A vision for Open Archaeology

Anthony Beck, School of Computing, University of Leeds, LS2 9JT, UK, a.r.beck@leeds.ac.uk

Cameron Neylon, Science and Technology Facilities Council, Rutherford Appleton Laboratory, Harwell Oxford, Didcot, OX11 0QX, cameron.neylon@stfc.ac.uk

Abstract

By unblocking knowledge bottle necks and enhancing collaborative and creative input 'open' approaches have the potential to revolutionise science, humanities and arts. 'Open' has captured the zeitgeist: but what is it all about? Is it about providing clear and transparent access to knowledge objects: data, theories and knowledge (open access, open data, open methods, open knowledge)? Is it about providing similar access to knowledge acquisition processes (open science)? Obviously it is, however, this is not the whole story. Open approaches require active engagement. This is not just engagement from the 'usual suspects' but engagement from a broader societal base. For example, primary data creators need the appropriate incentives to provide access to Open Data - these incentives will vary between different groups: contract archaeologists, curatorial archaeologists and research archaeologists all have different drivers. Equally important is that open approaches raise a number of issues about data access and downstream data re-use. This paper will discuss these issues in relation to the current situation in the UK and in the context of the DART project: an Open Science research project.

Keywords: Open Archaeology, Open Science, research, practice, policy, management, DART

Introduction

So what is Open Archaeology? The *archaeology* obviously refers to the enactment of activities within the domain of archaeology. The *open* relates to the philosophies and freedoms espoused by the communities that develop Open approaches. These groups promote free redistribution and access to the specifications, designs, implementation details, data, transformations and synthesis associated with a 'thing'. This includes the well-established groups that develop and promote Open Source Software, Open Standards, Open Formats and Open Data (for example: HM Government, 2012) and the more nascent developments within the communities that undertake Open Research, also known as Open Science (Wikipedia, 2012a). Open Archaeology shares much of the philosophy of these open approaches and is predicated on promoting open redistribution and access to the data, processes and syntheses generated within the archaeological domain. This is aimed at both the production and consumption of archaeological knowledge with the associated aim of maximising transparency, re-use and engagement whilst maintaining professional probity.

Current approaches to archaeological knowledge

Every landscape is at risk from a variety of natural and anthropogenic forces. The loss of some of these resources is inevitable, but the decision making process is a delicate balancing act between different cultural, industrial and economic issues which can be enacted at different scales: from the local to the international. Risk and benefit determination is crucial in determining impact and

informing policy led mitigation. A good quality up-to-date knowledge base, which collates all the appropriate information and data relevant to a specific topic and makes it available in a suitable manner, is an important resource for decision makers. The quality and currency of this knowledge base has an inevitable impact on the quality of, and the uncertainties associated with, the decision making process. Data, and the subsequent information and knowledge that is derived from an integrated data corpus, should be used effectively to produce realistic policy (i.e. policy which can be enacted and produces the desired impact) and develop effective governance strategies (be they local, regional, national or global).

Unfortunately, access to the corpus of knowledge (data, theory and practice) is fragmented and opaque. Historic Environment Information Resources (HEIRs) are distributed between different government agencies, universities, NGOs (national and international), charities and companies. In addition there are barriers to data sharing in the form of bureaucracy, formats (particularly paper archives) and incompatible structures and formats. Important data resides in silos which effectively block access to different stakeholders on the basis of awareness, privilege (access rights), cost and security. This lack of access to data leads to downstream fragmentation in the data transformation environments (theory, processing methods etc.), the knowledge structuring environments (classification framework: dating, pottery sequences etc.) and interpretations. Stakeholder requirements are not joined up: this means that decisions, policy or research are not enacted with 'perfect' knowledge[1].

The ramification is that research and decision making is undertaken on a sub-set of the corpus of data. This leads to a Rumsfeldian system of knowledge availability (see Figure 1):

- the known knowns: the accessible knowledge
- the known unknowns: the known but inaccessible knowledge
- the unknown unknowns: the potential knowledge advances gained by integrating all data, collaborating with different domains and future research avenues

