A Lockpick's Guide to dataARC: Designing Infrastructures and Building Communities to Enable Transdisciplinary Research

Rachel Opitz, Colleen Strawhacker, Philip Buckland, Jackson Cothren, Tom Dawson, Andrew Dugmore, George Hambrecht, Willem Koster, Emily Lethbridge, Ingrid Mainland, Tom McGovern, Anthony Newton, Gisli Palsson, Tom Ryan, Richard Streeter, Elisabeth Stade, Vicki Szabo, Polly Thompson

The North Atlantic Biocultural Organization (NABO) community initiated dataARC to develop digital research infrastructures to support their work on long-term human-ecodynamics in the North Atlantic. These infrastructures were designed to address the challenges of sharing research data, the connections between those data and high-level interpretations, and the interpretations themselves. In parallel, they were also designed to support the reuse of diverse data that underpin transdisciplinary synthesis research and to contextualise materials disseminated widely to the public more firmly in their evidence base. This article outlines the research infrastructure produced by the project and reflects on its design and development. We outline the core motivations for dataARC's work and introduce the tools, platforms and (meta)data products developed. We then undertake a critical review of the project's workflow. This review focuses on our understanding of the needs of stakeholder groups, the principles that guided the design of the infrastructure, and the extent to which these principles are successfully promoted in the current implementation. Drawing on this assessment, we consider how the infrastructure, in whole or in part, might be reused by other transdisciplinary research communities. Finally, we highlight key socio-technical gaps that may emerge as structural barriers to transdisciplinary, engaged, and open research if left unaddressed.

Code: https://github.com/castuofa/dataarc-source

Ecosystem explorer and documentation https://github.com/ropitz/dataarc-demo

1. dataARC's Aim: supporting transdisciplinary, open, engaged research

Transdisciplinary collaborations are integral to research addressing current challenges in sustainability, long-term human environment interactions, and landscape change (Becker *et al.* 1997; Theis and Tomkin 2015; Game *et al.* 2018). The impacts of these challenges are strongly felt in the North Atlantic (Jorgenson *et al.* 2019). The North Atlantic Biocultural Organization (NABO), working in the North Atlantic region, brings together a research community focused on long-term human-ecodynamics (L-T H-E) with a strong archaeological focus (Hambrecht *et al.* 2019; Hicks 2019). This research community's ambitions and practice have been visibly shaped by the potential relevance of their data and research outcomes to contemporary debates on the impacts of the climate crisis (Rockman and Hritz 2020), management of present-day ecosystems (Catlin 2016; Hambrecht *et al.* 2020), and the value ascribed to landscapes and communities (McGovern *et al.* 2019; Sigurðardóttir *et al.* 2019). NABO's work has also been influenced by the wider research community's increasing emphasis on public benefits of research (Belford 2020; Aitchison 2021).

This situation is by no means exclusive to the North Atlantic; arguably, it is a key characteristic of contemporary archaeological work. As noted by Belford (2020, 374), citing Gale (2003), 'the "idea that archaeological information and particularly understanding is uniquely ours" is no longer tenable, although a shift of position relies on the greatest accessibility of data and analysis'. Communicating our bundles of data, chains of inference, and interpretations based on them both to other researchers from different disciplines and to a wider public poses distinct challenges. In addition to excellent science communication, it requires thorough and harmonised documentation of research methods, as well as data and knowledge domain description.

The NABO community, together with its wider circle of collaborators (hereafter referred to as NABO+), aimed to develop tools to meet these challenges through dataARC. They identified the need to share research data and explain the connections between those data and higher-order interpretations and research outcomes. By doing so, they intended to support the reuse of data in new

transdisciplinary synthesis research. They further aimed to ground the wider public understanding of high-level interpretations more firmly in their evidence base. DataARC engaged community members in design work and user testing to develop an infrastructure to meet these needs.

In this article, we outline the core motivations behind dataARC, briefly introduce its socio-technical context, and summarise the tools, platforms and (meta)data products developed. We then undertake a critical review of the project's work, focusing on the needs of the stakeholder groups it was intended to serve, the principles that guided the design of the infrastructure, and the extent to which these principles are successfully implemented at present. Drawing on this review, we consider how the infrastructure, in whole or in part, might be reused by similar research communities.

In reflecting on dataARC, we highlight key socio-technical gaps which, if left unaddressed, may emerge as structural barriers to transdisciplinary, engaged, and open research. We emphasise the growing gap between the demands of the open data and impact agendas and the resources and infrastructures available to meet them. We further illustrate tensions between the open data and impact agendas. In framing these reflections, we draw on Fredheim's discussion of the relationships between engaged and participatory practices and open research in archaeology (Fredheim 2020). Fredheim has argued that, 'archaeologists must rethink their understandings of publics and expertise before we can begin to move towards a more critical understanding of how archaeology may be both "open" and ethical' (Fredheim 2020, 6). Critical reflection on how these relationships and tensions are embedded into our research infrastructures is imperative if, as researchers, we are to help bridge them and effectively engage in the work of 'addressing complex problems and research questions posed by global social challenges' (British Academy 2016, 9).

2. Designing dataARC: the socio-technical context

All infrastructures are influenced by the socio-technical context in which they are designed and developed. The contemporary (Western European and North American, Anglophone) research context vocally promotes transdisciplinarity, interdisciplinarity and synergistic research, evidenced by the frequency with which these terms appear in funding calls and descriptions of research frameworks (Kerr 2020). In parallel, we are called (if not required) to pursue open research, encompassing open data, methods, code and publications, as promoted by the FAIR (Findable, Accessible, Interoperable, Reusable) data paradigm (Wilkinson *et al.* 2016; Stall *et al.* 2019). The rise of open research is further evidenced by the increase in requirements for data management plans and open data publication conditions for recipients of research funding. At the same time, we are encouraged to engage in impactful research, engage with communities and stakeholders, and create public benefits through our work. This combined trend toward transdisciplinarity, openness, and engagement represents a major shift in expectations for the practice of research. The international significance of the impacts of these shifts has been underscored by Pétursdóttir, as highlighted at the

2018 European Association of Archaeologists conference in a roundtable on 'Pan-disciplinary research and the future role for archaeology'. This roundtable linked the future of the discipline to a research culture that, 'strongly encourages interdisciplinary collaboration, theoretical and methodological development, as well as outreach, openness and cooperation beyond the confines of academia' (Pétursdóttir 2020, 1).

Archaeology and Heritage are well served by research infrastructures and frameworks developed to support openness and collaboration within disciplinary academic communities. ARIADNEplus, the Pelagios Network, and the CIDOC CRM exemplify successful initiatives that actively promote open research practices and support coordinated, collaborative research by providing models, tools and platforms for data integration, rich annotation, and cross-mapping. University libraries, disciplinary and generalist data archives and portals such as Neotoma, the ADS, tDAR, and OpenContext provide essential support, training and guides to good practice for open data. Specific funding is available for impact work, for example from the UK AHRC. Dedicated organisations for interdisciplinary research such as the US National Socio-Environmental Synthesis Center provide further support for research across disciplinary lines. These organisations provided both inspiration and practical models for dataARC's design.

However, to undertake the kind of impact-led transdisciplinary work called for by Pétursdóttir, we must go beyond our current open and collaborative, but fundamentally disciplinary and academic, research practices. DataARC aims to adapt and extend the models provided by these projects to develop an infrastructure for this new context. It specifically builds on the ontology provided by the CIDOC CRM (Bekiari *et al.* 2021; Bruseker *et al.* 2017), tools and approaches for mapping to it developed under the aegis of ARIADNE and ARIADNEplus (Aloia *et al.* 2017; Meghini *et al.* 2017; Niccolucci and Richards 2019), and the governance model of the Pelagios Network (Barker *et al.* 2016; Barker 2020).

Reworking these models and tools to serve new aims highlighted a series of socio-technical challenges which, while not new, are being rearticulated. The challenges of the context that shaped dataARC, introduced here, will be returned to after presenting the infrastructure and its reception.

2.1 Socio-technical challenge: labour

It is undeniable that additional labour is required to accomplish the current research context's objectives. This additional, often unpaid, labour in academic pursuits tends to disproportionately affect women (Lawless 2018), early career researchers (Coin 2018), and people of colour (David-Chavez and Gavin 2018; Latulippe and Klenk 2020). The COVID-19 pandemic exposed and exacerbated these issues that continue to exist in science and academic research (Malisch *et al.* 2020; Deryugina *et al.* 2021; Krukowski *et al.* 2021).

Despite the challenges of increasing labour requirements expected from researchers, many individual researchers, groups, and communities, including the NABO+ community, choose to pursue this work because they genuinely believe in its

value. As noted by Belford ([2020](#)) in his discussion of the roles (and failings as implemented) of Research Frameworks, another kind of disciplinary infrastructure intended to frame and enable research practice, 'Research outputs need to be socially relevant and to deliver public benefit – both as part of the spatial planning system and as a coherent "biosphere" of knowledge production. This gives rise to two further issues that are familiar concerns for archaeologists: resources and power'. Who will do this work, what support and resources will they be able to access, and what rewards are returned are questions that now sit at the centre of debates on the role and operation of academic work.

2.2 Socio-technical challenge: communicating how data and interpretations connect

DataARC was motivated by the contention that we, as domain specialists and researchers, should be actively involved in communicating how data relate to complex ideas and how non-experts might approach them. Equally, it was motivated by the idea that supporting infrastructures are necessary to do this well and meet our professional and ethical obligations.

