of the data in the datasets about which the metadata were recorded, and ownership and intellectual property remains with the dataset owner/creator (unless they state otherwise). No data are reproduced in the MaDiH (مديح) CKAN catalogue. Care was taken to ensure the dataset owner and creator were recorded for each dataset.¹⁰⁰

Due to time constraints and to avoid confusion between the license of the metadata versus the original, recorded datasets, information on the licensing of the datasets was not systematically collected for all datasets during this phase of MaDiH; this is something that urgently needs to be added in a next phase. Nonetheless, it has



Amman, photo by Hisham Zayadnh on Unsplash

⁹⁷ Ministry of Planning and International Cooperation. 2018. *Fourth National Action Plan 2018-2020 under the Open Government Partnership Initiative (OGP)*.

https://www.opengovpartnership.org/wp-content/uploads/2019/01/Jordan_Action-Plan_2018-2020.pdf.

⁹⁸ OECD, *OECD Principles and Guidelines for Access to Research Data from Public Funding*, 2007, <http://www.oecd.org/sti/inno/38500813.pdf>.

⁹⁹ See MaDiH CKAN Terms of Use at <https://madih-data.kdl.kcl.ac.uk/>, accessed January 18, 2021.

¹⁰⁰ We also took reasonable care when adding datasets to our catalogue to not include those infringing copyright or intellectual property rights, but cannot guarantee that the data or material in the linked datasets are not doing so.

TECHNICAL WHITE PAPER

MaDiH: Mapping Digital Cultural Heritage in Jordan

become clear that, while many datasets have Creative Commons licenses or customised Terms of Use, a number of the datasets do not have any licensing, or do not clearly state what the copyright (if any) is - including large, prominent datasets like EAMENA.¹⁰¹ Assigning licenses to datasets is important in order to adhere to the FAIR data principles: in order to indicate which data can be used, in what ways, and if and how the data owner/creator should be acknowledged. For open type licensing of data, the Creative Commons suite of licenses can be used,¹⁰² while for online cultural heritage collections of datasets Rights Statements are also useful to indicate the copyright status of the datasets contained in the catalogue or repository.¹⁰³ A combination could be used, for example in MaDiH our metadata falls under a CC-NC-SA license, while we have used the 'Copyright Not Evaluated' Rights Statement.¹⁰⁴

CHALLENGES

- Licensing of datasets was not systematically recorded in the MaDiH CKAN catalogue, largely due to the difficulty of identifying the license and the fact that a considerable number of datasets lack clear licensing statements, impeding their open use and/or risking inaccurate acknowledgment of creators/owners.

RECOMMENDATIONS

- Encourage appropriate licensing of data, datasets, and metadata using a continuum approach to openness aligned to best practices such as Rights Statements and Creative Commons.

¹⁰¹ "EAMENA database", The EAMENA project, accessed January 19, 2021, <https://database.eamena.org/>.

¹⁰² "Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)", Creative Commons, accessed January 19, 2021, <https://creativecommons.org/licenses/by-nc/4.0/>. These are widely used, for example Wikipedia uses a Creative Commons Attribution-ShareAlike License.

¹⁰³ "Rights Statements: For Cultural Heritage Institutions", International Rights Statements Group, accessed January 15, 2021, <https://rightsstatements.org/en/>. For example used by Europeana.

¹⁰⁴ "Copyright not Evaluated", International Rights Statements Group, accessed January 15, 2021, <https://rightsstatements.org/page/CNE/1.0/?language=en>.



Multilingualism

Multilingualism is a central element of the MaDiH (مدىح) project.¹⁰⁵ Refer to the MaDiH (مدىح) Policy White Paper for data analysis. The issue has technical elements of interest to the project too, however. This is most clearly seen in the lack of Arabic language support in the CKAN data catalogue. This is a very widely used and well supported open source project but, as with many such projects, has a largely English-speaking development and user community. The quality of translations of Arabic in the core product is very low as a result, with even basic terms not translated correctly. MaDiH (مدىح) team members contributed translations to the project and there is clearly scope for more work, building connections between the Jordanian open source development and DCH communities with a leading international open source project.

It is reasonable to assume the issue is more fundamental than simple user-interface changes, however, given the need to design multilingual support into products at a field-level, potentially requiring fundamental architectural updates. This extends to the need for adequate data models, capable of supporting local requirements for representation of cultural knowledge. Ideally, Jordan's future national DCH data catalogue would be fully bi-lingual and conform to local cultural requirements for data access and sovereignty (issues closely related to multilingualism).

MaDiH (مدىح)'s data vocabulary goes some way to enabling that level of high quality digital multilingualism, but is only an indication of the direction that is needed. Ideally, users would be able to select a simple 'EN' or 'AR' button on the MaDiH (مدىح) user interface to toggle between equally rich and linguistically / culturally correct experiences.

¹⁰⁵ James Smithies, Fadi Bala'awi, Pascal Flohr, Shatha Mubaideen, Alessandra Esposito, Sahar Idwan, Carol Palmer, Issa Mahasneh, Shaher Rababeh, "MaDiH (مدىح): Mapping Digital Cultural Heritage in Jordan. Policy White Paper"(2021).



It may well be more beneficial, in the long-run, to contribute Arabic to existing international archival data standards, and initiatives such as ICA's Multilingual Archival Terminology project,¹⁰⁶ with a view to ensuring Jordanian cultural heritage data is captured and described multilingually 'at source'. This would be a major undertaking, perhaps requiring several generations, but would represent an exceptional contribution to global DCH and have the added (and more immediate) benefit of positioning Jordanian DCH at the centre of global efforts to improve multilingual DCH support. In time, Jordan could require all cultural heritage data to be developed multilingually, creating profound opportunities for downstream technical development and also regional archival federation.