---

[1] This term is borrowed from economics. It refers to the complete understanding of a domain of discourse which allows Perfect Competition: when every economic actor has complete knowledge of the marketplace then the most price effective form of competition can occur.
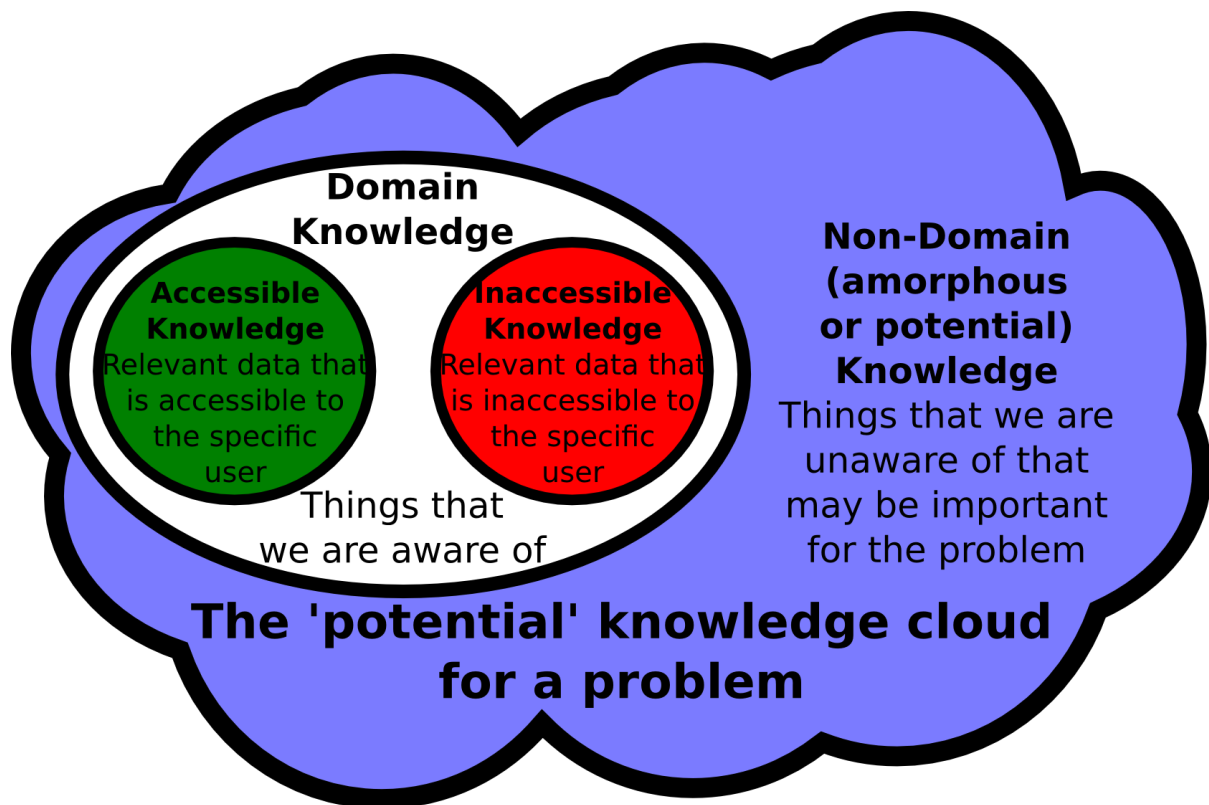
Ideally one would reduce the size of the latter two domains so that decisions, be they research, policy or management, can be formed from a position of 'perfect', or 'near-perfect', knowledge with information and data which is both accessible and well understood.

The underlying ethos is that better decisions are made when users can access appropriate parts of the knowledge base in a manner which is relevant to the problem they are solving. Providing a heritage naive decision maker with a complex synthesis of a bronze-age landscape for a planning enquiry does not improve the decision making capability of that individual (i.e. the synthesis is not fit for the purpose of aiding a planning decision). Providing mechanisms where 'source data' can be re-faceted, or transformed, to provide a 'fit-for-purpose' dataset or visualisation which is clear and unambiguous and can integrate into the consuming user's business process is essential to maximise utility. Subtleties may be lost in generalised derivatives, but as the underpinning data, information and knowledge is accessible then it can be mined for further detail if required. This entails timely and appropriate access to 'objects'[2] or resources required to make a decision, test a hypothesis, conduct research or undertake management. In the 21st century 'timely and appropriate access' can only be achieved through open activities: in this context 'Open Archaeology'.

This requires the development of two different strands: *open data* (which can be transparently re-used) and *dynamic data* (which dynamically reflects changes in the underlying data).

What would a dynamic and open archaeology look like?

Developments in information and communication technology are increasingly framing how we operate both professionally and socially. Cheap and accessible communication tools enable formal

---

[2] Objects is used as a catch all for data, information and knowledge 'things' be they from researchers, curators, the public or businesses.

and informal groups to coalesce around any shared interest, identity or activity. Interoperable access to data, services and syntheses provide the building blocks for a range of knowledge ecosystems. Shared processing (workflow) systems, hosted in cloud environments, can process and visualize these heterogeneous data via online services and automatically generate and maintain the processing metadata. The workflow and metadata provide unambiguous detail on how data are transformed into information. This is important for downstream data-centric applications where provenance information is critical. The semantic web and linked data have the potential to transform static archives into dynamic resources that fully articulate the impact of change as the supporting knowledge based is refined. Such frameworks that provide ubiquitous access to data and other resources will require communities to address such issues as access, accreditation, copyright and licensing.

Within such an environment it becomes easier to imagine an 'open archaeology' in the following way. Fine grained data downloaded from geophysics instruments or collected during, likely to be fully digital, excavations is placed into a virtual 'folder'. This folder synchronises the data with a cloud based repository. This invokes a variety of services which generate description and discovery metadata (essentially archiving the raw data). Using ontologies and other semantic web tools the data can interoperate as Linked Data with the corpus of data collected at this and other scales. Persistent Uniform Resource Identifiers (URIs) are allocated for each object for long term referencing. The underlying data quality can be evaluated and improved in relation to the network of data, information and knowledge in which the 'new' object exists. If and when aspects of the data are updated the changes re immediately reflected in the 'corpus'. When users want to query this resource they can either access it directly (as a service or a download) or utilise mediation interfaces which have been specifically designed to transform the resource into specific data and knowledge products that the user requires (again as a download or a service). The terminology may strip this of elegance, but these types of environments are currently under development and technically achievable. They represent a major conceptual shift from depositing 'static' documents, objects and data in archives to developing dynamic, rich and interlinked repositories of knowledge. The Archaeology Data Service (ADS) has traditionally hosted static material with a focus on long term preservation of content. However, the ADS are aware of the difference between re-use and preservation and are developing data dissemination, and other, services based on semantic web technology.