The challenges of integration across long-standing disciplinary boundaries and working with wider communities are diverse. They range from the intellectual divides created by seemingly incompatible inferential models to the pragmatic incompatibilities arising from expectations around funding, emphasis on outreach and impact, and publication (Crow and Dabars [2019](#); Fontana *et al.* [2020](#); Gibbs and Beavis [2020](#); Kerr [2020](#)). The structural barriers created by differences in the character of datasets, metadata designed for in-discipline use, and unconnected semantic frameworks are particularly difficult to break down because both intellectual and technical work is required. While the need for tools and methods that support cross-disciplinary data, metadata and semantic integration has been highlighted repeatedly, we lack concrete suggestions for how to design them (Thompson *et al.* [2017](#); Levin and Svenningsen [2019](#); Horcea-Milcu *et al.* [2020](#); Shen [2021](#)).

Crumley *et al.* ([2018](#), 283), discussing the challenges of interdisciplinarity in the context of human-environment interactions research, explain the importance of developing supporting systems for the multi-faceted work of addressing human-environment interactions. They emphasise the need for approaches that both lead to robust research and speak to contemporary challenges. They note that 'Research on such interaction processes often requires deep insights and detailed analyses in both the natural and social sciences. The research therefore needs interdisciplinary groups of researchers working closely together in a way that enables a lively exchange among disciplines and beyond them, and an invitation to collaborate with practitioners'. Their discussion of historical ecology specifically identifies several critical needs to enable this type of interdisciplinary research. These include targeted publication venues and scholarly organisations that should enable exchanges and promote collaboration.

2.3 Socio-technical challenge: the intersection of transdisciplinarity, openness, and engaged research

Crumley *et al.* ([2018](#)) equally emphasise the importance of collaboration outside academia, notably with members of Indigenous communities and producers of local and traditional knowledge. Their proposals fall broadly within the approaches found in Collaborative Action Research (Bennett and Brunner [2020](#)), reflecting contemporary academia's emphasis on engaged and impactful research. Here the intersection of transdisciplinarity, open research, and engaged research again creates a new requirement. There are active political debates around human-environment interactions, particularly in the context of the climate crisis discourse (van der Leeuw *et al.* [2011](#); Armstrong *et al.* [2017](#); Holm and Winiwarter [2017](#)). In this context, grounding interpretations and insights firmly and explicitly in contextualised data that is drawn from multiple domains is essential if those data are to be responsibly shared with a wider audience. The specific requirements of transdisciplinary data-intensive research, particularly where intellectual transparency is a priority, have been summarised recently (Alexander *et al.* [2020](#)). This work highlights the need to address epistemological, ethical and practical concerns – particularly around qualitative data sharing – to support these practices.

Essentially, the transdisciplinary and engaged research contexts, combined with the open research culture, asks us to open data to non-expert communities. In turn, it creates an expectation that we will enable these non-expert communities, whether comprised of extra-domain researchers or interested community members, to engage responsibly and gainfully with these data. DataARC thus set out to develop a research infrastructure that would not only enable the discovery of data, as done by existing archival infrastructures and domain-specific data discovery portals, but also communicate how we scaffold from research data to specific interpretations, and then from these interpretations to overarching understandings of the past.

3. Summarising What Was Built

The core element of the dataARC infrastructure is a search interface, a portal allowing a user to search for data by period, location, or tagged concepts. This interface is designed to promote the discovery of data from multiple disciplines and to provide contextual information about how that data is connected to high-level concepts as part of the search process (Figure 1). The search interface's back end ingests datasets and metadata via [GitHub](#). Combinations of data elements from these datasets are mapped to a central concept map, which is mapped in turn to the CIDOC CRM. The front-end search interface and customised search and analysis interfaces created in [Jupyter](#) notebooks access the data, concepts, and mappings via an [API](#) (Application Programming Interface) built on the openAPI [specification](#).
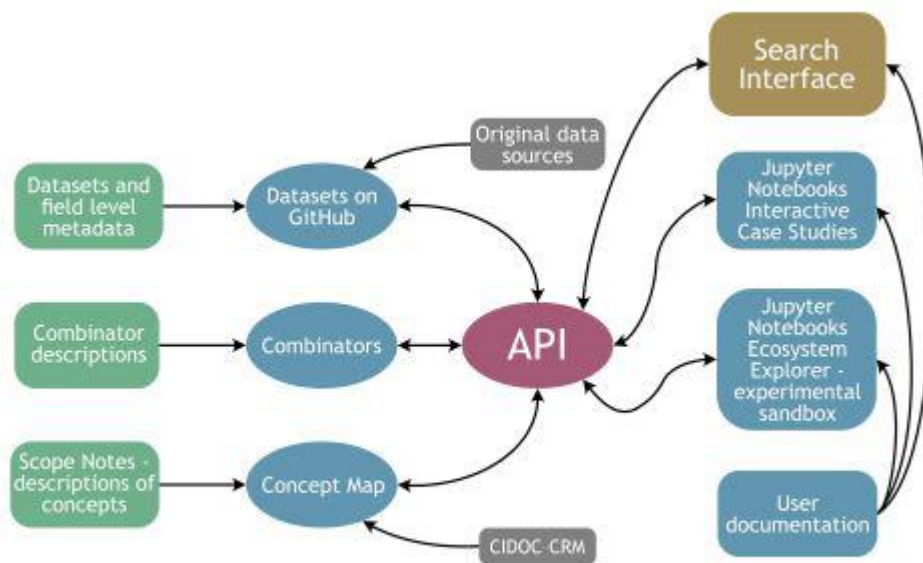
Figure 1: The dataARC infrastructure, whose main components are shown here, builds on the CIDOC CRM and leverages tools and platforms including GitHub, Jupyter notebooks, and the OpenAPI Specification.

The design of the tool is intended to encourage users to assess the potential relevance of datasets from outside their own area of expertise. The search interface provides three methods for filtering or querying data:

- Time (temporal filter, both calendar years and historical periods of interest)
- Space (spatial or geographic filter)
- Concepts (conceptual filter based on the community ontology, discussed below)

The results generated from these filters are structured to help users identify data-rich areas across multiple datasets. Most users will be familiar with data-rich areas in their own domain but are less likely to have detailed knowledge of complementary data from other domains.

In the dataARC infrastructure, a concept map (a domain ontology) enables the generation of search results designed to encourage cross-disciplinary exploration and discovery of data-rich areas from other domains. The concept map, an extract of which is illustrated in Figure 5, is built around the broad theme of 'the changing landscape'. It captures the core set of concepts that specialists would use in explaining how their data relate to the broad concept of landscape change. All the concepts appearing on the concept map are related to different degrees. The concept map is coded and visually represented as a graph, within which concepts appear as nodes and the reasons for connections between them appear as edges. The number of degrees of separation on the graph reflects the relatedness of concepts.

DataARC uses the 'graph' of these concepts, such as cattle and animal husbandry, to create points of direct and indirect connection between data used by different domains. Two data contributors could map to the same concept, such as animal husbandry. The same data contributors could map their data to two connected concepts: one could map to 'cattle' and the other to 'animal husbandry'.

As individual data contributors map their own data to the community concept map, they provide explanations for these mappings. These are referred to as 'combinators' in the infrastructure and its documentation. These explanations range from citations of core domain literature to commentary on areas of ambiguity or inference. These mappings from dataset-specific queries to concepts (nodes) on the shared concept map (graph) create connections between domain-specific data and shared concepts of interest (Figure 2). This structure reflects that a zooarchaeologist and an Icelandic saga specialist can provide information about the same concepts. However, neither researcher would wish to imply a direct relationship between their datasets. The structure, rather than pushing researchers to identify direct relationships between data, allows researchers to imply that both their datasets inform on a shared concept or on closely related concepts.
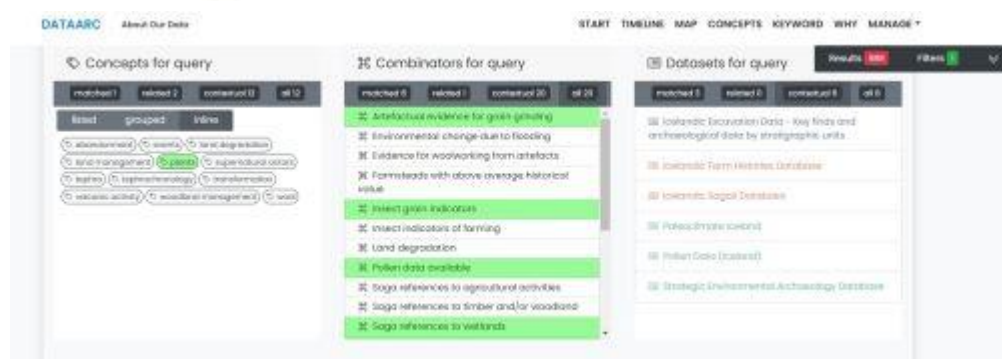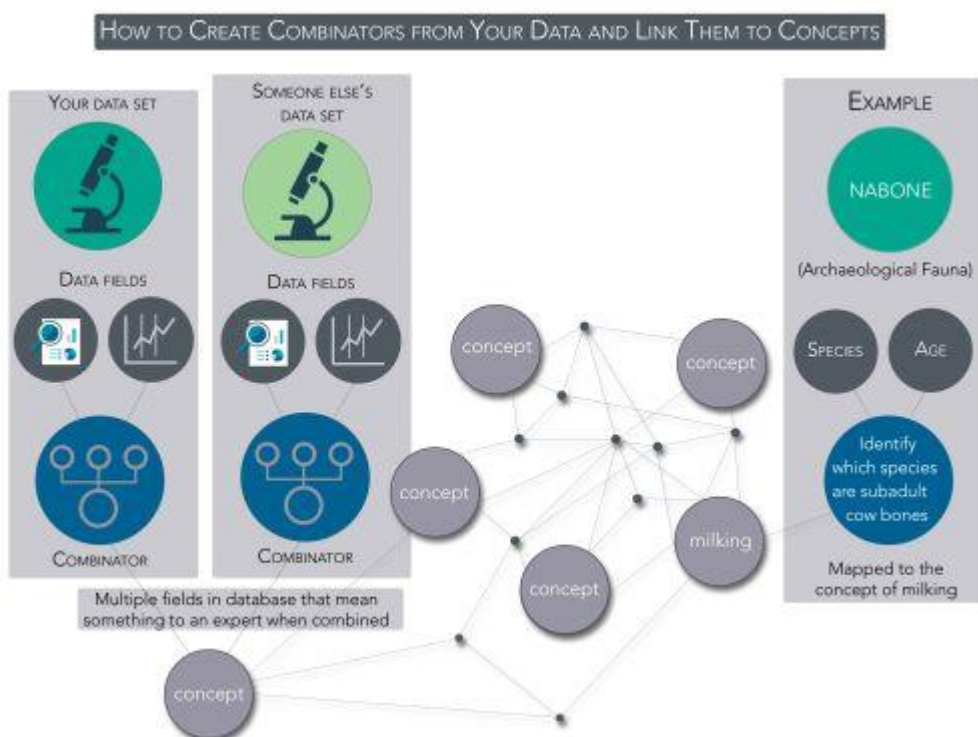
Figure 2: Concepts are linked to the results of queries on individual datasets via well-described mappings, referred to as 'combinators' within the dataARC infrastructure. (2a) A conceptual illustration of how combinators provide data to concept links (2b)