CHALLENGES

- The CKAN open source project would benefit from more contributions from Arabic speakers.
- Jordanian DCH professionals need clear technical guidelines to enhance multilingual datasets.

RECOMMENDATIONS

- Encourage Jordanians to help the CKAN community, and other open source communities, improve Arabic language support in the products.
- Contribute Arabic input to international data standards and archival initiatives, and encourage their use in Jordan.

¹⁰⁶ International Council on Archives. "Multilingual Archival Terminology". March, 2012, accessed December 04, 2020. <https://www.ica.org/en/online-resource-centre/multilingual-archival-terminology>.



Career development requirements

As Jordan's technology sector and the educational system feeding it are robust, a good foundation exists for further development of DCH careers. One way to achieve this would be to engage with the international Research Software Engineering (RSE) community,¹⁰⁷ and perhaps begin a national Jordanian initiative.¹⁰⁸ RSE is cross-disciplinary and focused on enabling research in the higher education sector, but (as with King's Digital Lab) is also becoming increasingly common in the arts, humanities, and cultural heritage sector. KDL's RSE training, provided to the MaDiH (مدىح) team, is a good place to start.¹⁰⁹ Refer to the MaDiH (مدىح) Policy White Paper for more information about education and training requirements. The Lab's career profiles provide examples of potential professional roles.¹¹⁰ The Jordan Open Source Association is also very well positioned to foster the development of technical DCH capacity, acting as a contact point between the related sectors and advocating for cultural heritage use cases and standards across Jordan's technology sector. It has been one of the main aims of the MaDiH project to increase contact points between the IT and cultural heritage sectors, but further fostering of development of collaboration is needed. Use could also be made of training and education initiatives across the Digital Humanities community focused on the Arab world, such as the NYU Abu Dhabi Winter Institute in Digital Humanities,¹¹¹ and the University of Michigan Library's resource list for librarians working with Middle East scholarship.¹¹² Ideally, Jordan's DCH community would start to contribute similar initiatives to the global community as well.

¹⁰⁷ "Society of Research Software Engineering", Society of Research Software Engineering, accessed January 19, 2021, <https://society-rse.org/>.

¹⁰⁸ Local RSE initiatives exist in a variety of countries, including the United Kingdom, United States, and Germany.

¹⁰⁹ "King's Digital Lab MaDiH's (مدىح) RSE Training", European DARIAH-Campus, accessed January 19, 2021, <https://campus.dariah.eu/resource/rse2019>.

¹¹⁰ James Smithies, "Research Software (RS) Careers: Generic Learnings from King's Digital Lab, King's College London" (Zenodo, 7 February 2019), <https://doi.org/10.5281/zenodo.2564790>.

¹¹¹ NYU Abu Dhabi Winter Institute in Digital Humanities. <https://wp.nyu.edu/widh/>. Accessed 08 April 2021.

¹¹² University of Michigan Library, Resource list for librarians working with Middle East scholarship. <https://guides.lib.umich.edu/NE/DSDH>. Accessed 08 April 2021.



CHALLENGES

- Non-technical cultural heritage staff need support from colleagues with technical skills as well as some cultural heritage domain knowledge.

RECOMMENDATIONS

- Foster the development of RSE careers and career pathways in collaborations between the government, technology sector, universities, and the cultural heritage sector.

FUTURE REQUIREMENTS (CONCEPTUAL)

Many of the requirements identified by the MaDiH (مديح) project are related to national policy and strategy, as detailed in the MaDiH (مديح) policy white paper, or listed throughout this paper and in Appendix C: Technical Challenges and Opportunities. Refer to Appendix B for detailed technical requirements, related to the MaDiH (مديح) CKAN data catalogue. The goal of the project has been to begin the process of landscape and requirements analysis, rather than defining a specific solution in the form of a technical product(s) or system(s), or even a high-level national architecture. All of those will ideally be produced eventually, but it is important to gain a general overview of the technical 'terrain' first. It has become clear, however, that the MaDiH (مديح) CKAN data catalogue is of particular value to multiple Jordanian communities and should be both sustained and extended. Although originally intended only as a proof of concept output, and (primarily) a tool to collect data to inform policy and strategic analysis, it quickly proved valuable as a national DCH datasets catalogue. For example, in March 2021 the catalogue had over a thousand unique visitors, with visitors from Jordan second after visitors from France. Once migrated to Jordan to ensure data sovereignty, and placed under long-term management, the catalogue would ideally be scaled

TECHNICAL WHITE PAPER

MaDiH: Mapping Digital Cultural Heritage in Jordan

towards a resource such as <https://www.europeana.eu/> or <https://digitalnz.org/>, both aggregating metadata and (where feasible) providing human and programmatic access to full content (i.e. enabling access to the data). In doing so, it could form the basis of a rich Linked Open Data ecosystem for Jordanian cultural heritage data, from both before and after 1750 CE and including tangible as well as intangible heritage content.¹¹³ There is obvious potential for such a resource to scale towards other relevant domains, such as natural heritage and the social sciences, but early phases would presumably focus on archaeological and cultural heritage content.

CHALLENGES

- Maintaining and sustaining the current version of the MaDiH (مدىح) CKAN data catalogue, both in terms of system and content.