Open Data as a dynamic resource

If archaeological data is to be viewed as a dynamic resource it is important to consider the relationships between the different data, information and knowledge objects. Archaeology deals with a variety of different data which are integrated to gain insights into the past. Some of these are measurements and observations. Others are classifications and groupings which help structure the observations and provide significant detail on process and cultural significance. Built on top of both of these are interpretations. For simplicity these have broken these down into three groups:

- Physical observations – measurement and observation data (many of which can not be replicated again) including sensor and laboratory measurements (such as geophysical and soil measurements). This data does not change over time. For example a sherd of pottery may have a location, context, dimensions and can be attributed with form and fabric classification descriptors.
- The structuring knowledge/classification environment – the knowledge (or classification) frameworks which can be attributed to the physical observations. These frameworks reflect clusterings and groupings in the data. This data is dynamic. For example the pottery form and fabric sequence (and their associated dating implications) change when new data is added to the corpus or specialists refine their points of view. A similar position is observed with radiocarbon dates and the body of knowledge that provides calibration for the raw

dates. There should be a tight relationship between these relationships and the physical observations which together represent a dynamic knowledge base which can be fed into research, policy, practice and management. Currently this relationship is not formally represented and is normally decoupled.

- Analysis, interpretations and synthesis – a layering of multiple interpretative points of view based on hypotheses and bodies of theory and evidence.

It should be noted that in recent years this contrast has rightly been subject to revision, as various commentators have noted that all stages of archaeological practice involve theory-laden assumptions, and hence that data collection and interpretation are closely entwined. Irrespective of this broader debate, each of these components require additional metadata that describes the methods of observation, collection and analysis so that subsequent re-users can understand issues pertaining to scale, uncertainty and ambiguity that are essential when datasets are integrated. Being able to link, and therefore integrate and query, the different data resources means that heterogeneous resources can be treated as one. This is close to becoming Linked Data (2012). A Linked Data approach preserves much, if not all, of the underlying structure and semantics of the source data by employing ontology[3] and other Knowledge Organisation Systems (KOS). This allows the delivery of richer data and can provide more complete answers to queries.

Open processes and methods

Open data, and synthesis in open access publications, do not provide the full supply chain for open and reproducible science. What is missing is the methodology, the process whereby data is transformed into information and information is transformed into knowledge. Methodology can encompass any linear or reflexive transformative process, from algorithms to a body of theory, and transcend the barriers between science and the humanities (Beck, In Press). Open Processing includes providing access to code, processing chains (workflows) and descriptive metadata. Workflows are an ideal way to explicitly document and share metadata about data transformation and processing. A popular open source workflow enactment environment is Taverna Workbench (Taverna, 2012). Taverna allows workflows to be shared with other people through myExperiment (2012), the social web site for scientists. myExperiment represents a distinctly social and open approach to collaborating and conducting science, where loosely coupled communities can interact and share data, workflows and experiences.

What are the benefits of an Open Archaeological approach?

An example of the potential is probably useful at this juncture. In addition to many other things pottery provides essential dating evidence for archaeological contexts. However, pottery sequences are developed on a local basis by individuals with an imperfect knowledge of the whole. This means there is overlap, duplication and conflict between different pottery sequences developed by individuals which are periodically reconciled (your Type IIb sherd is the same as my Type IVd sherd and hence the dating range can be refined). This is the perennial processes of lumping and splitting inherent in any classification system. The semantic integration of localised sequences can potentially support more robust pottery frameworks. In addition, as the pottery data is linked (and not decoupled and stale) then updated pottery classifications and dating implications immediately update the dating probability density function for a context or group. One can also reason over the data to find out which contexts, relationships and groups are impacted by a change in the dating sequences either by proxy or by logical inference (a change in the date of a context produces a logical inconsistency with a stratigraphically related group). As an aside, if the data is stored as RDF triples then the logical consistency in the physical and stratigraphic relationships can be

---

[3] an ontology is a formal representation of a set of concepts within a domain and the relationships between those concepts. It is used to reason about the properties of that domain, and may be used to define the domain.

automatically verified using reasoning software (such as Prolog). As all of the data is stored as Linked Data, this means that all the primary data archives are linked to their supporting knowledge frameworks (such as a pottery sequence). When a knowledge framework changes the implications are propagated through to the related data dynamically. This means that, in theory, the implications of minor changes in the structuring knowledge environment can be tracked dynamically through to the underlying observations and their impact observed in interpretations. The feedback mechanism implied by Linked Data can profoundly alter the way archaeologists, and others, engage with their data and derived resources. This also means that any policy, management, curatorial and research decisions are based upon data that reflects the most-up-to date information and knowledge.

Deposition is no longer the final act of the excavation process: rather it is where the dataset can be integrated with other digital resources and analysed as part of the complex tapestry of heritage data. The data does not have to go stale: as the source data is re-interpreted and interpretation frameworks change these are dynamically linked through to the archives. Hence, the data sets retain their integrity in light of changes in the surrounding and supporting knowledge system. This introduces a greater reliance on explicitly exposing the techniques and methods used to process the linked open data.

What are the challenges of a dynamic and open archaeology?

This paper argues that Open Archaeology can have a significantly positive impact on many of the modes of archaeological consumption. The challenge is one of how the environment can be engineered to increase the likelihood of this happening. This is about incentives and how individuals and organisations embed Open Archaeology principles into their work practice.

Data access

Access to data is obviously important, and is mandated by professional and legislative frameworks throughout the world. In the UK the majority of archaeological excavation work is undertaken as part of the planning process. The National Planning Policy Framework (NPPF: Department for Communities and Local Government, 2012) describes the planning policies on the conservation of the historic environment in England and Wales. One of the specific objectives of NPPF is to enhance the corpus of archaeological evidence:

*Local planning authorities should **make information about the significance of the historic environment gathered as part of plan-making or development management publicly accessible**. They should also require developers to record and advance understanding of the significance of any heritage assets to be lost (wholly or in part) in a manner proportionate to their importance and the impact, and **to make this evidence (and any archive generated) publicly accessible**. (Department for Communities and Local Government, 2012)*

