FAIMS Stocktaking Workshop 16-19 August 2012, UNSW

Date: 23 August, 2012

Submitted by: Adéla Sobotková, Shawn Ross, Penny Crook, Brian Ballsun-Stanton

1 Workshop Report

1.1 Purpose

The Stocktaking Workshop was organised to solicit input about development directions for the FAIMS project and to inform decisions about the structure of the project, technologies, modules to be supported. It was also convened to gauge the level of convergence on data standards across academia, industry and state agencies as well as across different archaeological subdisciplines (survey, excavation, artefact recording and sciences). The workshop also served to bring together the members of Steering Committee.

Originally the workshop was envisioned to have 35-40 participants with 10 international participants discussing over two days. The response to the initial call for papers, however, was overwhelming and the duration of workshop was extended to four days from Thursday 16 August to Sunday 19 August. A webpage for the workshop was created using OCS and linked to the FAIMS webpage (see: www.fedarch.org/ocs). Registration, as well as the communication and information related to the workshop was channeled through this webpage. The plenary session abstracts and workgroup objectives are available in the attachment to this report.

1.2 Attendance

The workshop attracted 80 attendees from NSW, Victoria, ACT, Queensland, SA, and WA as well as overseas. The numbers stayed high on Thursday and Friday (16-17 August), and dropped gradually over the weekend. The least attended session on Sunday had 40 participants. Participants included members of archaeological community, both academic and public, industry and state agency officials as well as representatives from each of the FAIMS developer organisations and NeCTAR.

1.3 Workshop Structure

The workshop was structured as a stocktaking enterprise with plenary sessions driving the agenda of smaller afternoon discussion workgroups. The Plenary session speakers presented successful initiatives in digital archaeology and broadened the horizons of the FAIMS debates. Plenary sessions were recorded and filmed.

After the inspiration provided by the presentation overseas and domestic case studies, the FAIMS stakeholders broke into small workgroups, where they discussed recordkeeping standards, requirements for the mobile systems, federation strategies and other specific issues that would inform the FAIMS project development. The groups addressed Archaeological Survey, Excavation, Artefacts (split into Ceramic and Lithic recording), Sciences, Federation, Sustainability and Sensitive Data. Each group had the opportunity to meet multiple times during the workshop.

Each of the groups was assigned a leader who was responsible for ensuring that the results of the discussion were reported back daily during evening wrap-up plenary session. The leader also received an information package that provided some background and outlined goals for each group. Individual

workgroup participants were encouraged to post the minutes of their discussions on Google Groups set up for that purpose by FAIMS. Wi-fi connection was provided free of charge for all participants at the Workshop. Several of the groups were supplemented by a separate focus group, facilitated by an experienced focus group leader. These focus groups were recorded and are currently being transcribed and analysed to supplement the information gathered in other fora including the project survey and workgroup discussions.

1.4 Workgroup Purpose and Outcomes

1.4.1 Purpose

The working groups were intended to generate the basis for the requirements of the FAIMS project. Each group was assigned a specific topic, with a number of questions to guide their discussion. Group membership was on an ad-hoc basis, with group interaction continuing after the conference through a group-specific mailing list (groups.fedarch.org).

1.4.2 Outcomes

The notes below are preliminary, derived from the working group notes submitted to us. They do not yet include focus group transcripts, or further reflections by group members. Fieldwork groups (artefacts, excavation, survey) also provided some mapped vocabularies and minimum threshold requirements for recording, which have not been reproduced here (these vocabularies and requirements will be made available on the website shortly).

1.4.2.1 Artefact group

In their report, the artefact group suggested:

- modular design that would accommodate regional and typological differences
- input systems based on recognised, existing standards and practices (producing a library of possible input systems)
- recording based on artefact recording workflow
- smooth image handling
- easy customisation
- use of controlled vocabularies with aliases
- comprehensive management of metadata
- maximal automation

- versioning and tracking
- recording of uncertainty (either binary, as a flag, or scaled)

In addition, they recommended that any system be able to manage both group and individual records, produce rudimentary statistics in the field, provide search and query in the field, and connect to a range of external data recording devices (present and near future). Concerning vocabularies and data standards, the Artefact group requested a discussion forum to discuss properties of each attribute.