These relationships propagate through into the results returned by a search in the system. The results of any search are structured into three sections: matched, related, and contextual results. They are further grouped into archaeological, textual, and environmental categories (Figure 3). The matched results section displays the number of records directly returned from one or more applied filters. The data that are mapped to first-degree (related) and second-degree (contextual) connections on the concept map are then returned.

Importantly, this design choice results in the search outcomes that are more likely to include data contributed by community members working in multiple domains. Different domains tend to use different, but related, vocabularies when discussing their data. Therefore, different domain specialists are likely to map their data to different, but related, concepts. Inclusion of related and contextual results allows domain specialists to continue using vocabularies that are specific to their disciplines, while building bridges to other domains that use closely related concepts. Similarly, different domains tend to see their data as having different primary areas of application and individual specialists tend to map their data to the most relevant concepts. Returning related and contextual results to a domain specialist running a search is intended to point them to unexpected potential connections and related data. It provides an explicit mechanism to prevent the unintentional creation of disciplinary silos through search result sets.

Figure 3: An illustration of how the results returned by the dataARC search tool are organised. They are grouped into 'matched', 'related' and 'contextual' results and further organised into 'archaeological', 'environmental' and 'textual' types, and then by data source.

In addition to providing conceptual connections, dataARC search results include a 'Why' section, intended to provide further explanation of the search results. This 'Why' section, illustrated in Figure 4, allows a user to explore the data-to-concept (combinator) mappings as well as descriptions of concepts. This section therefore exposes relationships that exist among the different concepts used in the search and between concepts and related data.

Concepts are described in Scope Notes, which are modelled on those produced for the CIDOC CRM (Gill 2004). Scope Notes are written by the data creators and then made available for further contribution and discussion. They provide semantic documentation for the concepts and are an integral part of the infrastructure. While Scope Notes for concepts are relatively common in research data infrastructures, it is less usual to include further documentation describing the connections between concepts and queries on data. In dataARC, this additional layer of description and semantic documentation is provided for the combinator mappings. These are made visible in the 'Why' section of the search tool. While Scope Notes may be written by groups, combinator documentation is written by individual data contributors through a separate view within the user interface. Metadata for the datasets themselves are

generated using an XML template and guidance provided within the project's documentation. These metadata, together with the datasets, are maintained in GitHub. This allows for revisions and updates to be made within a versioned workflow and preserves versioning history.
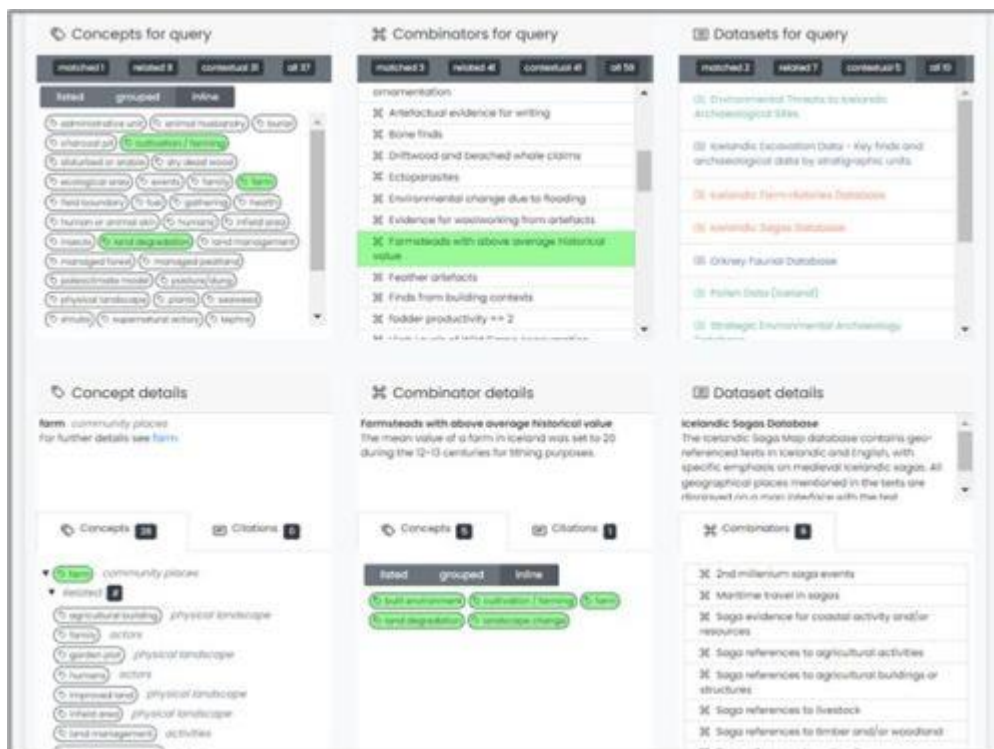

Figure 4: The 'Why' section of the search tool exposes the connections between concepts used in a search, data returned, and combinators, providing an explanation to an end user of how their search results were generated and why the data are considered relevant.

The search tool's interface is intended to serve as the primary means through which many users would begin to explore the data available. It provides basic search functions, access to explanations of the search results and documentation, and the ability to save and download search results. More advanced queries are possible through the project's API.

The dataARC API connects the back-end infrastructure and the search interface. It can be accessed via http, rest, or graphql protocols. It provides the technical mechanisms to ingest new datasets and combinators, and to produce responses to queries that include data provided by specialists in different domains. It also provides information about the connections between these datasets, together with background information about the data sources and how they are typically used. The latter is contained in linked documentation. The same API functions may be called from alternative interfaces, enabling advanced users to extend the functionality available through the search tool. It also allows users to create different analyses or visualisations of the data and documentation connected through dataARC. To illustrate how the API can be used, the project developed two example Jupyter notebooks that interactively illustrate the use of the infrastructure's linked concepts and data model through use cases, referred to as Research Data Explorations (Bankoff and Mejía Ramón 2021; Mejía Ramón and Bankoff 2021). These Research

Data Explorations demonstrate how data and connected concepts can be accessed via the API, integrated with external datasets, analysed and visualised. Two further exemplar notebooks are referred to as the Ecosystem Explorer. These are designed to help a potential data contributor understand the impacts of adding their data and of making new conceptual connections within the system. For a user interested in meta-analysis of the knowledge domain represented by dataARC, these notebooks also provide a demonstration of a network analysis approach to studying the connections between the concepts and data.

The search tool and other elements of the infrastructure are documented in user guides to running and interpreting searches, contributing data, developing data mappings, and using the Ecosystem Explorer. Further technical documentation for the infrastructure is maintained on GitHub. Given the non-standard structure of the search results, the documentation for the use of the search tool was viewed as integral to the infrastructure.

4. Reflections on dataARC's Design and Implementation

In the later stages of dataARC we undertook a critical review of the infrastructure, considering the problem(s) it tries to address, the character of its user community, and the principles that underpinned its design. Both the internal project team and external colleagues were consulted. The reflections presented here are based on these consultation exercises and on our own assessment of how the tool works. This assessment is thus grounded in discussions with the project's data contributors about their own experiences using the system and with the software development team across the course of the project. It is further informed by user observation and discussion with users during workshops held at several stages in the infrastructure's development. These include workshops at the Computer Applications in Archaeology Conference in Tubingen in Spring 2018, at the Applied Environmental Anthropology Conference in Edinburgh at the end of 2018, and at the Center for Advanced Spatial Technologies (*CAST*) in early 2021. In addition to gathering feedback from users, participants were surveyed and interviewed in later stages of dataARC. This feedback was gathered with the users' written consent, following research ethics protocols in place at CAST, which led the workshops where the data was gathered.

A robust evaluation of how the dataARC platform enables or constrains specific behaviours and activities awaits a formal assessment of the experiences of a larger, more diverse user community. We emphasise this lag in assessment of the long-term utility of dataARC because many of the principles behind the design are concerned with long-term patterns of activity, such as the incremental improvement of data mappings over time. We also recognise that unintended uses of dataARC will likely emerge over time as the community of users expands. The value of the initial evaluation presented here lies in insights gained and 'lessons learned' through the development process. These assessments are intended to encourage adaptation or reuse of the infrastructure's overall model, or its individual components in other research contexts. They are also intended to support continued experimentation with the design of data infrastructures for transdisciplinary research.