RECOMMENDATIONS

- Migrate the MaDiH CKAN catalogue to Jordan to ensure data sovereignty while ensuring continued access to the international and national research and heritage sector communities.
- Scale the MaDiH CKAN catalogue (or equivalent data management platform) to enable access to data instead of only metadata, developing towards a Linked Open Data ecosystem for Jordanian cultural heritage.
- Eventually, scale the MaDiH CKAN catalogue (or equivalent data management platform) to include natural heritage, and potentially other relevant domains such as the social sciences.

¹¹³ All these categories are currently included in the MaDiH catalogue. See the Policy White Paper for more information.

RECOMMENDED DESIGN & ENGINEERING PROCESS



There is no need to describe best practice in software engineering to a Jordanian technical readership given its maturity, reflected in the MaDiH (مديح) project in the Jordan Open Source Association and their national community. The expectation is that any future technical development of the MaDiH (مديح) CKAN data catalogue would use modern Agile software design and engineering practices, with hosting and maintenance according to best practices in systems and security maintenance.

There are wider issues to consider, however, related to the potential larger strategic infrastructure vision the MaDiH (مديح) project touches on in its policy white paper. That context is national, with alignment to international standards and technical infrastructure, requiring careful alignment of strategic and technical requirements. If this path is chosen - if a decision is made to use the MaDiH (مديح) CKAN repository as the key asset in a wider programme of DCH technical and infrastructural development - care should be taken to invest in a robust concept development process, involving primarily local but also international experts in software and systems design and the development of DCH infrastructure. The use of one or more external consultants is recommended, including international consultants. It is important any design and delivery of such an infrastructure is the responsibility of appropriately skilled Jordanian Research Software Engineers but (as in infrastructure development in any country), consideration should be given to international expertise and global infrastructure (such as large-scale cloud infrastructure) where appropriate. Significant attention should be paid to international DCH initiatives and infrastructure, aligning to them or leveraging them directly where appropriate, and significant effort should be made to understand (through user research and requirements elicitation) the nature of the user community and what their needs are.



CHALLENGES

- Ensuring further development, hosting, and maintenance of the MaDiH (مديح) CKAN data catalogue occurs according to software engineering (including analysis and design) and management best practice.
- Ensuring any larger vision for Jordanian DCH is well designed and achievable.
- Ensuring any larger vision for Jordanian DCH supports local needs and communities.
- Ensuring any larger vision for Jordanian DCH aligns to and where appropriate reuses key international initiatives.

RECOMMENDATIONS

- Use expertise across the Jordanian technology sector to ensure best practice in design and engineering in the further development and maintenance of the MaDiH (مديح) CKAN data catalogue.
- Engage external consultants, including international consultants, in the conceptual development of any larger vision for Jordanian DCH and infrastructure.
- Use expertise across the Jordanian technology sector to design, deliver, and maintain any larger Jordanian DCH and infrastructure.
- If necessary, leverage international expertise and (for example cloud) infrastructure to design, deliver, and maintain any larger Jordanian DCH and infrastructure.
- Ensure any larger vision for Jordanian DCH is well aligned to existing international DCH initiatives and infrastructure.

APPENDIX A: MADIH (مديح) CKAN CATALOGUE, ENHANCEMENT RECOMMENDATIONS



CKAN is an open source platform with an active community, and therefore many endorsed or informal extensions are available that might solve the issues below.¹¹⁴

Type	Issue	Recommendation	Priority
User experience	Front end (user interface) not attractive and clunky	Develop high-quality interface	High
Data model/ data entry	Currently only two customised fields, so information added as tags, causing multiple issues (prone to error: e.g. typos, difference in spelling, forgetting to add information)	Add additional fields, one for each type of information, e.g. online/offline; digital/analogue; access; dataset location; dataset language; dataset subject (different levels); dataset coverage (different levels). Plus more to standardize information currently in free-text description field. Potential other new fields: copyright year; the time the photograph (etc) was taken (year/period of data collection). Each with controlled vocabulary.	High
	Currently not possible to have nested values (different levels in one field)	Develop the possibility of having nested values (e.g. for data type; period).	High
	Too many levels nested in one field	As above: more customised fields and better nesting of fields to avoid this. The lay-out and naming should also be improved.	High
	No spatial information nor geocoding recorded (only as text values in tags).	Add possibility to add coordinates and draw points/lines/polygons on map to indicate dataset coverage, where relevant (also adapt search to enable searching by location, see below).	Medium

¹¹⁴ "CKAN Extensions", CKAN, accessed April 7, 2021, <https://extensions.ckan.org/>

TECHNICAL WHITE PAPER

MaDiH: Mapping Digital Cultural Heritage in Jordan



Type	Issue	Recommendation	Priority
Data model / data entry	Greater clarity about (technical) information openness	Add https://5stardata.info/en/ or similar to data set entries.	Low
	Greater clarity about licensing and copyright of datasets. A field exists in CKAN to indicate the licensing of a dataset; however, the project needs to gain more understanding about recommendations for its use.	Add information (e.g. to the field name) to clarify. An overall license for the MaDiH metadata has already been added, but it might be worth adding an additional field (if further investigation demonstrates it isn't currently available) specifying this for each dataset (currently all the same, but potential to differ between datasets in future).	High
	Cultural period, not specific dates	possibility to enter specific start and end dates to place on timeline (but needs to be discussed if this is making things too detailed for our generic catalogue)	Low (needs discussion first)
	Format recorded as resource level, not at dataset level. So only some of the formats are recorded.	Format field at dataset level.	Low
Data model / Vocabulary	Full complexity of datasets not captured. E.g. missing difference between large collections of data (e.g. MEGA-Jordan) and data from single archaeological sites, or difference between single journal paper and journal catalogue.	Re-assess when possibility of additional fields exists and based on feedback from within and outside team. E.g. additional fields for data type. Ensure to keep balance between broad enough categories for meaningful analysis and capturing full complexity.	Medium
Search	Search not user-friendly	Develop user-friendly search interface, including for advanced search (OR, NOT, combinations)	High