Academic and other research archaeologists also make a contribution to the corpus of archaeological evidence but a much greater contribution to the theoretical and analytical debate as well as the different modalities of interpretation. The majority of research is directly or indirectly funded by the public purse through the government funded research councils. Research Councils UK (RCUK, 2012a) is a strategic partnership of the UK's seven Research Councils and undertakes cross cutting and co-ordinating activities. The RCUK have established common principles on data policy applicable across the funding councils (RCUK, 2012b). The relevant principle states that *'Publicly funded research data are a public good, produced in the public interest, which **should be made openly available with as few restrictions as possible in a timely and responsible manner that does not harm intellectual property**.'*

One of the problems is that although publically accessible archives are advocated the guidance does not stipulate repositories, structures and licences that maximise re-use. Furthermore, As Bradley

(2006) states 'The idea of preservation really refers to the written report' with the worrying comment that 'The vast majority of field projects undertaken in the United Kingdom are never published in any form other than the client report.' Client reports and other 'synthetic' derivatives do not provide enough information on source detail and processing methods that allows the scientific re-appraisal of technique nor are they in a format that facilitates the re-use of content embedded in the documents.

There is also a gap between the amount of work commissioned and the number of outputs submitted for deposition and archiving. For example, Online Access to the Index of Archaeological Investigations (OASIS: Hardman, 2009) is an ideal environment for the deposition of 'grey literature' client reports and is becoming part of the business process of contract units (as part of their best practice process) and the planning authorities (as part of their tender compliance process). It is estimated that OASIS only contains approximately 10% of the archaeological reports from the first decade of the 21st Century. Hence, gaining access to the appropriate synthetic reports can still be difficult (Bradley, 2006; Ford, 2010). The size of this gap for excavation data and other resources needs quantifying.

One of the problems in academia is that the reward system is based upon publication output and not on the output of other, more fundamental, research objects. It is important that this balance is redressed and that appropriate checks are put in place to confirm that the outputs are in deposited in appropriate repositories. The Wellcome Trust, for example, is strengthening the manner in which it enforces its open access policy: failure to comply with the policy could result in final grant payments being withheld and non-compliant publications being discounted when applying for further funding (Wellcome Trust, 2012). The Royal Society goes even further. The report 'Science as an open enterprise' (Royal Society, 2012) describes not only how open data is beneficial to science, society and policy but recognised that changes in culture and communication are required to maximise impact.

Contract archaeology has different drivers. Whilst NPPF states that the underlying data should be made publically accessible, the mechanism of deposition does not mean that the resources can be easily re-used. The current commercial business process is predicated on delivering reports and depositing 'paper archives' that satisfy both client and curatorial requirements so that the process of excavation can be signed off and the rest of the development can proceed. The problem here is that the nature of the object (data) has moved from a paper, analogue, record to a digital record whilst many curatorial systems are still designed with analogue data in mind. If the surrounding data re-use and analysis environment is generated then it is arguable that the very act of depositing structured data into the corpus is enough to satisfy the requirements of the planning process as this data will be immediately available for review and analysis by multiple stakeholders. This means that the contract unit would no longer be required to produce a hand-crafted report that includes textual summaries of data (this kind of synthesis, if necessary, can be built directly from the deposited data itself). Instead one could generate alternative analytical derivatives (for example to examine the impact of the work on regional agendas) and synthetic outputs (for example producing popular synthesis which encourages people to re-use the data). As many large UK contract units are also charities with an educational remit this repositioning would enhance engagement and provide a catalyst for public access to open data. This should make the excavation process cheaper for the commercial archaeological contractor, and by proxy, the client. The incentives in the commercial framework are more about changing the execution of the business process rather than incentivising individuals.

The culture of change
Although contract and academic archaeologists will be the creators of the majority of data, curatorial archaeologists are key stakeholders and mediators of knowledge. The technologies and approaches described in this paper have the potential to significantly disrupt this sector. Although

many benefits have been described the very process of change has social implications and would need effective management. For at least the past decade UK curatorial services at a regional and national level have been significantly cut. This has damaged morale. The introduction of systems that fundamentally change working practice are likely to create issues around role perception, recognition and job security. These will need effective management. Whilst these archaeologists will still produce curatorial and planning advice they will start to base this advice on a variety of data sources rather than just the HEIR databases they maintain. This may be seen to erode prestige and relevance: this will also need managing. However, rather than spending time maintaining and updating a dataset based on at least secondary synthesis, curatorial archaeologists will be directly accessing the full corpus of archaeological knowledge to address regional and national policy and research agendas. This has the potential to be more engaging and mean that the sector has more relevance to both planning, research and the public. It could be argued that large scale projects that employ many archaeological contractors, such as the Channel Tunnel Rail Link, could see many benefits by opening the data from these projects at an early stage so that the curatorial archaeologists can work more effectively with the consultant, contract and research archaeologists to develop more nuanced strategies and outputs.

Engagement and ownership

There are also issues of public engagement and accountability. By making the results of academic and contract archaeology open then the public can engage with the resource more effectively and link planning and research policy actions directly to outcomes (a form of process transparency). A number of incumbent governments have had aspirations to make local and national decision making systems more transparent. There is also another facet, and one which is followed by the authors: open is the 'right thing to do'. This is particularly relevant to academics who are funded through the public purse: it's a duty to communicate their research and its social and ethical implications both to policy-makers and to the non-specialist public. There is an increasing view that these types of non-academic engagements are richer, participatory, relationships that go far beyond simple outreach (MORI, 2000; Royal Society et al., 2006; Poliakoff and Webb, 2007).