In our reflexive analysis, we explicitly consider and discuss the socio-technical context within which the dataARC infrastructure developed. This context is characterised, as outlined previously, by a significant shift in research culture to emphasise transdisciplinarity, open research methods, FAIR data, research impact, engagement and the generation of public benefits. We specifically explore how this wider socio-technical context impacts on the capacity of individuals and groups to engage with the infrastructure as they intend or desire. Within these reflections, we highlight the working and knowledge production practices of communities engaged in transdisciplinary data-intensive research and note where these practices gain support and, conversely, meet resistance. This reflexive discussion highlights the specific challenges of designing, constructing and using research infrastructures for emergent territories that cross disciplinary and institutional boundaries related to the broad domain of L-T H-E work and enable practices aligned with the imperatives of contemporary research.

5. Reflecting on the Community: who are we building for and what are their needs?

At its conception, the dataARC infrastructure aimed to address two ambitions of domain researchers. First, they wanted their data to be better connected with data created by their cross-disciplinary peers. Second, they wanted their data to be available and useful to wider publics. These wider publics include domain researchers or students exploring a new area or topic and seeking to identify potentially relevant data. They also include an individual from outside the project's core community using the data for purposes not envisaged by the project. This last group was originally conceived of as a researcher from a domain further from the defined problem space of L-T H-E, a person working on science policy or advocacy related to L-T H-E, or an interested member of the public. Data scientists later emerged as additional interested participants, seeking new approaches to working with diverse and disparate data and responding to the challenges of designing and maintaining community-owned infrastructures and resources (see [Section 6](#)).

Characterising these user communities to understand how they might engage with the infrastructure was a critical task. We initially solicited information on user needs through discussions with members of the wider NABO+ research community. We later developed user personas, which provided a clearer picture of the research habits in question. These exercises also began to reveal the character of the problems that these users encountered when trying to carry out their L-T H-E research and follow good practice in open research, with an eye toward future research impact.

Typical creators and users of research data who might contribute their data to the dataARC infrastructure used existing, simple tools. Most data creators used a spreadsheet together with a coding manual, or some other standard in use within their particular sub-domain's community. This standard was followed or disregarded

depending on immediate circumstances and individual data literacy (Kansa and Kansa 2021). They sometimes used more complex, less flexible, tools. Most common were databases that attempt to enforce coding standards when data are created. These databases were also intermittently updated and adapted to meet the immediate needs of specific situations. Most users were well aware of the problems introduced using simple, flexible data creation tools like spreadsheets. In summary, when we use simple unconstrained tools, we often end up bemoaning the inconsistencies of the data produced. Then in attempting to produce knowledge from data, we berate our past selves when we must work with the data within the confines of analytical tools and methods that demand consistency. Alternatively, when we use norm-enforcing tools, we often find ourselves bemoaning the faults in the design of the system, and later berate our past selves when we have used 'creative workarounds' to subvert the system's restrictive controls, all the while blaming the tool's creators, who seemingly left us with no other option than to do so. The prolonged, angst-ridden discourse around excavation diaries, context sheets, and their digital successors maps directly onto the problems identified in dataARC (Thomas 1991; Berggren and Hodder 2003; Opitz 2018; Richards and Hardman 2008).

In producing these data, users draw extensively on and reproduce tacit, domain-specific knowledge. This tacit knowledge ranges from basic items like the meanings of abbreviations or codes, as noted in Kansa and Kansa's 'We all know that a 14 is a sheep' (Kansa and Kansa 2013), to an understanding of the limitations of interpretations that may be made from a pollen core. While some of this tacit knowledge may be explained in coding manuals, undergraduate textbooks and course materials, workshop prompts, or the methods sections of postgraduate dissertations, it largely remains uncodified and unexplained for future users. This lack of documentation is, in part, because tacit knowledge is exceptionally difficult to capture coherently or succinctly. This tacit knowledge plays an important role in allowing us to operate with our simple tools. It directly supports the use of loosely enforced coding norms and unexplained, but well-understood, limitations on interpretation. The tools used do not have to do the work of enforcing these boundaries and behaviours, because that work is done through the domain-specific social norms transmitted through tacit knowledge. In other words, it is done that way because everybody knows that is how it is done. While this is functional for disciplinary work, these simple tools do not meet the requirements of synthesis research and transdisciplinary work because domain-specific tacit knowledge can't be relied upon across a whole transdisciplinary area. To support the latter, we must design tools that capture and summarise tacit knowledge.

Proposals to invest in tools intended to capture and summarise tacit knowledge can seem somewhat suspect (Garrick and Chan 2017). Domain researchers have invested heavily in accruing tacit knowledge and are conscious of its complexity and nuance. Summarising this nuance can seem like a poor investment of time and effort. Effectively, we are asking experts to repeat the work done in accruing tacit knowledge in the first place. The task may also seem bound to produce unsatisfactory results as summaries often fail to capture subtleties. On the other hand, the arguments for making this investment are well rehearsed, particularly in the context of interdisciplinarity, where cross-domain translation is necessary (Rust 2004; MacLeod 2018; Reed and Meagher 2019; Bammer *et al.* 2020). Beyond

the question of the value of the exercise, we must consider that domain researchers are, on the whole, under-resourced and time poor. These limitations are not intended as complaints, but as a central and essential design consideration. The dismay expressed by a participant in one of the project's workshops at the prospect of having to learn to do anything substantively new during their data creation process, no matter how easy or worthwhile, reflects the lived experience of many data creators. Adding a 'documentation of tacit knowledge' step to the data creation process, even if worth doing, requires labour for which there is little resource.

This lived experience presents a real challenge to developing research infrastructures, given the assumptions of the benefits of open data, open code, open science, and open knowledge. These assumptions include that everyone benefits from the collective investments and the greatest success emerges from a virtuous cycle of making generous intellectual investments without immediate reward (Marwick *et al.* 2017). This situation is even more acute in the context of transdisciplinary research. The scale of the generous investment required is increased because the data must be made not only findable and accessible, but open in an intellectual sense. This requires that tacit knowledge is documented and shared, semantically interoperable, and made reusable by contextualising the data. This aspect of dataARC's experience echoes comments that mapping data to the CIDOC CRM may be overly complex and too burdensome for some use cases (Kansa 2014).

While being increasingly encouraged to engage in transdisciplinary research, researchers have found that contributing their data sources comes with additional work. Data contributors for one domain are asked to make yet another investment as domain researchers to ensure their data are 'more-than-FAIR.' These investments are, in the culture of a research community, necessary to produce 'valid knowledge' from data.

This challenge is exacerbated for a domain researcher from further afield, or a person working on science policy or advocacy, or an interested member of the public. What investment is required from them to engage with the infrastructure in a way that is valuable to themselves (Manuel-Navarrete *et al.* 2021)? The dataARC platform ostensibly provides an initial gateway for a naïve searcher into the interdisciplinary domain of L-T H-E. Pragmatically, it primarily allows for the identification of who is in the conversation (that is, the data creators), rather than the ability to get to grips with the substance of the conversation through co-mapped data and shared concepts. For this group of users, the tool instead allows them to identify researchers who have expertise in different subjects that can speak to the broader implications of L-T H-E research. That said, the core design elements of the infrastructure do not prioritise the needs of this community. We read this (im)balance as reflective of the L-T H-E community's interest in engaging with the impact agenda, which is strong, but which stops short of enshrining this aim as a central driver when designing their tools and systems. This transitional situation is characteristic not only of the L-T H-E research community but of academic research communities more widely, as we collectively seek to reorient ourselves to align with the recent shift in expectations around how research is done.

6. Implications: reflecting on design principles

The dataARC infrastructure was designed through a collaboration between the domain scientists who are the primary group of intended end users, data scientists and software developers. While a formal co-design method was not discussed at the outset of the project, the software development team participated in research meetings led by domain and data scientists. In turn, domain scientists participated in design meetings about the tool and tested prototypes. The project broadly followed a co-design model, in that the tool was designed with the user community's active involvement at multiple junctures.

The design of the dataARC infrastructure reflects a set of core principles. These shaped the development of both the data search tool, the creation of associated elements of the infrastructure such as the Ecosystem Explorer, and the Scope Notes collection. These principles reflect the needs and aspirations of the infrastructure's community, as expressed by the domain scientists involved, translated through the data science and development teams into the language of design.

7. dataARC's Design Principles

- **Make data FAIR** – In 2014-2016, as dataARC emerged in parallel with the crystallisation of the FAIR data agenda (Wilkinson *et al*. 2016), Lynn Yarmey of the Research Data Alliance encouraged the foregrounding of open data and FAIR data practice in the early stages of design work and prototyping. The project's interdisciplinary scope and the decision to not manage archiving of the datasets, which remains with the dataset contributors, led to an emphasis on technical and semantic interoperability and their roles in enabling reuse.
- **Embrace ambiguity** – Rather than trying to get diverse disciplinary experts to adhere to a single strict definition of a term (e.g., the definition of 'animal husbandry'), the system revealed the process of negotiation among different disciplinary perspectives. The system also highlights the uncertainties present in inference and interpretation through explanations of data-to-concept mappings.
- **Show your workings** – Explanations of the data are needed to make ambiguity workable. The system was designed to provide places for contributors and users to explain their thinking. This includes also allowing them to provide references to domain literature to contextualise the decisions made in creating mappings between data and concepts. This is intended to aid non-specialists in understanding the data and how to interpret them. This emphasis on explanation is increasingly common in archaeological research infrastructures and systems. Specifically, dataARC's implementation of spaces for explanation echoes the approach taken in Perio.Do (Aloia *et al*. 2017) and the emphasis on mapping parallels the approach taken in ARIADNE (Meghini *et al*. 2017).
- **Encourage exploration outside of disciplinary boundaries** - DataARC was designed to highlight search results from multiple disciplines. The presentation of the search results, grouped into textual, environmental and archaeological categories, enables users to push themselves beyond their intellectual comfort zones by promoting and highlighting the potential relevance of data from related domains.