TECHNICAL WHITE PAPER

MaDiH: Mapping Digital Cultural Heritage in Jordan



Type	Issue	Recommendation	Priority
	Search of controlled vocabulary (and potentially of everything) is capital sensitive.	All queries should be non-capital sensitive to avoid missing datasets due to (non)capitalization.	High
	Search error does not indicate what is wrong.	Search error warning messages to indicate what the issue is.	Medium
	Currently no nested value search supported.	Enable more advanced faceted search. I.e. when searching for "Roman" datasets, all the children of "Roman" such as "Roman Early" and "Roman Late" should be included in the results.	High
Data entry	Naming of "source" field confusing	Rename to "source of dataset" or something along those lines; also move the field (see point above).	Medium
	No controls on field length	Add possibility to add such controls, to allow for greater consistency (e.g. description field).	Low
	No controls on date format	Add controlled date format (e.g. always yyyy-mm-dd).	Medium
	Order always alphabetical in drop-down menus (or set from start?).	Allow for different order in drop-down menus, e.g. chronological for Cultural Period.	High
Language	Many characters (non-standard in English, or special symbols) currently not supported, including characters needed for Arabic transliteration.	Make it possible to use non-standard characters, e.g. diacritics.	High
	Bilingual functionality not implemented, not customised.	Develop bilingual interface, i.e. user presses 'EN' or 'AR' button at top of screen to switch. So not each field showing both languages all the time.	High

TECHNICAL WHITE PAPER

MaDiH: Mapping Digital Cultural Heritage in Jordan



Type	Issue	Recommendation	Priority
	Core translation of Arabic in CKAN but quality very low.	High-quality translation for all fields and controlled vocabulary terms; where possible add translations for free-text. Contribute to CKAN language mapping.	High
User Interface/ User Experience	Some fields would be clearer when in a different order/position.	Design clearest lay-out and move the relevant fields.	Medium
	Filter: difficult to find term for fields with lengthy terms due to cut off and order by number	Enable alphabetic (or other sorting); improve naming of fields; allow to make left (filter) column wider.	Medium
	Not possible to keep string of text in the same line (e.g. formatting of title in English and Arabic).	Allow displaying of text over two lines.	Not relevant if bilingual interface is developed
	Lists of collaborators take up most of the space, taking attention away from other important fields	Change lay-out, ideally show only one or two with the option to expand to see all. Or if too difficult, minimise and expand on demand.	Low/Medium
Other	Export not possible or not by user	Use CKAN default API to develop user-friendly export interface. To csv, if spatial data also added then also to shp.	Medium/Low

APPENDIX B: DIGITISATION SUGGESTIONS



The following suggestions imply the need for a larger digitization strategy, involving user engagement, assessment, and technical feasibility.

Name	Type	Purpose	File format	Comments	In MaDiH
Analogue reports/ Paper archive	Documents	Documentation, Conservation	pdf, docx, doc	DoA, etc.	Yes
Photographic archive	Photographs	Documentation, Conservation, Promotion	jpeg, tiff, jpg, pdf, png	DoA, Private collections, international institutions, etc.	Yes
Maps archive	Maps	Documentation, Conservation	tiff, jpeg	RJGC, DoA, etc.	Yes
Newspapers, letters and books	Documents	Documentation, Conservation	tiff, jpeg, pdf,	Private collections	Yes
Museum databases	Objects/ Documents	Documentation, Conservation	DBASE, MS Access, Oracle, MySQL, etc.	DoA museums, Royal Tank Museum,	Yes
Aerial Photographs	Photographs	Documentation, Conservation	tiff, jpeg, pdf,	RJGC	Yes
Testimonies and stories	Audiovisual recording	Documentation, Conservation, Promotion	WAVE, MP3, MP4, Flash, Opus, MIDI, AVI, WMA, SWF etc.	Royal Tank museum	Yes
Video archive	Audiovisual recording: video	Documentation, Conservation, Promotion	mp4	Royal Tank museum	Yes
Architecture/Heritage	Buildings	Documentation, Conservation, Promotion	TBD	GAM	Yes
Fine arts	Objects/ Paintings	Documentation	tiff, jpeg, pdf	The Ministry of Culture	No

APPENDIX C: TECHNICAL CHALLENGES



1. Improve awareness of and adherence to international metadata standards.
2. Improve national information management standards, in support of government open data policies.
3. Fund gaps in the technical landscape identified through the MaDiH (مدىح) CKAN catalogue.
4. Fund training and study opportunities for the next generation of DCH professionals.
5. Involve Jordanian stakeholders in the early design of datasets and user interfaces.
6. Improve awareness of the technical requirements and value of open access strategies.
7. Maintain project continuity after funding for Phase 1 ends.
8. Maintain the CKAN catalogue after funding ends.
9. Ensure Jordanian data sovereignty is enabled.
10. The project is well positioned to scale up but requires additional funding to do so.
11. The project could be scaled to support other sectors.
12. The project could be used as a model for similar repositories supporting other sectors.
13. The majority of the recorded datasets are of or contain a non-structured or -encoded text data type, which will make integration (e.g. through Linked Open Data) complicated.
14. Almost a third of the data types recorded were databases, but only a minority was available (to us) in this format.
15. While the majority of datasets we recorded were (partly) publicly available, the majority of these were of '1-star' or below open data quality.
16. Datasets varied in data types, formats, and quality of data: a challenge for

TECHNICAL WHITE PAPER

MaDiH: Mapping Digital Cultural Heritage in Jordan

creating one searchable catalogue and Linked Open Data.