What is not clear here is when to deposit: the principles do not prescribe when, in the lifecycle of a research project, data should be released. It is recognised that those who collect data should, if they choose, have 'a limited period of privileged use of the data they have collected to enable them to publish the results of their research'. Most projects should deposit their data openly at the end of a traditional project life cycle (end of research or contract) subject to ethical, commercial and exploitation caveats described in the RCUK principles. This obviously has implications for those projects that are on-going or have a longitudinal dimension. However, there is also an inverse. There are people and organisations who want to make their data openly available as soon as possible after collection. This is at the extreme end of Open Science. Data is made available irrespective of whether the research or analysis team has tools to process this data or has written peer review journal papers. This means that the community can get access to the data at a very early stage. Whilst this may appear to put the data collectors at a disadvantage there are a number of benefits:

Data credit: the data, although open, is not in the public domain, it is released under an open licence. If a 'by attribution' clause is utilised then the data collectors have to be acknowledged when the data is used. It is likely that references for data will become more important for career progression.

Building communities and networks: access to good quality data designed to be re-used is not always straightforward. Writing grants to collect good data to produce publications is also time consuming. Making good quality data available to a community means that research teams can evolve around the data. As the data is open it also allows multi-disciplinary re-use. The practice of

conducting research can be enhanced as more researchers can analyse the data. Interesting collaborations can develop.

Open Science advocates opening access to data, and other scientific objects, at a much earlier stage in the research life-cycle. Open Scientists argue that research synergy and serendipity occurs (i.e. greater research and scientific advances) through openly collaborating with other researchers (more eyes/minds looking at the problem (Johnson, 2011; Nielsen, 2011)). They would further argue that such open approaches could lead to the earlier identification of processes or strategies which will have a profound impact on effective policy or curation programmes. Of great importance is the fact that the scientific process itself is transparent and can be peer reviewed: by exposing data and the processes by which this data is transformed into information other researchers can replicate and validate techniques (Beck, In Press). Open Science removes the barriers around science 'objects' which in turns means that Citizen Science (Hand, 2010; Wikipedia, 2012b) and Crowdsourcing (Cooper et al., 2010; Wikipedia, 2012c) techniques can be used to collect and analyse data and techniques can be found to harness latent micro-expertise to solve particular scientific problems. Citizen Science, Crowdsourcing and the exploitation of micro-expertise enhance public and professional participation and engagement and should result in effective 'science in society'. As a consequence it is believed that collaboration is enhanced and the boundaries between public, professional and amateur are blurred.

Open Archaeology in society

To recap, Open Archaeology promises to make digital archaeological content available in a manner which facilitates re-use and under licences that allow people to do anything with the content. There are already established environments where this data is used and will play a crucial role (i.e. research frameworks, planning control etc.). However, there are enormous benefits that could stem from public accessibility and the different public, professional and mixed communities that develop around the data as they collaborate on solving specific problems or re-use issues. These are communities of practice:

Communities of practice are groups of people who share a concern, a set of problems, or a passion about a topic, and who deepen their knowledge and expertise in this area by interacting on an ongoing basis (Wenger et al., 2002).

Communities of practice are not new: they have always been around. Recognition of the importance of communities of practice coupled with improvements in global communication that underpins both the internet and social media means that loosely coupled communities including different stakeholders can be created around any shared interest. Development in communication technology and access to resources means that the nature of the collaborations will be different (Townsend et al., 2009). This will be part of a continuing expansion of 'heritage in society' and the redefining of relationships between different heritage stakeholders. Many of the collaborators will be amateurs. Many field-activities have built and sustained relationships with volunteers, however amateurs represent a vastly under-utilized resource for many curatorial and research activities. Much like the citizen scientists at Galaxy Zoo (2012), Ancient Lives (2012) or Old Weather (2012) a legion of amateurs can help to reduce the size and enhance the utility of the "little-used and inaccessible data mountain" (English Heritage, 1995). Open approaches encourage serendipity: amateurs and professionals utilising and 'mashing up' resources for exploratory purposes. Interestingly Johnson (2011) argues that networked open systems are the most important drivers of innovation and serendipity plays an important role.

However, there are practical and ethical problems particularly if one views citizen scientists simply as a resource: the 21st century version of 'trowel fodder'[4]. This can rapidly diminish goodwill and returns and lead to the impression that the professional community wishes to exploit the amateur rather than undertaking serious engagement with the aim to develop mutually beneficial collaborative frameworks (Hand, 2010). Haklay (2011) criticises this passive role and argues for more inclusive and nuanced collaboration. The use of citizen scientists in any process almost demands that the results of that process become open: surely the people who freely contribute to the development of a project should also have access to the information (Wiggins and Crowston, 2011)?

The ramifications of opening archaeology

Like other disciplines, such as ecology, there are many potential issues in increasing access to archaeological data and other knowledge resources. Many of these issues revolve around professional practice and in particular the treatment of human remains (Wikipedia, 2012d). Most issues full under one of two themes:

Access to data and derivatives that still have resonance for a community or cultural group (indigenous groups, 20th Century war excavations and forensic archaeology).

Providing access to data that facilitates looting and the international trade in illegal artefacts.

The distinction between these two points is mainly one of boundaries and certainty. In respect of point 1 it is reasonably clear when an activity may have ethical implications, even in retrospect. Remedial activities can be taken and the corpus of evidence can be redacted or obfuscated for different audiences accordingly. However, point 2 is significantly more ambiguous. The same data which is used to conduct archaeological research or curatorial management can be used to target looting activities (for example, night hawking: the illegal act of metal detecting without the landowners consent or on scheduled sites (Oxford Archaeology, 2009)). In the latter example it is difficult to know what to do – removal of access rights to the archaeological knowledge base may reduce the problem but has a profound impact on legitimate activities. In respect of access to digital data this is effectively the position adopted by many curatorial authorities. Data can be accessed by visiting curatorial facilities but is generally not accessible on-demand externally.