- **Prioritise connectivity** – A community-owned central ontology - a conceptual map of the shared research space - provides a web of connections between the perspectives and data worlds of diverse users of the infrastructure (Figure 5). Adding more connections to the ideas, explanations, and data provided by other community members is valuable for individual users and for the group.

Figure 5: A subsection of the concept map illustrates the diverse connections between related and contextual concepts, which are used to generate search results.

- **Give people credit for and control over their data** – Individual contributors of datasets should be recognised as the creators of those data. They should have a significant role in how those data are displayed, explained, and connected to shared frames of reference. This builds on the promotion of data citation in scientific publications, which encourages data creators to share and document their data (Parsons *et al.* 2010).
- **Favour deep explorations over speed** – While not explicitly drawing on a 'slow data' framing (Caraher 2015; 2019), the system encourages the creation of well-documented data and connections. Users are encouraged to spend time considering the data in depth. Experts' interpretations of it should support users and motivate them to explore further, rather than prioritising quick data discovery.
- **Be tolerant** – Acknowledge that an imperfect contribution of data and set of connections is preferable to no contribution. The user should allow for gaps in data and explanations, making it clear there is a gap, but not making this a barrier to participation. The platform provides a space for the data to continually improve and evolve its linkages to other datasets.
- **Enable the use of infrastructure and data for other projects** – Design the infrastructure to allow others to build on the data and tools, to take data out of the

system, to adapt code, and to pursue aims outside the project's original scope. This follows the ethos of reuse and remixing of open code communities gathered on platforms such as GitHub or, more strongly, of malleable software (Tchernavskij 2019).

- **Balancing valuing expertise with encouraging wider use of data** - Design the system to highlight the expertise behind mappings between data and concepts in order to underscore the importance of expertise in arriving at sound interpretations. At the same time, allow anyone to access and use the data.

8. Key Design Choices in Retrospect

While the use of design principles and practices can empower us to attempt to 'reconcile[s] our objectives, the terrain, and the social context in which we work' (Carver 2016, 35), the results of a design, when implemented, can diverge from the designers' aims (JafariNaimi *et al.* 2015). The gaps and tensions that revealed themselves through our initial evaluation and reflexive exercises highlight areas for future work on the dataARC infrastructure. They equally highlight areas where attempts at adaptation to the current research context are encountering what Huvila would characterise as 'sticky problems' (Huvila 2019). In each of the analyses of the implementation of design principles below, successes, partial successes, and failures are considered in the context of the wider current shift in research culture.

The F in FAIR: The findability of data from multiple domains is demonstrably improved by the dataARC infrastructure simply by aggregating data from domains that do not share repositories onto a single platform, while limiting the set of domains involved to maintain a coherent scope. Cross-disciplinary findability is also, arguably, improved through the presentation of the search results described above. Because almost any search on the system will return results from multiple datasets that are derived from research in different domains, the search results are more diverse than those generated by most search portals. These search results are an important part of the design of dataARC, because most infrastructures aim to return the results that are closest, or the best fit, to a query. In the dataARC infrastructure, the 'best' fit is defined as diverse matches to terms in the conceptual neighbourhood of the search terms. This approach to increased Findability is possible because of the dataARC infrastructure's well-defined scope. This scope allows it to sit between the domain-specific research data infrastructures, such as the European Pollen Database, and domain agnostic ones, for example Zenodo or the Harvard DataVerse. This allows for transdisciplinary search results tailored to a research theme. Cross-domain projects that are leveraging the availability of cloud-based databases or linked open data tools to bring together legacy datasets or present their own data can benefit from dataARC's illustration of how to increase 'Findability' across semantic gaps between domains.

Embracing ambiguity and showing your work: As open data initiatives encourage more researchers to add their data to archives and tag them with keywords or concepts in search portals, the web of implied connections woven by shared tags inevitably grows increasingly tangled. How users of these portals should understand the meaning of a tag, why it has been applied, and if a shared tag indicates a meaningful connection between two datasets remain important, open questions. This

is particularly important if we want to be able to generate coherent cross-disciplinary data assemblages.

The CIDOC CRM's use of ontologies and Scope Notes, adopted by dataARC, are attempts to allow expressions of ambiguity, include explanations, and indicate whether shared tags should imply a meaningful connection. DataARC's experience reflects that seen in the wider CIDOC CRM community: in constrained use cases such as mapping between multiple museum collections where there is a good degree of shared domain knowledge and a common intellectual framework, the CIDOC model works well. As the data types to be connected become more dissimilar, the exercise grows more fraught (Isaksen *et al*. 2009; Kansa 2014).

It is therefore unsurprising that the task of summarising potential sources of ambiguity in data providing explanations of connections was characterised as difficult by cross-disciplinary data contributor teams. The number of empty or minimally completed combinator descriptions at present reflects the impact of this perceived difficulty. In parallel, when users are interpreting the results of their queries, again we observed some discomfort caused by ambiguity. This discomfort was inferred from users' limited use of the 'Why' section.

In sum, we must conclude that dataARC had limited success in encouraging users to embrace ambiguity. That said, there may be confounding factors: Because data contributors have not fully embraced developing combinator descriptions, their incompleteness likely limits the usefulness of the 'Why' section. Further, some users immediately exited the tool by exporting their data after the results were returned and never interacted with the 'Why' section within the tool. The documentation of ambiguity presented in the 'Why' section is contained in the data export, so it is also possible this is engaged with outside the dataARC infrastructure.

Balancing valuing expertise with encouraging wider use of data: The dataARC infrastructure includes elements that highlight the need for expertise to interpret data correctly. The primary mechanism for this is the documentation of combinators, where data contributors provide commentary on limitations and uncertainties. The metadata provided for each dataset, which include field level documentation, further underscore the importance of expertise in the construction and use of these data. In parallel, the data are made freely available and downloadable, a necessary, if insufficient, first step to encouraging its wider reuse. The same documentation and metadata designed to highlight the importance of expertise can, in another light, be seen as encouragement for the interpretation of the data by non-specialists. Thus, while the infrastructure both highlights the value of expertise and encourages data reuse, its design does not explicitly attempt to communicate the trade-offs between these two imperatives. The documentation for the search tool and broader project materials on the website do underscore the balance between cautioning and encouraging non-experts in their reuse of data. Striking this balance will be essential for the imperatives of open data and transdisciplinarity to be reconciled.

Prioritising connectivity: The 'Why' section of search results displays 'related' and 'contextual' results when selecting a single concept. This functionality attempts to foreground the idea that datasets are interconnected. However, the density of connections between datasets via shared concepts depends on how individual

contributors have mapped their data to concepts. For example, data contributors may or may not choose to intentionally map to the same concepts selected by other contributors or to closely related concepts. Unless contributors intentionally co-map their data, some datasets could remain isolated from other related datasets. At present, the data contributor does not see how much connectivity is created when mapping data to concepts within the back-end data contributor interface. This could be accomplished using the Jupyter notebook-based tools within the Ecosystem Explorer. However, this process would require a level of scripting expertise that cannot be expected of the community contributing data and mappings. The infrastructure, therefore, is only partially effective at prioritising connectivity and improving the visualisation of connectivity is an important consideration for any reuse.

Favour deep explorations over speed: The system, while allowing rapid access to data and conceptual connections, does not provide a simple path to integrated data. Many users found creating a search challenging in comparison with their experiences using tools with a single Google-style search box and button. To search effectively, preparation and an understanding of how the tool functions are needed. This is an initial investment not required by comparable interfaces. Further, straightforward explanations of the connections implied within a search result collection are not easily found. The search results are presented in a 'browse' view in the 'Results' section, to be inspected individually before downloading for manipulation in aggregate. This interface feature constrains the speed at which data can be considered while within the search interface. Similarly, the 'Why' section, laid out in multiple connected panes that highlight individual connections among different elements of the search results, operates as a tool for 'browsing' concepts and data-query to concept mappings. This presentation of the results underscores that the search interface remains a data discovery tool to be used in the context of active research. The scrutiny and analysis of datasets, the assessment of the connections implied by the data mappings, and the creation of further inferential connections remains the work of the researcher. For the dataARC community, whether the deeper, slower exploration of connections will be valuable as a research tool remains an open question.

Give people control over their data: The current design provides a relatively open system for data contributors. Data contributors may contribute many mappings to concepts or only a few. They may describe these mappings in detail or not at all, add or edit many scope notes or none. They may attempt to build connections to other datasets or avoid this part of data contribution. The system assumes all users act in good faith, contributing good data and working to provide well-resourced combinator mappings. As the user community grows, a peer-review recommendation system, along the lines of that used in Wikipedia, may be needed to ensure the system remains anchored in usable data. Adding a review process can maintain standards set by the community. However, it may have unintended effects on the character of the community of contributors because of the additional effort needed to participate. This process also requires more investment on the part of the infrastructure's community, who would need to manage this peer review (Tennant *et al*. 2017). In academic publishing, how this labour is rewarded remains contested, as does the question of how labour is rewarded in community-owned infrastructures in diverse contexts (Hendricks *et al*. 2020; Pia *et al*. 2020). Because the dataARC infrastructure

is already, in the judgement of the project team, one that requires a high level of effort by data contributors, any adaptations that imply additional work should be considered carefully.