1.4.2.2 Excavation:

In their report, the excavation group reiterated some of the suggestions of the artefact group, and added:

- excavation unit needs to be flexible and adjustable (shovel test to stratigraphy)
- spatial recording (simultaneous recording of xyz coordinates for single finds and 3D excavation units)
- incorporate vector sketches and raster imagery (photographs, site plans, drawings, etc.)
- allow audio-recording of voice notes (not necessarily with text-to-speech capability)

1.4.2.3 Survey

In their report, the survey group reiterated some of the suggestions of the other groups, and added:

- Globally unique identifiers for all fields, with local human-readable aliases
- work across up to 5 layers of devices: mobile, in field server, lab server, research server, repository
- synchronisation must be invisible instantaneous, transparent and seamless, and must allow versioning

1.4.2.4 Federation / technical summary

It is critical that all developed recording applications be modular. The applications cannot hard code any records or means of achieving those records. While there is no opposition to an AR application, little thought has been devoted to how it can be used. Overall it is difficult for the archaeological community to break from normal "everything's a paper chart" methodologies. We must not feel constrained by existing systems and approaches and technologies. We need to think to future developments like Google Glass and Augmented Reality devices, at least for inspiration. VeRSI and Brian propose the development and demonstration of a "killer app" for this purpose.

Given modular data requirements, the mobile data standard established was essentially that of an XML templating engine. The application should be able to generate input systems based on XML

schema. Furthermore the schema should provide for the ability to interface with external data-collection devices, such that it can collate and absorb the outputs of the external devices. Non-relational data architectures could serve the community extremely well, as the current relational structure require hacks to get around the limitations of relational databases, hobbling data collection and analysis efforts.

Applications should collect an internal-audit log (engaging cameras, acceleration, position, gait analysis, etc.) to indicate who the data-collector is and as much of their mental/physical state as possible. The program should also have a highly accurate place-time record, as well as linking that to visual records of the site and any appropriate events.

Much User-Centred Design and Scenario Testing will be needed to refine these requirements.

ADS, tDAR, OpenContext, Heurist, and Kora should be considered for the first round of interoperability.

1.4.2.5 Sustainability and Sensitive Data

We are still awaiting reports from the Sensitive Data and Sustainability Workgroups.

In brief, the Sensitive Data workgroup agreed that technical solutions to the sensitive data are fairly straightforward. It explored different ways of handling Sensitive data beyond merely restricting access, hiding or ambiguating location. One of its main outcomes focused on facilitating access to meaningful derived information, while hiding actual sensitive data.

Sustainability group suggested a wide range of approaches to funding the FAIMS project, combining user-driven approaches to continued seeking of grant funds.

1.5 Workshop conclusions

The principal purpose of this report was to outline best practice(s) in record-keeping and data and metadata standards. What we learnt from the workshop, however, is that on the one hand, a range of standards will need to be supported, but on the other hand, there is common ground across this range of standards. We plan to offer modules that are based on existing published standards and current fieldwork best practices. As a result, what we have clarified (and present below) is a strategy or an approach rather than a list of standards. The lists of standards that will form the basis of our mobile applications and federation efforts will need to be derived from publications and current fieldwork practices, the research of which the FAIMS team will undertake over the course of the next month. We will build a system that can accommodate a range of standards and will develop examples for a select few.

- Overall, the high-level overall features seen as important by participants included:
- All developed systems and applications need to be modular
- Rather than an extensive core data collection system we need a library of data collection tools that are based on existing published standards and existing paper recording systems. Each of these

needs be modifiable, extensible and shareable. To the extent possible, these tools need to be mapped against one another (possible because of similarities in vocabularies and ontologies).

- Controlled vocabularies with aliases need to be implemented throughout these systems
- Field–available reference collections and spatially aware applications are in high demand and should be provided. Support, training, and tool tips need to exist for the designed tool features.
- Unique identifiers for each piece of data collected need to be generated with human-readable aliases.
- While we recognise that existing state registry requirements need to be taken into consideration, we have been encouraged by the SC and others to design comprehensive high quality recording systems and promote these to the state registries, mining companies and other stakeholders.
- Systems developed need to accommodate existing and future-data collecting devices (total stations, dGPS, digital calipers, etc.). Devices used in field and laboratory need to sync seamlessly and transparently.
- Metadata needs to be comprehensively collected and managed. As much as possible it needs to be collected automatically and derived from existing system components (eg. Login, Timestamps, reference collections).
- In all cases, data creation needs to follow workflows with options to adjust for variations in workflow.
- We were encouraged to find demonstration data sets in various realms and pursue these as a greater proof of concept for the project
- Australian implementations of a repository (probably tDAR) and a dataset publication service (probably OpenContext) are envisioned.
- Cloud-based implementations of data manipulation and analysis tools are also envisioned (likely Heurist and Kora).

1.6 Workshop Feedback

So far 18 people have responded to the online Evaluation of the workshop. The overall feedback was very positive. Of the current respondents 53% are likely to be more involved in the FAIMS project, while the remainder 47% are just as likely to remain involved as they had been so far.