17. The MaDiH (مديح) project has not been able to accurately assess the sustainability of the datasets described in its CKAN catalogue.
18. Permanent staff are needed to keep Jordan's DCH data assets up to date, and attend to issues of system sustainability, link rot, and bit rot.
19. Significant value could be gained from targeted digitisation of content identified by the MaDiH (مديح) project, but funding is required for this.
20. Licensing of datasets was not recorded in the MaDiH (مديح) CKAN catalogue, largely due to the difficulty of identifying the license and a considerable number of datasets lack clear licensing, impeding their open use and/or risking improper acknowledgment of creators/owners.
21. The CKAN open-source project would benefit from more contributions from Arabic speakers.
22. Jordanian DCH professionals need clear technical guidelines to enhance multilingual datasets.
23. Non-technical cultural heritage staff need support from colleagues with technical skills as well as some cultural heritage domain knowledge.
24. Maintain and sustain the current version of the MaDiH (مديح) CKAN data catalogue.
25. Ensuring further development, hosting, and maintenance of the MaDiH (مديح) CKAN data catalogue occurs according to software engineering and management best practice.
26. Ensuring any larger vision for Jordanian DCH is well designed and achievable.
27. Ensuring any larger vision for Jordanian DCH supports local needs and communities.
28. Ensuring any larger vision for Jordanian DCH aligns to and where appropriate reuses key international initiatives.



APPENDIX D: TECHNICAL RECOMMENDATIONS



1. Align to RI and DCH infrastructure in the UK and Europe.
2. Use the MaDiH (مديح) CKAN catalogue as the foundation of a national DCH initiative capable of enhancing government, higher education, and private sector activities.
3. Leverage Jordan's capabilities in IT to position itself as a world leader in DCH.
4. Attract a new generation of professionals to DCH, through study, training, and employment opportunities.
5. Enhance Jordanian data sovereignty and long-term data sustainability through innovative partnerships.
6. Build on the awareness gained by the MaDiH (مديح) project to enhance open access initiatives and standardise national DCH datasets.
7. Transition a proof of concept DCH catalogue into a long-term national asset hosted in Jordan.
8. Bolster Jordanian data sovereignty by demonstrating best practices in migrating the MaDiH (مديح) data catalogue to Jordan.
9. Use the processes and vocabulary developed by the MaDiH (مديح) project for future data identification and collection activities in Jordan.
10. Contribute to the ckan.org open source project, to improve Arabic language support.
11. If scaled up, the project should remain focused on the cultural heritage sector.
12. Modest future funding should be used to upgrade the existing MaDiH (مديح) data catalogue with the requirements listed in Appendix C.
13. Major future funding should use MaDiH (مديح) as a model, but undertake significant preliminary engineering and design processes before embarking on major enhancements.
14. Focus future data quality and integration work on datasets containing

TECHNICAL WHITE PAPER

MaDiH: Mapping Digital Cultural Heritage in Jordan


structured data, i.e. avoiding more text- and image-only journal papers or websites but include compilations.



15. Aim to digitise and/or 'structure' key datasets that are not yet available in that format.
16. Encourage producers of CH data to follow specific, recognised styles, centralised ontologies, standardised controlled vocabularies (ideally compatible with Linked Open Data gazetteers), and structured, open formats.
17. Obtain resources for 'cleaning' legacy data in existing databases.
18. Resolve data quality and integration issues through long-term planning, aligned to central government policies.
19. Jordan's DCH community should use a triage system, balancing the need for system and data maintenance, to manage its DCH assets.
20. Jordan should invest in permanent technical careers, and develop graduate career pathways, to ensure infrastructure is maintained and link and bit rot is minimised.
21. Future funding bids should include funding for targeted digitisation.
22. Cost-benefit analysis, to ensure alignment to national policy goals and identify 'low-hanging fruit', should be used to help select content for digitisation.
23. Future digitisation should be conducted in collaboration with well-placed Jordanian organisations and institutions, such as the Royal Hashemite Documentation Centre.
24. Digitisation should be undertaken using international technical standards for digitisation and subsequent cataloguing, and initiatives such as 5 Star Open Data.
25. Encourage appropriate licensing of data, datasets, and metadata.
26. Encourage Jordanians to help the CKAN community improve Arabic language support in the product.
27. Contribute Arabic input to international data standards and archival initiatives, and encourage their use in Jordan.