Be tolerant: The implemented infrastructure is permissive. For example, it allows incomplete or blank descriptions of data to concept mappings. It asks data contributors to provide metadata but does not block contributions that lack it. Equally creating sparse mapping to concepts is possible. The system does not attempt to clean data for users and, while encouraged, this is not required to contribute data. While this approach allows community members to contribute to dataARC without being 'all in', the effect on the overall usability of the infrastructure as its community grows remains to be seen. This functionality requires those searching for data to be tolerant of contributions that may not be fully documented or mapped to concepts and combinators. As discussed in relation to contributors' control over data, a peer-reviewing mechanism may be appropriate to balance an ultimately tolerant system, as demonstrated by Wikipedia-style platforms. However, again this implies additional effort to actively maintain the infrastructure.

9. Scrapyard: how to take apart and reuse/adapt the infrastructure

Because the user community for any given specialised digital infrastructures is limited, enabling the adaptation and reuse of the dataARC infrastructure by future projects was considered essential. DataARC's prioritisation of reuse potential in its design could be seen as the inevitable result of broader trends in the open-source software community. These trends are influential, but the strong conviction that research infrastructures are made to be shared is also influenced by the project's specific academic context. It reflects the scale of investment by institutions and consortia, the communities using these infrastructures, and the value of the research they produce.

The use of containerised or modularised components in the design of the interface and data model is intended to facilitate this future adaptation and reuse. The front-end interface design is itself modular, with each component (map, timeline, concept map, results, 'why' section) existing in its own software container. This design choice allows for the replacement of one, several, or all, front-end components. This functionality is demonstrated through the project's Jupyter notebooks, which effectively act as a parallel interface for the search tool, using data inputs from the API. The back-end interface for creating concept and combinator mappings could be redesigned along similar lines, again using inputs from the API and Jupyter notebooks. The infrastructure's code base, data and documentation are maintained on GitHub. Beyond the convenience of this platform for the development team, its use is intended to encourage the adaptation of the infrastructure. However, managing the connections between modular components when adapting the system for reuse presents some challenges. These are readily illustrated by planning out specific adaptations that reuse elements of the system.

Reuse adaptation example 1: The current system is designed to have data mapped against a single, versioned concept map. It is possible to rewrite the API to support the inclusion of multiple concept maps. Doing so could be valuable because interchangeable concept maps could represent multiple evolving perspectives. The API and the data-combinators-concepts model are tightly linked, in that the design of the API reflects the assumptions made in the concept map. Consequently, the API to data-combinator-concepts model links must be considered if redesigning the tool to support multiple concurrent concept maps or an editable concept map.

Reuse adaptation example 2: In the current infrastructure, each element's spatial coordinates and temporal range enable searches using the map and timeline in the interface. Many data do not have well-defined spatial or temporal ranges and it might be desirable to adapt the system to accommodate them. In this use case, it would be possible to remove the timeline and map-based components of the interface. Searches would then rely entirely on the concept map. Without any temporal or spatial constraints, it seems likely that numerous results would be returned by searches, particularly for the 'related' and 'contextual' result sets. This overload of results is not helpful to a user seeking to identify relevant data. The data model might be adapted to reflect changes in the interface, for example by adding another property that could be used for filtering.

Reuse adaptation example 3: Visualisations of individual datasets are likely to vary between systems. To support flexibility in the representation of data, templates are used to style individual datasets. These templates can be reconfigured by data contributors at any stage after loading data into the system. As the template used to represent data is configured at the per-dataset level, making changes to the data level display is relatively straightforward from a technical perspective. More challenging is the selection of visualisations that will prove understandable to both generalist users and domain experts. While the design challenges briefly illustrated above must be managed, they are not insurmountable. The important question, within the academic socio-technical context, is how we will adapt to shared ownership as social norm and manage expectations for ongoing investments of labour in research infrastructures. These questions have also been raised in discussions of community owned infrastructures such as iNaturalist (Basman and Tchernavskij 2018; Morell *et al.* 2021).

10. Concluding Remarks: implications for others

DataARC is designed to provide a toolkit for open, data intensive, impact-orientated, transdisciplinary research. It is planned to be maintained by the community that uses and benefits from it. The intent is to enable and encourage exchanges of ideas and data and the links between them, making it easier to work together in meaningful and sustainable ways. It is specifically designed to enable work that draws on the knowledge of multiple experts with different ways of knowing, aligned with research practices actively encouraged in the current academic context. DataARC aims to provide a model that is adaptable and reusable by other research communities. The

project's investment in building tools, together with this initial evaluation of their functionality, is intended to provide a starting point for other projects' technical work.

The project's work, moreover, highlights the need for co-design of research infrastructures and for investing in building the relationships among researchers, developers and designers. At the same time, it reveals how the current emphasis on community-led work, co-design, impact, public benefits, and open research paradigms has effectively devolved responsibility for substantive aspects of research infrastructures from institutions to self-organised groups, like dataARC. Notably, within L-T H-E research communities, emphasis on co-design and broad engagement, while beneficial in numerous ways in its involvement of users in the design process, requires their labour. Equally, the push for community ownership of research infrastructures in these domains, together with the benefits of giving communities more independence and control, comes with the additional burdens of maintaining and updating these infrastructures. In sum, significant shifts in what our research infrastructures need to be designed to do are in progress. Equally important, the bargains being made around the work of community building, design, development, testing, maintenance, updating, and extension for research infrastructures like dataARC are changing. In continuing to evaluate dataARC as a useful model for those investing in developing infrastructures for transdisciplinary, open, impactful, data-rich research, we must now consider if we are satisfied with the terms of these new bargains.

Acknowledgements

Bibliography

Aitchison, K. 2021 'No more polluter pays principle: opportunities and challenges of public benefit provision in UK development-led archaeology', *Internet Archaeology* **57**. https://doi.org/10.11141/ia.57.8

Alexander, S.M., Jones, K., Bennett, N.J., Budden, A., Cox, M., Crosas, M., Game, E.T., Geary, J., Hardy, R.D., Johnson, J.T., Karcher, S., Motzer, N., Pittman, J., Randell, H., Silva, J.A., da Silva, P.P., Strasser, C., Strawhacker, C., Stuhl, A. and Weber, N. 2020 'Qualitative data sharing and synthesis for sustainability science', *Nature Sustainability* **3**(2), 81–88. https://doi.org/10.1038/s41893-019-0434-8

Aloia, N., Binding, C., Cuy, S., Doerr, M., Fanini, B., Felicetti, A., Fihn, J., Gavrilis, D., Geser, G., Hollander, H., Meghini, C., Niccolucci, F., Nurra, F., Papatheodorou, C., Richards, J., Ronzino, P., Scopigno, R., Theodoridou, M., Tudhope, D., Vlachidis, A. and Wright, H. 2017 'Enabling European archaeological research: the ARIADNE E-Infrastructure', *Internet Archaeology* **43**. https://doi.org/10.11141/ia.43.11

Armstrong, C.G., Shoemaker, A.C., McKechnie, I., Ekblom, A., Szabó, P., Lane, P.J., McAlvay, A.C., Boles, O.J., Walshaw, S., Petek, N., *et al.* 2017 'Anthropological contributions to historical ecology: 50 questions, infinite prospects', *PloS One* **12**(2), e0171883. https://doi.org/10.1371/journal.pone.0171883

Bammer, G., O'Rourke, M., O'Connell, D., Neuhauser, L., Midgley, G., Klein, J.T., Grigg, N.J., Gadlin, H., Elsum, I.R., Bursztyn, M *et al.* 2020 'Expertise in research integration and implementation for tackling complex problems: when is it needed, where can it be found and how can it be strengthened?', *Palgrave Communications* **6**(1), 1–16. https://doi.org/10.1057/s41599-019-0380-0

Bankoff, R.J. and Mejía Ramón, A.G. 2021 *Quantifying an ever-changing landscape*. https://www.data-arc.org/my-front-page/example-data-explorations/quantifying_vignette/ [Last accessed: 3 May 2021].

Barker, E., Cañamares, P. de S., Isaksen, L. and Simon, R. 2016 'The Pleiades gazetteer and the Pelagios project' in L. Berman, R. Mostern and H. Southall, *Placing Names: Enriching and Integrating Gazetteers*, 97–109. http://dx.doi.org/10.17613/M6233X

Barker, E. 2020 'What would Indy do? Resisting post-truth through the practice of annotation' in C. Roosevelt (ed) *Spatial Webs: Mapping Anatolian Pasts for Research and the Public*, Istanbul: Koç University Research Centre for Anatolian Civilizations, 5–31. http://oro.open.ac.uk/69161/

Basman, A. and Tchernavskij, P. 2018 'What lies in the path of the revolution', *Proceedings of the 29th Annual Workshop of the Psychology of Programming Interest Group (PPIG 2018)*. https://ppig.org/files/2018-PPIG-29th-basman.pdf

Becker, E., Jahn, T., Stiess, I. and Wehling, P. 1997 *Sustainability: A cross-disciplinary concept for social transformations*, Vol. 6, Paris: Unesco.