Unfortunately every stakeholder can produce compelling arguments as to why they should get preferential access to everyone else's data whilst at the same time they should place embargoes on their own. The underlying issue is one of data embargoes and how these are implemented, or not. In this respect the archaeological discipline is in a difficult position. On the one hand many of the statutory, legislatory and funding frameworks call for public access of the products of archaeological research and practice. On the other hand there are concerns about the implications of sharing any archaeological data as it can be used by looters. These two positions are polarised and obviously conflicting.

It is important that the risks and benefits associated with providing access to heritage data are considered from a pragmatic basis. Some data are sensitive: it can be argued that access to such data should be restricted for reasons of security and ethics (particularly where these issues are clearly defined – such as indigenous archaeologies). It should also be recognised that data sensitivities and risks change over time. A data set which may be deemed sensitive whilst under excavation (at risk whilst it is accessible) is less sensitive once built over (at less risk as it is less accessible). Furthermore, providing access to data does not always mean providing access to raw data: data can be obfuscated by generalising spatial, temporal and other attributes. This implies

---

[4] A term used in UK excavation for an inexperienced archaeologists who is burdened with the majority of the manual tasks (http://www.doubletongued.org/index.php/dictionary/trowel_fodder/)

some form of user accreditation. Whilst not ideal, and definitely not open, access based on user accreditation does mean that different user groups get access to data that might otherwise have been embargoed. The Portable Antiquities Scheme (PAS, 2012) has successfully adopted this approach for accessing data at different degrees of granularity, including a 'research' grade level. This user accreditation framework allows a more nuanced approach to knowledge segmentation which provides a re-use position based upon accreditation and trust. By understanding these issues fit-for-purpose systems can be developed that deploy data in a timely and effective way.

It is also important to consider the changing modes of operation. Until very recently there has been a mix between analogue and digital techniques with an emphasis on analogue. However, the discipline is rapidly moving towards completely digital collection, analysis and dissemination workflows. The publication process is predicated on pre-20th Century communication metaphors which do not fit well in the internet-age of the 21st Century. It is not hard to imagine when the vast majority of archaeological enquiry is conducted and communicated *in-silico* (Wikipedia, 2012e). At this time it will be easier for all data creators to deposit digital evidence and synthesis as part of business and academic best practice. However, the issue here is not about the technology but about the social, organisational and policy issues that underpin modes of practice. The policy and statutory guidelines indicate that digital content should be openly shared and this can be easily facilitated through a domain, or non-domain, data repository. This is not happening. Why? The technical barriers are not a problem: Figshare (2012) and 'the Data Hub' (2012) provide free, or low cost, hosting environment. Do archaeologists just feel uncomfortable making their fine grained data available to a mass audience without going through an organisation with an authoritative reputation such as the ADS? Is there an informal belief that if data is deposited with a repository then the repository also takes the ethical responsibility if the data is released and used abusively? These social and operational problems require addressing. It is critical that current flaws in the deposition and dissemination of mixed analogue and digital archaeological content do not become the norm when archaeology processes are predominantly digital.

Whatever the answer the point remains: archaeologists, for right or wrong, consider the implications of placing fine grained digital data in the public domain and downstream abuse been identified as a 'barrier' to deposition. However, there appears to be limited guidance as to how to resolve these issues. This means that many archaeologists are re-inventing the wheel. What is missing is co-ordination. There is a need to produce guidance material approved by national heritage organisations and standards bodies that provides clarity on the ethical issues, responsibilities and pragmatics of depositing digital content. Improving access to digital data which can easily be re-used is essential to improve understanding and governance. This debate is too important to be based solely on perceived self interest.

The DART Project: an example of Open Archaeology

The Detection of Archaeological residues using Remote Sensing Techniques (DART) project (www.dartproject.info) focuses on analysing factors that influence archaeological contrast dynamics. DART aims to determine how different remote sensing technologies detect contrast caused by different underlying factors under dynamic environmental conditions. This understanding will allow the optimal deployment of the different sensors. By combining the results from a battery of sensors, each optimally deployed when the archaeological residues have the greatest likelihood of being detected, the knowledge about the range and scope of extant archaeological remains can be maximised. To examine the complex problem of heritage detection DART has attracted a consortium consisting of 25 key heritage and industry organisations and academic consultants and researchers from the areas of computer vision, geophysics, remote sensing, knowledge engineering and soil science.

DART is a data rich project: in-situ soil moisture, soil temperature and weather data is collected at least once an hour, geophysical surveys and spectro-radiometry transects are conducted at least monthly, aerial surveys collecting hyperspectral, LiDAR and traditional oblique and vertical photographs are taken throughout the year and laboratory analyses and tests are conducted on both soil and plant samples. The data archive itself is in the order of terabytes. Communicating this data between the different teams is a challenge in its own right. In addition the data collected by DART is of relevance to a broad range of different communities. DART took the unusual decision to adopt an open-science position. DART has decided to forego the '...limited period of privileged use of the data they have collected to enable them to publish the results of their research'. However, it has had a concomitant impact on licence negotiating particularly for third party data providers[5]. Open Science was adopted with a twofold aim:

To maximise the research impact by placing the project data and the processing algorithms into the public domain as soon as was practicable.

To build a community of researchers and other end-user around the data so that collaboration, and by extension research value, can be enhanced.