	Very valuable	Valuable	Not very valuable	N/A	Total
Plenary Sessions	14	3	0	0	17
Discussion	10	3	1	3	17
Focus Groups	6	1	2	7	16
Social Events	12	4	0	1	17

1.6.1 Selected participant comments:

"Said it above: wish I'd brought some colleagues down! What a rich resource. I think that the workshop has brought out a number of issues, at all scales. I wonder if there is a useful publication to be developed from it already???"

"I found it quite valuable personally, as I have not had much experience in this sort of area (other than the databases I use for my personal research). It was wonderful to have international guests as well, I think they were critical. I am excited to see what is adopted and how the FAIMS project goes forward. Thank you so much for inviting me and for doing such a great job on putting the workshop together."

"A big congratulations to Shawn, Adela, Brian and the volunteer team for an exceptionally well run and executed workshop. Well done."

"It was fabulous. Adela and Shawn, you are to congratulated for dreaming it up, organizing it and working so hard to make everything useful and keep everyone happy."

"A very valuable exercise. Congratulations to Shawn, Adela, Brian and Penny for pulling off such a successful event in a short timeframe."

"There is no substitute for bringing people together face-to-face. This is particularly critical at the commencement of a major project such as FAIMS where international collaborators will be working together for several years mainly using asynchronous media to discuss the project."

"Overall the FAIMS workshop was well organised and run. The diversity of archaeological sub-disciplines represented was generally a strength. Also I was appreciative of the sensitivity that generally was shown, in recognising that our workgroup discussions would not always be easy."

2 Abstracts

2.1 Keynote

2.1.1 Capacity Building in (Digital) Archaeology

Kansa, E. (UC Berkeley)

Abstract: Twenty-first century archaeology must make data preservation and dissemination a regular partof its workflows. Better data dissemination can promote analytic rigor and transparency, reduce inefficiencies and duplication of effort, and open new research opportunities for larger scale and multidisciplinary inquiry. At a time of cutting fiscal austerity, such efforts can reduce costs and expand the equity and effectiveness of archaeological communication.

While programs in archaeological data preservation and access continue to attract a great deal of excitement and funding, they face steep challenges. Long-term sustainability is a major concern, as are issues in interoperability, semantics, data longevity, and professional incentives. In order to face these challenges, we need to better understand the emerging social, professional and technical context of digital archaeology. One of the most important features of this context is the Web. The rise of the World Wide Web represents one of the most significant transitions in communications since the printing press or even since the origins of writing. To many researchers, the Web offers great opportunities for expanding the accessibility, scale, diversity, and quality of archaeological communications both for current research and, through data preservation efforts, for future generations. Understanding the Web, and how it interfaces with the changing and more fluid nature of professional life, is a vital first step in securing a vibrant future for the digital past.

2.2 Mobile Applications

Respondents: Rabinowitz, A.; Watrall, E.; Wilson, A.

We would like to have the speakers present your experience with mobile applications, particularly where you drew the line in what was recorded digitally and what was recorded on paper, and what made you decide on one type of data creation. For example, in our own presentation we will present a couple of vignettes from our own experience with customized ArcPAD survey and remote sensing application (deployed on PDAs).

Second, give us a vision of your ideal tool. What capabilities do you need to see to shift the line so that more data is created digitally?

2.2.1 Mobile GIS for Excavations in Andean Archaeology: The Good, Bad, and the Ugly

Wernke, S. (Vanderbilt University)

Abstract: Given the relative underdevelopment of 3D GIS and the complexities of excavation data, mobile GIS for archaeological excavations remains a relatively underdeveloped application domain, but for the same reasons, it is also the domain of archaeological research in which an integrated information management system would be most beneficial. This talk reviews our efforts to implement such systems in the highland Andes using modified commercial software products (ESRI ArcPad and Garafa GISPro) on different hardware systems (PocketPCs and iPads), with mixed results. Based on these experiences, it is argued that a successful FAIMS outcome will require at least five essential characteristics:

1. a hardware and software interface that minimizes the impedances of human/machine interaction (and thus, minimizes the friction of data registry),

2. a flexible and field-editable relational database design that can handle a wide variety of data types,

3. an imaging subsystem that permits in-field image geo-/ortho-rectification,

4. direct communication with total stations and other field mapping instrumentation, and

5. a datasynchronization system that can be easily maintained by project managers.