Bekiari, C., Bruseker, G., Doerr, M., Ore, C.-E., Stead, S. and Velios, A. 2021 *Volume A: Definition of the CIDOC Conceptual Reference Model*, (Version 7.1.1). https://doi.org/10.26225/FDZH-X261

Belford, P. 2020 'Borderlands: rethinking archaeological research frameworks', *The Historic Environment: Policy and Practice* **11**(2–3), 359–81. https://doi.org/10.1080/17567505.2020.1737777

Bennett, H. and Brunner, R. 2020 'Nurturing the buffer zone: conducting collaborative action research in contemporary contexts', *Qualitative Research*, 1468794120965373. https://doi.org/10.1177/1468794120965373

Berggren, Å. and Hodder, I. 2003 'Social practice, method, and some problems of field archaeology', *American Antiquity* **68**(3), 421–34. https://doi.org/10.2307/3557102

British Academy 2016 *Crossing Paths: Interdisciplinary Institutions, Careers, Education and Applications*, report compiled by D. Soskice. https://www.thebritishacademy.ac.uk/publications/crossing-paths/

Bruseker, G., Carboni, N. and Guillem, A. 2017 'Cultural Heritage Data Management: The Role of Formal Ontology and CIDOC CRM' in M.L. Vincent, V.M. López-Menchero Bendicho, M. Ioannides and T.E. Levy (eds) *Heritage and Archaeology in the Digital Age: Acquisition, Curation, and Dissemination of Spatial Cultural Heritage Data*, Cham: Springer International Publishing. 93–131. https://doi.org/10.1007/978-3-319-65370-9_6

Caraher, W. 2015 'Slow archaeology', *North Dakota Quarterly* **80**(2), 43–52.

Caraher, W. 2019 'Slow archaeology, punk archaeology, and the "Archaeology of Care"', *European Journal of Archaeology* **22**(3), 372–85. https://doi.org/10.1017/eaa.2019.15

Carver, M.O.H. 2016 *Making Archaeology Happen: design versus dogma*, Routledge.

Catlin, K.A. 2016 'Archaeology for the Anthropocene: scale, soil, and the settlement of Iceland', *Anthropocene* **15**, 13–21. https://doi.org/10.1016/j.ancene.2015.12.005

Coin, F. 2018 'When love becomes self-abuse: gendered perspectives on unpaid labor in academia' in Y. Taylor and K. Lahad (eds) *Feeling Academic in the Neoliberal University*, Cham: Palgrave Macmillan. 301–20. https://doi.org/10.1007/978-3-319-64224-6_14

Crow M.M. and Dabars, W.B. 2019 'Restructuring research universities to advance transdisciplinary collaboration' in K.L. Hall, A.L. Vogel and R.T. Croyle (eds) *Strategies for Team Science Success: Handbook of Evidence-Based*

Principles for Cross-Disciplinary Science and Practical Lessons Learned from Health Researchers, Cham: Springer International Publishing. 477–88. https://doi.org/10.1007/978-3-030-20992-6_37

Crumley C.L., Lennartsson, T. and Westin, A. 2018 *Issues and Concepts in Historical Ecology: The Past and Future of Landscapes and Regions*, Cambridge University Press.

David-Chavez, D.M. and Gavin, M.C. 2018 'A global assessment of Indigenous community engagement in climate research', *Environmental Research Letters* **13**(12), 123005. https://doi.org/10.1088/1748-9326/aaf300

Deryugina T., Shurchkov, O. and Stearns, J. 2021 'COVID-19 disruptions disproportionately affect female academics', *AEA Papers and Proceedings* **111**, 164–68. https://doi.org/10.1257/pandp.20211017

Fontana, M., Iori, M., Montobbio, F. and Sinatra, R. 2020 'New and atypical combinations: an assessment of novelty and interdisciplinarity', *Research Policy* **49**(7), 104063. https://doi.org/10.1016/j.respol.2020.104063

Fredheim, L.H. 2020 'Decoupling "Open" and "Ethical" archaeologies: rethinking deficits and expertise for ethical public participation in archaeology and heritage', *Norwegian Archaeological Review* **53**(1), 5–22. https://doi.org/10.1080/00293652.2020.1738540

Gale, F. 2003 *East and Northeast Wales – Later Prehistoric. Research Framework for the Archaeology of Wales*. [Last accessed: 11 September 2021]. PDF

Game, E.T., Tallis, H., Olander, L., Alexander, S.M., Busch, J., Cartwright, N., Kalies, E.L., Masuda, Y.J., Mupepele, A.-C., Qiu, J *et al.* 2018 'Cross-discipline evidence principles for sustainability policy', *Nature Sustainability* **1**(9), 452–54. https://doi.org/10.1038/s41893-018-0141-x

Garrick, J. and Chan, A. 2017 'Knowledge management and professional experience: the uneasy dynamics between tacit knowledge and performativity in organizations', *Journal of Knowledge Management* **21**(4), 872–84. https://doi.org/10.1108/JKM-02-2017-0058

Gibbs, P. and Beavis, A. 2020 *Contemporary Thinking on Transdisciplinary Knowledge: What Those Who Know, Know*, Springer Nature.

Gill, T. 2004 'Building semantic bridges between museums, libraries and archives: the CIDOC Conceptual Reference Model', *First Monday* **9**(5). https://firstmonday.org/ojs/index.php/fm/article/download/1145/1065

Hambrecht, G., Feeley, F., Smiarowski, K., Hicks, M., Harrison, R., Brewington S., Cesario G. and Gibbons, K. 2019 'A millennium of Icelandic archaeological fish data examined against marine climate records', *Quaternary Research*, 1–17. https://doi.org/10.1017/qua.2019.35

Hambrecht, G., Anderung, C., Brewington, S., Dugmore, A., Edvardsson, R., Feeley, F., Gibbons, K., Harrison, R., Hicks, M., Jackson, R., Ólafsdóttir, G.Á., Rockman M., Smiarowski, K., Streeter, R., Szabo, V. and McGovern, T. 2020 'Archaeological sites as Distributed Long-term Observing Networks of the Past (DONOP)', *Quaternary International* **549**, 218–26. https://doi.org/10.1016/j.quaint.2018.04.016

Hendricks, G., Tkaczyk, D., Lin, J. and Feeney, P. 2020 'Crossref: the sustainable source of community-owned scholarly metadata', *Quantitative Science Studies* **1**(1), 414–27. https://doi.org/10.1162/qss_a_00022

Hicks, M. 2019 *Community, Ecology, and Modernity: Faunal Analysis of Skútustaðir in Mývatnssveit, Northern Iceland*, Doctoral Dissertation, City University of New York. https://academicworks.cuny.edu/gc_etds/3316/

Holm, P. and Winiwarter, V. 2017 'Climate change studies and the human sciences', *Global and Planetary Change* **156**, 115–22. https://doi.org/10.1016/j.gloplacha.2017.05.006

Horcea-Milcu, A.-I., Martín-López, B., Lam, D.P.M. and Lang, D.J. 2020 'Research pathways to foster transformation: linking sustainability science and social-ecological systems research', *Ecology and Society* **25**(1), art13. https://doi.org/10.5751/ES-11332-250113

Huvila, I. 2019 'Management of archaeological information and knowledge in digital environment' in M. Handzic and D. Carlucci (eds) *Knowledge Management, Arts, and Humanities: Interdisciplinary Approaches and the Benefits of Collaboration*, Cham: Springer International Publishing, 147–69. https://doi.org/10.1007/978-3-030-10922-6_8

Isaksen, L., Martinez, K. and Earl, G. 2009 'Archaeology, formality & the CIDOC CRM'. Presented at the *Interconnected Data Worlds: Workshop on the Implementation of the CIDOC CRM, Berlin, Germany (23-24 November 2009)*. https://eprints.soton.ac.uk/69707/

JafariNaimi, N., Nathan, L. and Hargraves, I. 2015 'Values as hypotheses: design, inquiry, and the service of values', *Design Issues* **31**(4), 91–104. https://doi.org/10.1162/DESI_a_00354

Jorgenson, A.K., Fiske, S., Hubacek, K., Li, J., McGovern, T., Rick, T., Schor, J.B., Solecki, W., York, R. and Zycherman, A. 2019 'Social science perspectives on drivers of and responses to global climate change', *WIREs Climate Change* **10**(1), e554. https://doi.org/10.1002/wcc.554

Kansa, E. 2014 'Open context and linked data', *ISAW Papers* **7**(10). http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/kansa/

Kansa, E.C. and Kansa, S.W. 2013 'Open archaeology: we all know that a 14 is a sheep: data publication and professionalism in archaeological communication', *Journal of Eastern Mediterranean Archaeology and Heritage Studies* **1**(1), 88–97. https://doi.org/10.5325/jeasmedarcherstu.1.1.0088

Kansa, E. and Kansa S.W. 2021 'Digital data and data literacy in archaeology now and in the new decade', *Advances in Archaeological Practice* **9**(1), 81–85. https://doi.org/10.1017/aap.2020.55

Kerr, S. 2020 'The future of archaeology, interdisciplinarity and global challenges', *Antiquity* **94**(377), 1337–48. https://doi.org/10.15184/aqy.2020.138

Krukowski, R.A., Jagsi, R. and Cardel, M.I. 2021 'Academic productivity differences by gender and child age in Science, Technology, Engineering, Mathematics, and Medicine Faculty during the COVID-19 pandemic', *Journal of Women's Health* **30**(3), 341–347. https://doi.org/10.1089/jwh.2020.8710

Latulippe, N. and Klenk, N. 2020 'Making room and moving over: knowledge co-production, Indigenous knowledge sovereignty and the politics of global environmental change decision-making', *Current Opinion in Environmental Sustainability* **42**, 7–14. https://doi.org/10.1016/j.cosust.2019.10.010

Lawless, B. 2018 'Documenting a labor of love: emotional labor as academic labor', *Review of Communication* **18**(2), 85–97. https://doi.org/10.1080/15358593.2018.1438644