From a practical point of view DART is implementing its own repository based upon the Open Source D-Space repository framework. Using WebDAV the distributed team members can share their data with the group on the server. The data is then checked and pre-processed. Python scripts are then used to bulk ingest the data into the repository with their metadata. This framework allows both the data and other research objects to be made available as downloadable objects with associated metadata. The metadata is used to both document the raw data (improving re-use) and to enhance its discovery (make it easier for people to find it). The metadata held within the repository can be harvested by external portals using the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). This means that external discovery portals can be used to automatically search the archives held within the DART data repository. The majority of material is made available as Open Data Commons By Attribution licences (http://opendefinition.org/licenses/odc-by/), for data, and Creative Commons By Attribution licences (CC-By: http://opendefinition.org/licenses/cc-by/), for everything else. In respect of data the repository is designed to hold the data in its earliest incarnation, i.e. as close to raw sensor readings as possible, but to also make available derivatives of the data which are easier for others to use. For example some data comes in binary formats which require proprietary tools, such as some of the logger data, or requires extensive pre-processing to facilitate other analysis, such as the EAGLE and HAWK hyperspectral data provided by NERC. In addition to the repository the team are developing an environment which will articulate all the data in an integrated manner and expose these as data services (however, the public release of this facet is likely to be delayed). The aim is to make it easier for the research team, and others, to data mine the resources in order to identify changes in feature contrast and link these through to environmental or land-management processes.

The data held within the repository and analysis environment has enormous research potential for many different areas for years to come. In addition, it has educational potential: many archaeologists find it difficult to access these types of baseline resources especially when it comes to hyperspectral data and the supporting ground spectro-radiometry readings. DART will be producing education packs and other teaching and learning resources to build on the rich data.

However, the approaches adopted by DART have raised a number of issues. It has become clear that the term 'Open Science' is interpreted differently in the different disciplinary bases. The different

---

[5] All third part data providers have agreed to the open licence and re-use conditions favoured by the project. NERC Airborne Research and Survey Facility have also been told to not embargoe the data they collect for the project.

interpretations arise from the term "as soon as practicable"' – from an engineering perspective it could be argued that it would only be practicable once the IP has been secured (were it development of a widget being aimed for) or intellectual IP has been protected. But in other domains, researchers regularly release, for example, software prototypes for re-use without restriction, at least for non-commercial use, as soon as the prototype has been completed. Moreover, recent experience of one member of the consortium serving on the editorial board of the premier journal in geotechnical engineering has revealed that if interpreted data have been published in the public domain, they are not eligible to be published in the journal. This naturally gives rise to a concern for the DART geotechnical researchers that they may be precluded from publishing their findings in journals if they release data on the web openly. This introduces further concerns about the nature of the PhD research process and if, in some domain, Open Science conflicts with undertaking novel research. The other disciplines represented in DART do not share these concerns, as they are confident that simply releasing data does not in itself preclude subsequent journal publication or research novelty. The precise position on the issue of how these traditional engineering journals accommodate 'Open Science' is being explored.

A second issue that differs across disciplines concerns the different interpretations placed on the ideas of data, information and ultimately rigorous data sets. Creating experimental equipment and taking readings does not mean that the outcomes are what are intended. Considerable refinement of the equipment and the methods of interpretation (i.e. via modelling) is typically needed before confidence can be placed in the outcomes, i.e. before the data are known to be accurate. Different positions can be taken as to whether it is sensible to release such unverified data or not. Once more these issues are being explored within the project.

Even with these caveats and the fact that the tools and techniques are still in development, the initial results have been encouraging. Different communities at the Royal Agricultural College in Cirencester are using data from DART for a variety of activities including mapping and examining carbon sequestration in field boundaries. Our collaborative network has extended: Other researchers have offered to analyse the full wave-form LiDAR data collected by NERC on the basis that analysing it in conjunction with the associated environmental data should provide better insights into contrast detection. This enhances the skills portfolio of the researchers dramatically. DART is providing open re-use teaching and learning packs for the EU wide ArcLandscapes project, other UK universities and a specific training school ran in Poznan in July 2012 which provided many archaeological researchers with their first opportunity of using hyperspectral data with supporting ground survey. Because of the 'open' position and licences adopted and negotiated by DART these students were provided with data and software they could take home, analyse and publish. This is a significant step forward in knowledge transfer.

Conclusion: moving toward more open and dynamic archaeology

The archaeological knowledge base should be, by definition, dynamic: It is predicated on the complex relationship between the corpus of knowledge, theory and classification systems. These relationships are fluid and contain many interlinked dependencies which means that variations in one constituent part can have complex repercussions. Conceptually, this can be used to more accurate model the past than any number of decoupled and generalised HEIRs. A better understanding of archaeological hermeneutics will occur as stakeholders become more used to dealing with a dynamic archaeological corpus. The interpretative interplay between theory, practice and data as part of a dynamic knowledge system will be re-established; theory will influence practice, which will change the nature of the data, which will impact on interpretative frameworks, which will provide a body of knowledge against which theory can be tested. This may provide the opportunity to question the orthodoxy of excavation and interpretative practice as data can be used to dynamically test hypothesis demanding more question focussed, rather than formulaic, practice.

The point of Open Archaeology is to have a transparently accessible knowledge base that can be used for many different scales of enquiry by many different audiences. The ability to turn these data into knowledge for a variety of different communities will be transformative and lead to greater, and sometimes unanticipated, impact. This will not only change the way we engage with, research into and manage our shared heritage but on the organisations that mediate engagement. From a cultural perspective we need to explore the different traditions and approaches to data access and determine what barriers exist to opening data and what incentives, or legislation, is required to improve data access. This is not just about how we publish, cite and integrate data (and how contributors can gain professional credit) but also about who owns data and how owners can or can not influence down-stream exploitation (and the ownership of downstream derivatives). There will remain areas where data, information and knowledge remain inaccessible, however, the reasons need to be made clear. It is wrong to silo data on the basis of dogma and worse when this is done in organisations that are publically funded. The benefit should not be calculated on how much the data can be sold for but rather based on the improvements in policy, research and public engagement. Improving access to data will improve the knowledge base which improves research and governance: think provision, not possession (Isaksen, 2009). As stated by English Heritage:

Knowledge is the prerequisite to caring for England's historic environment. From knowledge flows understanding and from understanding flows an appreciation of value, sound and timely decision-making, and informed and intelligent action. Knowledge enriches enjoyment and underpins the processes of change(English Heritage, 2005)

If knowledge is open everyone benefits.