TECHNICAL WHITE PAPER

MaDiH: Mapping Digital Cultural Heritage in Jordan

28. Foster the development of RSE careers and career pathways in collaborations between the government, technology sector, universities, and the cultural heritage sector. 
29. Migrate the MaDiH CKAN catalogue to Jordan to ensure data sovereignty while ensuring continued access to the international and national research and heritage sector communities.
30. Scale the MaDiH CKAN catalogue to enable access to data instead of only metadata, developing towards a Linked Open Data ecosystem for Jordanian cultural heritage.
31. Eventually, scale the MaDiH CKAN catalogue to include natural heritage, and potentially other relevant domains such as the social sciences.
32. Use expertise across the Jordanian technology sector to ensure best practice in design and engineering in the further development and maintenance of the MaDiH (مدیح) CKAN data catalogue.
33. Engage external consultants, including international consultants, in the conceptual development of any larger vision for Jordanian DCH and infrastructure.
34. Use expertise across the Jordanian technology sector to design, deliver, and maintain any larger Jordanian DCH and infrastructure.
35. If necessary, leverage international expertise and (for example cloud) infrastructure to design, deliver, and maintain any larger Jordanian DCH and infrastructure.
36. Ensure any larger vision for Jordanian DCH is well aligned to existing international DCH initiatives and infrastructure.

APPENDIX E: BIBLIOGRAPHY



- Ababsa, Mariam, Jordan Documentary Heritage: Part 1, UNESCO Amman Office (2014),
http://www.unesco.org/new/fileadmin/MULTIMEDIA/FIELD/Amman/pdf/Jordan_Documentary_Heritage.pdf
- AHRC, "Newton-Khalidi Cultural Heritage and Sustainable Development in Jordan", accessed January 20, 2021, <https://ahrc.ukri.org/funding/apply-for-funding/archived-opportunities/cultural-heritage-and-sustainable-development-in-jordan/>.
- Ancient World Mapping Center, the Stoa Consortium, and the Institute for the Study of the Ancient World, "Pleiades", accessed 18/01/2021, <https://pleiades.stoa.org/>.
- Berners-Lee, Tim, "Linked Data", 2006, accessed January 18, 2021, <https://www.w3.org/DesignIssues/LinkedData.html>.
- British Council, "Training in Endangered Archaeology", accessed February 3, 2021, <https://www.britishcouncil.org/arts/culture-development/cultural-protection-fund/projects/endangered-archaeology>.
- Ciula, Arianna, "Exposing Legacy Datasets" (2020), accessed 17/03/ 2021, <https://kdl.kcl.ac.uk/blog/legacy-project-datasets/>.
- Creative Commons, "Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)", accessed January 19, 2021, <https://creativecommons.org/licenses/by-nc/4.0/>.
- DataCite Metadata Working Group. "DataCite Metadata Schema Documentation for the Publication and Citation of Research Data". Version 4.0. DataCite e.V. (2016) <http://doi.org/10.5438/0012>.
- DCMI, "Dublin Core™ Metadata Initiative", accessed February 24, 2021, <https://dublincore.org/>.
- Department of Antiquities, the Getty Conservation Institute and World Monuments Fund, "MEGA-Jordan", accessed January 19, 2021, <http://megajordan.org/>.

TECHNICAL WHITE PAPER

MaDiH: Mapping Digital Cultural Heritage in Jordan



- Department of Antiquities, "Museums", accessed February 24, 2021, <http://www.doa.gov.jo/All-Museums.aspx>.
- EAMENA project, "The Cultural Protection Fund", accessed February 3, 2021, <https://eamena.org/cultural-protection-fund>.
- EAMENA project, "EAMENA", accessed January 19, 2021, <https://eamena.org/>.
- EAMENA project, "EAMENA database", accessed January 19, 2021, <https://database.eamena.org/>.
- Esposito, Alessandra, et al. "MaDiH (مدىح) Mapping the Digital Cultural Heritage in Jordan Project. Datasets Identification and Publication Protocol", Zenodo (2020), p. 4. 10.5281/zenodo.4146756.
- European Commission, "European Strategy Forum on Research Infrastructures (ESFRI)", accessed January 19, 2021, <https://www.esfri.eu/>.
- European DARIAH-Campus, "King's Digital Lab MaDiH's (مدىح) RSE Training", accessed January 19, 2021, <https://campus.dariah.eu/resource/rse2019>.
- Europeana, "Europeana Collections", accessed January 19, 2021, <https://www.europeana.eu/en/collections>.
- Flohr Pascal, et al., "Mapping Digital Heritage in Jordan (MaDiH): What are the data telling us?" in preparation (2021).
- Forum on Information Standards in Heritage (FISH), "FISH Vocabularies", accessed February 24, 2021, <http://www.heritage-standards.org.uk/fish-vocabularies/>.
- GAM, "Greater Amman Municipality", accessed November 20, 2021, <https://www.ammancity.gov.jo/ar/main/index.asp>.
- GeoNames Team, "GeoNames", accessed January 19, 2021, <http://www.geonames.org/>,
- Getty Conservation Institute and World Monuments Fund, "Arches", accessed January 19, 2021, <https://www.archesproject.org/>.
- Getty Conservation Institute, "Categories for the Description of Works of Art", accessed February 24, 2021, http://www.getty.edu/research/publications/electronic_publications/cdwa/1object.htm