Levin, G. and Svenningsen, S.R. 2019 'Digital transdisciplinarity in land change science – integrating multiple types of digital data', *Geografisk Tidsskrift-Danish Journal of Geography* **119**(1), 1–5. https://doi.org/10.1080/00167223.2019.1585893

MacLeod, M. 2018 'What makes interdisciplinarity difficult? Some consequences of domain specificity in interdisciplinary practice', *Synthese* **195**(2), 697–720. https://doi.org/10.1007/s11229-016-1236-4

Malisch, J.L., Harris, B.N., Sherrer, S.M., Lewis, K.A., Shepherd, S.L., McCarthy, P.C., Spott, J.L., Karam, E.P., Moustaid-Moussa, N., Calarco, J.M., Ramalingam, L., Talley, A.E., Cañas-Carrell, J.E., Ardon-Dryer, K., Weiser, D.A., Bernal, X.E. and Deitloff, J. 2020 'Opinion: in the wake of COVID-19, academia needs new solutions to ensure gender equity', *Proceedings of the National Academy of Sciences* **117**(27), 15378–15381. https://doi.org/10.1073/pnas.2010636117

Manuel-Navarrete, D., Charli-Joseph, L., Eakin H. and Siqueiros-Garcia, J.M. 2021 'Applying technologies of the self in transformation labs to mobilize collective agency', *Social Innovations Journal* **5**. https://socialinnovationsjournal.com/index.php/sij/article/view/696

Marwick, B., d'Alpoim Guedes, J., Barton, C.M., Bates, L.A., Baxter, M., Bevan, A., Bollwerk, E.A., Bocinsky, R.K., Brughmans, T., Carter, A.K., Conrad, C., Contreras, D.A., Costa, S., Crema, E.R., Daggett, A., Davies, B., Drake, B.L., Dye, T.S., France, P., Fullagar, R., Giusti, D., Graham, S., Harris, M.D., Hawks, J., Health, S., Huffer, D., Kansa, E.C., Kansa, S.W., Madsen, M.E., Melcher, J., Negre, J., Neiman, F.D., Opitz, R., Orton, D.C., Przstupa, P., Raviele, M., Riel-Savatore, J., Riris, P., Romanowska, I., Smith, J., Strupler, N., Ullah, I.I., Van Vlack, H.G., VanValkenburgh, N., Watrall, E.C., Webster, C., Wells, J., Winters, J. and Wren, C.D. 2017 'Open science in archaeology', *SAA Archaeological Record* **17**(4), 8–

14. http://onlinedigeditions.com/publication/?i=440506&article_id=2888896&view=articleBrowser&ver=html5

McGovern, T.H., Hambrecht G. and Hicks, M. 2019 'Historical ecology and longitudinal research strategies around Lake Mývatn, Iceland' in C. Ray and M. Fernández-Götz (eds) *Historical Ecologies, Heterarchies and Transtemporal Landscapes*, Routledge. 32-42.

Meghini, C., Scopigno R., Richards, J., Wright, H., Geser, G., Cuy, S., Fihn, J., Fanini, B., Hollander, H., Niccolucci, F *et al.* 2017 'ARIADNE: a research infrastructure for archaeology', *Journal on Computing and Cultural Heritage (JOCCH)* **10**(3), 1–27. https://doi.org/10.1145/3064527

Mejía Ramón, A.G. and Bankoff, R. 2021 *Data mining the past*. https://www.data-arc.org/my-front-page/example-data-explorations/history_vignette/ [Last accessed: 3 May 2021].

Morell, M.F., Cigarini, A. and Hidalgo, E.S. 2021 'A framework for assessing the commons qualities of citizen science: comparative analysis of five digital platforms', *SocArXiv*, Submitted May 25. https://doi.org/10.31235/osf.io/pv78g

Niccolucci, F. and Richards, J. 2019 *The ARIADNE Impact*, Hungary: Archaeolingua Foundation. https://doi.org/10.5281/zenodo.3476712

Opitz, R. 2018 'Publishing archaeological excavations at the digital turn', *Journal of Field Archaeology* **43** (sup. 1), S68–S82. https://doi.org/10.1080/00934690.2018.1505409

Parsons, M.A., Duerr, R. and Minster, J.-B. 2010 'Data citation and peer review', *Eos, Transactions American Geophysical Union* **91**(34), 297–98. https://doi.org/10.1029/2010EO340001

Pétursdóttir, Þ. 2020 'Collaborative, Public, Participatory: Pros and Cons of an Open Archaeology', *Norwegian Archaeological Review* **53**(1), 1–4. https://doi.org/10.1080/00293652.2020.1782463

Pia, A., Batterbury, S., Joniak-Lüthi, A., LaFlamme, M., Wielander, G., Zerilli, F., Nolas, M., Schubert, J., Loubere, N., Franceschini, I., Walsh, C., Mora, A. and Varvantakis, C. 2020 *Labour of Love: An Open Access Manifesto for Freedom, Integrity, and Creativity in the Humanities and Interpretive Social Sciences*, Copyright, Fair Use, Scholarly Communication, Etc. https://digitalcommons.unl.edu/scholcom/159

Reed, M. and Meagher, L. 2019 'Using evidence in environmental and sustainability issues' in A. Boaz, H. Davies, A. Fraser and S. Nutley (eds) *What Works Now? Evidence-informed Policy and Practice*, Policy Press.

Richards, J.D. and Hardman, C.S. 2008 'Stepping back from the trench edge: an archaeological perspective on the development of standards for recording and publication' in M. Greengrass, and L. Hughes (eds) *The Virtual Representation of the*

Past, Farnham, Surrey and Burlington, USA: Ashgate Publishing Company, 101-12. https://eprints.whiterose.ac.uk/7795/

Rockman, M. and Hritz, C. 2020 'Expanding use of archaeology in climate change response by changing its social environment', *Proceedings of the National Academy of Sciences* **117**(15), 8295–8302. https://doi.org/10.1073/pnas.1914213117

Rust, C. 2004 'Design enquiry: tacit knowledge and invention in science', *Design Issues* **20**(4), 76–85. https://doi.org/10.1162/0747936042311959

Shen, Y. 2021 'Transdisciplinary convergence: intelligent infrastructure for sustainable development', *Data Intelligence* **3**(2), 261–273. https://doi.org/10.1162/dint_a_00063

Sigurðardóttir, R., Newton, A.J., Hicks, M.T., Dugmore, A.J., Hreinsson, V., Ogilvie, A.E., Júlíusson, Á.D., Einarsson, Á., Hartman, S., Simpson, I.A *et al.* 2019 'Trolls, water, time, and community: resource management in the Mỳvatn District of Northeast Iceland' in L.R. Lozny and T.H. McGovern (eds) *Global Perspectives on Long Term Community Resource Management*, Springer. https://doi.org/10.1007/978-3-030-15800-2_5

Stall, S., Yarmey, L., Cutcher-Gershenfeld, J., Hanson, B., Lehnert, K., Nosek, B., Parsons, M., Robinson, E. and Wyborn, L. 2019 'Make scientific data FAIR', *Nature* **570**(7759), 27–29. https://doi.org/10.1038/d41586-019-01720-7

Tchernavskij, P. 2019 *Designing and Programming Malleable Software*, These de doctorat, Université Paris-Saclay (ComUE). https://www.theses.fr/2019SACLS499

Tennant, J.P., Dugan, J.M., Graziotin, D., Jacques, D.C., Waldner, F., Mietchen, D., Elkhatib, Y.B., Collister, L., Pikas, C.K., Crick, T., Masuzzo, P., Caravaggi, A., Berg, D.R., Niemeyer, K.E., Ross-Hellauer, T., Mannheimer, S., Rigling, L., Katz, D.S., Greshake Tzovaras, B., Pacheco-Mendoza, J., Fatima, N., Poblet, M., Isaakidis, M., Irawan, D.E., Renaut, S., Madan, C.R., Matthias, L., Nørgaard Kjær, J., O'Donnell, D.P., Neylon, C., Kearns, S., Selvaraju, M. and Colomb, J. 2017 'A multi-disciplinary perspective on emergent and future innovations in peer review', *F1000 Research* **6**. https://doi.org/10.12688/f1000research.12037.3

Theis, T. and Tomkin, J. 2015 *Sustainability: A comprehensive foundation*, University of Indiana Open Textbook Press. https://openlibrary-repo.ecampusontario.ca/jspui/handle/123456789/306

Thomas, R. 1991 'Drowning in data? - publication and rescue archaeology in the 1990s', *Antiquity* **65**(249), 822–28. https://doi.org/10.1017/S0003598X00080546

Thompson, M.A., Owen, S., Lindsay, J.M., Leonard, G.S. and Cronin, S.J. 2017 'Scientist and stakeholder perspectives of transdisciplinary research: early attitudes, expectations, and tensions', *Environmental Science and Policy* **74**, 30–39. https://doi.org/10.1016/j.envsci.2017.04.006

van der Leeuw, S., Costanza, R., Aulenbach, S., Brewer, S., Burek, M., Cornell, S., Crumley, C., Dearing, J., Downy, C., Graumlich, L., Heckbert, S., Hegmon, M., Hibbard, K., Jackson, S., Kubiszewski, I., Sinclair, P., Sörlin, S. and Steffen, W. 2011 'Toward an integrated history to guide the future', *Ecology and Society* **16**(4). https://doi.org/10.5751/ES-04341-160402

Wilkinson, M.D., Dumontier, M., Aalbersberg, Ij.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J. and Mons, B. 2016 'The FAIR Guiding Principles for scientific data management and stewardship', *Scientific Data* **3**(1), 160018. https://doi.org/10.1038/sdata.2016.18