Acknowledgments

References

Ancient Lives, 2012. Ancient Lives | Help us to Transcribe Papyri [WWW Document]. URL http://ancientlives.org/

Beck, A.R., In Press. The Practice of Collaboration, in: Opitz, R., Cowley, D.C. (Eds.), Interpreting Archaeological Topography – Airborne Laser Scanning, 3D Data and Ground Observation.

Bradley, R., 2006. Bridging the Two Cultures – Commercial Archaeology and the Study of Prehistoric Britain. The Antiquaries Journal 86, 1–13.

Cooper, S., Khatib, F., Treuille, A., Barbero, J., Lee, J., Beenen, M., Leaver-Fay, A., Baker, D., Popović, Z., players, F., 2010. Predicting protein structures with a multiplayer online game. Nature 466, 756–760.

Department for Communities and Local Government, 2012. National Planning Policy Framework (Publication (Legislation and policy)).

English Heritage, 1995. Research Frameworks: project report (3rd   draft). London.

English Heritage, 2005. Making the past part of our future.

Figshare, 2012. figshare - credit for all your research [WWW Document]. URL http://figshare.com/

Ford, M., 2010. Archaeology: Hidden treasure. Nature News 464, 826–827.

Galaxy Zoo, 2012. Galaxy Zoo: Hubble [WWW Document]. URL http://www.galaxyzoo.org/

Haklay, M., 2011. Extreme Citizen Science – ExCiteS [WWW Document]. Po Ve Sham - Muki Haklay's personal blog. URL http://povesham.wordpress.com/2011/03/07/extreme-citizen-science-excites/

Hand, E., 2010. Citizen science: People power. Nature News 466, 685–687.

Hardman, C., 2009. The Online Access to the Index of Archaeological Investigations (OASIS) Project: Facilitating Access to Archaeological Grey Literature in England and Scotland. The Grey Journal Volume 5.

HM Government, 2012. data.gov.uk: opening up Government [WWW Document]. URL http://data.gov.uk/

Isaksen, L., 2009. Pandora's Box: The Future of Cultural Heritage on the World Wide Web.

Johnson, S., 2011. Where good ideas come from : the seven patterns of innovation. Penguin,, London :

Linked Data, 2012. Linked Data | Linked Data - Connect Distributed Data across the Web [WWW Document]. URL http://linkeddata.org/

MORI, 2000. The Role of Scientists in Public Debate. Wellcome Trust.

myExperiment, 2012. myExperiment [WWW Document]. URL http://www.myexperiment.org/

Nielsen, M.A., 2011. Reinventing discovery : the new era of networked science. Princeton University Press, Princeton, N.J.

Old Weather, 2012. Old Weather - Our Weather's Past, the Climate's Future [WWW Document]. URL http://www.oldweather.org/

Oxford Archaeology, 2009. Nighthawks & Nighthawking: Damage to Archaeological Sites in the UK & Crown Dependencies caused by Illegal Searching & Removal of Antiquities (No. OA Job No: 3336). Oxford Archaeology.

PAS, 2012. Welcome to the Portable Antiquities Scheme website [WWW Document]. URL http://finds.org.uk/

Poliakoff, E., Webb, T.L., 2007. What Factors Predict Scientists' Intentions to Participate in Public Engagement of Science Activities? Science Communication 29, 242–263.

RCUK, 2012a. Welcome to Research Councils UK - RCUK [WWW Document]. URL http://www.rcuk.ac.uk/pages/home.aspx

RCUK, 2012b. RCUK Common Principles on Data Policy - RCUK [WWW Document]. URL http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx

Royal Society, 2012. Science as an open enterprise. The Royal Society.

Royal Society, RCUK, Wellcome Trust, 2006. Science Communication: Survey of factors affecting science communication by scientists and engineers.

Taverna, 2012. Taverna - open source and domain independent Workflow Management System [WWW Document]. URL http://www.taverna.org.uk/

The Data Hub, 2012. Welcome - the Data Hub [WWW Document]. URL http://thedatahub.org/

Townsend, A., Soojung-Kim Pang, A., Weddle, R., 2009. Future Knowledge Ecosystems: The Next Twenty Years of Technology-Led Economic Development (No. IFTF Report Number SR-1236).

Wellcome Trust, 2012. Wellcome Trust strengthens its open access policy [WWW Document]. URL http://www.wellcome.ac.uk/News/Media-office/Press-releases/2012/WTVM055745.htm

Wenger, E., McDermott, R.A., Snyder, W., 2002. Cultivating communities of practice: a guide to managing knowledge. Harvard Business School Press,, Boston, Mass. :

Wiggins, A., Crowston, K., 2011. From Conservation to Crowdsourcing: A Typology of Citizen Science, in: System Sciences (HICSS), 2011 44th Hawaii International Conference On. pp. 1 –10.

Wikipedia, 2012a. Open research [WWW Document]. URL http://en.wikipedia.org/wiki/Open_research

Wikipedia, 2012b. Citizen science [WWW Document]. URL http://en.wikipedia.org/wiki/Citizen_science

Wikipedia, 2012c. Crowdsourcing [WWW Document]. URL
        http://en.wikipedia.org/wiki/Crowdsourcing
Wikipedia, 2012d. Archaeological ethics [WWW Document]. URL
        http://en.wikipedia.org/wiki/Archaeological_ethics
Wikipedia, 2012e. In silico - Wikipedia, the free encyclopedia [WWW Document]. URL
        http://en.wikipedia.org/wiki/In_silico