TECHNICAL WHITE PAPER

MaDiH: Mapping Digital Cultural Heritage in Jordan



- GPIA Amman, "Documentation of the Objects in Jordanian Archaeological Museums (DOJAM)", Amman, accessed January 19, 2021, <https://www.zitadelle-amman.de/projekt/>.
- Gray, Jonathan, *What Do Data Portals Do? Tracing the Politics of Public Information Infrastructures on the Web*, UK: Data Publics, Lancaster University, 2017.
- International Council on Archives, *General International Standard Archival Description ICA*, 1999, https://www.ica.org/sites/default/files/CBPS_2000_Guidelines_ISAD%28G%29_Second-edition_EN.pdf.
- International Council on Archives. "Multilingual Archival Terminology". March, 2012, accessed December 04, 2020. <https://www.ica.org/en/online-resource-centre/multilingual-archival-terminology>.
- International Council of Museums & International Committee for Documentation (ICOM CIDOC), "Conceptual Reference Model", accessed February 24, 2021, <http://www.cidoc-crm.org/>.
- International Rights Statements Group, "Copyright not Evaluated", accessed January 15, 2021, <https://rightsstatements.org/page/CNE/1.0/?language=en>.
- International Rights Statements Group, "Rights Statements: For Cultural Heritage Institutions", accessed January 15, 2021, <https://rightsstatements.org/en/>. For example used by Europeana.
- Kim, James M., and Michael Hausenblas. "5-star Open Data". 2012 (updated 2015), accessed January 18, 2021, <https://5stardata.info/en/>.
- King's College London, "Statement on Use of Personal Data in Research.", accessed 23 November 2020, <https://www.kcl.ac.uk/research/support/rgei/research-ethics/kings-college-london-statement-on-use-of-personal-data-in-research#>.
- Król, Karol, and Dariusz Zdonek, 'Peculiarity of the Bit Rot and Link Rot Phenomena', *Global Knowledge, Memory and Communication* 69, no. 1/2 (1 January 2019): 20–37, <https://doi.org/10.1108/GKMC-06-2019-0067>.
- MaDiH project, "The National Project for the Documentation of Intangible Heritage", accessed February 3, 2021, <https://madih->

TECHNICAL WHITE PAPER

MaDiH: Mapping Digital Cultural Heritage in Jordan



data.kdl.kcl.ac.uk/dataset/the-national-project-for-the-documentation-of-intangible-heritage.

MaDiH project, "Amman Architectural Heritage Record", accessed February 3, 2021, <https://madih-data.kdl.kcl.ac.uk/dataset/amman-heritage-houses-record>.

MaDiH Team, "The MaDiH Vocabulary", accessed February 24, 2021, <https://doi.org/10.6084/m9.figshare.12950879.v1>.

MaDiH Team, "The MaDiH data collection form", Zenodo (2020), <https://zenodo.org/record/4289919#.X74g4mgzY2w>.

McKeague, Peter, Rein van't Veer, Isto Huvila, Anne Moreau, Philip Verhagen, Loup Bernard, Anwen Cooper, Chris Green, and Niels van Manen. "Mapping Our Heritage: Towards a Sustainable Future for Digital Spatial Information and Technologies in European Archaeological Heritage Management." *Journal of Computer Applications in Archaeology* 2, no. 1 (2019): 89-104. <http://doi.org/10.5334/jcaa.23>.

Millar, Paul, James Smithies, Chris Thomson and Jennifer Middendorf, "UC CEISMIC Canterbury Earthquakes Archive", accessed February 24, 2021, <http://www.ceismic.org.nz/>.

Ministry of Culture, Heritage Directorate, "Intangible heritage of Jordan publications", accessed January 19, 2021, <http://ich.gov.jo/node/52332>.

Ministry of Culture, "Museums", accessed February 24, 2021, <https://culture.gov.jo/museums>.

Ministry of Planning and International Cooperation, *The Fourth National Action Plan 2018 - 2020 under the Open Government Partnership Initiative (OGP)*. 2018: 27, https://www.opengovpartnership.org/wp-content/uploads/2019/01/Jordan_Action-Plan_2018-2020.pdf.

Myers, David, and Alison Dalgity, "The Middle Eastern Geodatabase for Antiquities (MEGA): an open source GIS-based heritage site inventory and management system." *Change Over Time* 2, no. 1, (2012).

National Library of New Zealand, "DigitalNZ", accessed February 24, 2021, <https://digitalnz.org/>.

TECHNICAL WHITE PAPER

MaDiH: Mapping Digital Cultural Heritage in Jordan



- NOAA, "Contributing Data", accessed January 14, 2021, <https://www.ncdc.noaa.gov/data-access/paleoclimatology-data/contributing>.
- OECD, *Principles and Guidelines for Access to Research Data from Public Funding*, OECD, 2007, <http://www.oecd.org/sti/inno/38500813.pdf>.
- Palumbo, Gaetano, "JADIS: the Jordan Antiquities Database and Information System; A Summary of the Data." (1994).
- Puglia, Steven, Jeffrey Reed, and Erin Rhodes. *Technical guidelines for digitizing archival materials for electronic access: Creation of production master files--raster images*. Digital Library Federation, 2005. <https://www.archives.gov/files/preservation/technical/guidelines.pdf>.
- Pelagios Network. "The Pelagios Network", accessed January 18, 2021, <https://pelagios.org/>.
- Rabinowitz A., Shaw, R., Buchanan, S., Golden P., Kansa, E. "Making sense of the ways we make sense of the past". *History and Landscape* 59(2): 42-55. (2016) <https://doi.org/10.1111/j.2041-5370.2016.12037.x>
- Rabinowitz A., Ryan Shaw and Patrick Golden, "A gazetteer of periods for linking and visualizing data", accessed January 19, 2021, <https://perio.do/>.
- Rankin R., M. Gordon, and R. Potter, "Implementing a Federated Data Archive with Asynchronous Data Query, Gathering and Analysis Capabilities", *AGU Fall Meeting Abstracts* 33 (1 December 2009): SM33B-1570.
- Royal Hashemite Documentation Center, "RHDC home page", accessed January 19, 2021, <https://www.rhdc.jo/>.
- Rücknagel, Jessika, Paul Vierkant, Robert Ulrich, Gabriele Kloska, Edeltraud Schnepf, David Fichtmüller, Evelyn Reuter et al. "Metadata schema for the description of research data repositories." (2015) https://gfzpublic.gfz-potsdam.de/rest/items/item_1397899_6/component/file_1398549/content.
- Smithies, James, Paul Millar, and Christopher Thomson, "Open Principles, Open Data: The Design Principles and Architecture of the UC CEISMIC Canterbury Earthquakes Digital Archive", *Journal of the Japanese Association for Digital Humanities* 1, no. 1 (2015): 10–36.

TECHNICAL WHITE PAPER

MaDiH: Mapping Digital Cultural Heritage in Jordan



- Smithies, James, "Research Software (RS) Careers: Generic Learnings from King's Digital Lab, King's College London". Zenodo, February 7, 2019, <https://doi.org/10.5281/zenodo.2564790>.
- Smithies, James, Fadi Bala'awi, Pascal Flohr, Shatha Mubaideen, Alessandra Esposito, Sahar Idwan, Carol Palmer, Issa Mahasneh, Shaher Rababeh, "MaDiH (مديح): Mapping Digital Cultural Heritage in Jordan. Policy White Paper"(2021).
- Society of Research Software Engineering, "Society of Research Software Engineering", accessed January 19, 2021, <https://society-rse.org/about/>.
- Tanasi, Davide. "The digital (within) archaeology. Analysis of a phenomenon." *The Historian* 82, no. 1 (2020): 22-36. <https://doi.org/10.1080/00182370.2020.1723968>.
- Tudhope, Douglas et al., "Connecting Archaeological Data and Grey Literature via Semantic Cross Search", *Internet Archaeology*, no. 30 (2011), <https://doi.org/10.11141/ia.30.5>.
- Waibel, Günter, Ralph LeVan, and Bruce Washburn, "Museum Data Exchange: Learning How to Share", *D-Lib Magazine* 16, no. 3/4 (March 2010), <https://doi.org/10.1045/march2010-waibel>.
- Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg et al. "The FAIR Guiding Principles for scientific data management and stewardship." *Scientific data* 3, no. 1 (2016): 1-9. <https://doi.org/10.1038/sdata.2016.18>.
- UK National Archives, "Cataloguing archive collections", accessed February 24, 2021, <https://www.nationalarchives.gov.uk/archives-sector/advice-and-guidance/managing-your-collection/developing-collections/cataloguing-archive-collections/>.
- UK National Archives, *Digitisation at The National Archives*, (UK, 2016), accessed February 24, 2021, <https://nationalarchives.gov.uk/documents/information-management/digitisation-at-the-national-archives.pdf>.
- UKRI, *The UK's research and innovation infrastructure: opportunities to grow our capability*, UKRI, 2020, <https://www.ukri.org/files/infrastructure/the-uks-research-and-innovation-infrastructure-opportunities-to-grow-our-capacity-final-low-res/>.

APPENDIX F: FUNDING



The MaDiH (مديح): Mapping Digital Cultural Heritage in Jordan project was funded by the Cultural Heritage and Sustainable Development in Jordan AHRC Newton / Khalidi call. Ref. AH/S011722/1.

APPENDIX G: ACKNOWLEDGEMENTS



The *MaDiH* (مديح): *Mapping Digital Cultural Heritage in Jordan* project is dedicated to Dr Andrea De Silva Zerbini (07.07.1984 - 12.07.2019), our inspiration and driving force. The project team would like to thank all of our partners, and in particular the Jordanian cultural heritage community who have offered essential support, encouragement, and advice over the course of the project. Particular thanks go to HRH Dana Firas and Steering Group members. Arianna Ciula provided important contributions to the project's data collection and management process and white papers, and delivered RSE training with Neil Jakeman (King's Digital Lab), Adrian Evans and Tom Sparrow (University of Bradford), and Dominic Oldman and Diana Tanase from the British Museum's ResearchSpace project. King's Digital Lab implemented and hosted the CKAN data catalogue.

The Steering Committee members:

- HRH Princess Dana Firas, President of ICOMOS Jordan and Petra National Trust.
- HE Nayef Al Fayez, Minister of Tourism and Antiquities.
- HE Ahmad Hanandeh, Minister of Digital Economy and Entrepreneurship.
- HE Prof Fawwaz M. Al-Abed Al-Haq, President of the Hashemite University.
- HE Ahmad Al-Shami, Acting Director General of the Department of Antiquities.
- Joel Bubbers, Director Jordan & Levant Cluster Lead at British Council.
- Prof Graeme Earl, Vice Dean External Relations, Faculty of Arts & Humanities, King's College London.
- Dr Robert Bewley, Chairman of the Council for British Research in the Levant Board.
- Prof Bill Finlayson, EAMENA project director, School of Archaeology, University of Oxford, UK.
- Dr Will Wootton, Senior Lecturer in Roman Art, King's College London.

The former Steering Committee members:

- HE Majd Shwiekeh, former Minister of Tourism and Antiquities.
- HE Mothana Gharibeh, former Minister of Digital Economy and Entrepreneurship.
- HE Prof Kamal Al Din Bani Hani, former President of the Hashemite University.
- HE Yazid Elayan, former Director General of the Department of Antiquities.