2.2.2 Keeping paper or going digital? Optimising the record-keeping during archaeological field survey

Sobotkova, A. (UNSW)
Ross, S. (UNSW)

Abstract: This presentation illustrates two record-keeping strategies deployed at the Tundzha Regional Archaeological Project, a regional surface survey in Bulgaria. To reduce the time devoted to record-keeping and improve accuracy, our project tested real-time digital record keeping on GIS-linked PDAs (using customised instances of ESRI ArcPad). The use of PDAs proved highly efficient during legacy data verification and ground truthing of satellite remote sensing results. During these tasks a team of two traveled between legacy sites or suspect SRS features, and created or verified and updated the entire record digitally. During full-coverage pedestrian survey with a team of five or more, long processing times on the PDA caused considerable delays to survey progress; efficiency considerations required the continued use of some paper documentation to complement the digital record.

2.2.3 Mobile Data Capture Tool Demonstration

Scriffignano, J. (Dynamic Spatial Solutions)

Abstract: A demonstration of mobile data capture tools for cultural heritage management focusing on survey management, site recording and management, test pitting & AHIMS site card extraction.

These tools have been developed for use in ArcPAD and ArcGIS in a single user environment with the purpose of supporting the greater cultural heritage management process and agreements. These tools are currently being redeveloped to reside within a multiuser environment based on an enterprise SQL Server geodatabase supporting versioning.

2.3 Online Repositories

Respondents: Wells, J.

We ask panel members to define and explain the different strategies for designing and building a repository, and the different repositories themselves e.g. what are the primary differences between tDAR, Open Context, ADS, AKB, Kora? Why were they designed this way? Whose needs to they meet?

In our context of archaeological data storage, what is curation?

How do we encourage data reusability? What is open-linked data and how do we produce it Is open linked archaeological data unique?

How can we encourage convergence between repositories for cultural heritage management and research?

In five minutes, what would the ideal future repository look like?

2.3.1 Unpacking the definition of a "digital repository"

Brin, A. (Arizona State University)

Abstract: A short discussion about what it means to be a repository online, unpacking the definition and requirements of a "digital repository".

2.3.2 Online repositories: the Archaeology Data Service 15 years on

Richards, J. (Archaeology Data Service)

Abstract: The UK's Archaeology Data Service is the longest standing discipline-based archive in the sector, and recently enjoyed its 15th birthday. The ADS was established in 1996 as one of the five discipline-based service providers making up the UK Arts and Humanities Data Service (AHDS). It is hosted by the University of York. It began with two members of staff, but now has 15. The ADS is the mandated place of deposit for archaeological research data for a number of research councils and heritage organisations and makes all its holdings freely available for download or online research. At the last count it provides access to over 17,000 unpublished fieldwork reports (the so-called grey literature) and over 500 data-rich digital archives. All reports and archives are allocated a permanent means of citation, or Digital Object Identifier, in collaboration with the British Library and Datacite.

In 2011 the ADS was awarded the Data Seal of Approval, a kite-mark for trusted digital repositories, making it the second UK repository to receive this accreditation, after the UK Data Archive.

2.3.3 AKB-National Registry of Archaeological Sites in Bulgaria

Kecheva, N. (National Institute of Archaeology and Museum, Bulgarian Academy of Sciences)

Abstract: A short presentation about the structure and purpose of the Archaeological Map of Bulgaria.

2.3.4 Digital Repositories from the Perspective of a Web-based Consumer of Repository Services

Kansa, E. (UC Berkeley)

Abstract: Online repositories tend to be oriented toward services for researchers managing largely offline data. This emphasis is appropriate and justifiable, since most of the research community does not manage their own data on the Web. In contrast, Open Context is a Web-based data dissemination system with a different set of repository needs. As a publisher of data, these needs include branding and identity, persistent identifiers for citation, and the need to version and archive dynamic data. Open Context is not alone in these needs. As more scholarship migrates online and as more researchers adopt the Web as their primary dissemination platform, digital repositories will need to expand their range of services to meet these new needs.

2.3.5 KORA: Open Source Digital Repository Platform for Cultural Heritage

Watrall, E. (Department of Anthropology MATRIX: Center for the Arts, Letters, and Social Sciences Online Michigan State University)

Abstract: There exists a startlingly wide variety of options for archaeologists and cultural heritage professionals seeking software and tools to store, preserve, and share digital archaeological data. The options range from commercial to non-commercial, proprietary to open source, and centralized to decentralized. It is within this context that this talk will introduce and discuss KORA, an open source digital repository platform originally designed to store, share, preserve, publish, and distribute cultural heritage content. Developed by Michigan State University's MATRIX: The Center for the Arts, Letters, and Social Sciences Online, KORA has been used in a wide variety of cultural heritage digital library, archive, and repository projects over the past 10 years. This talk will introduce KORA as an ideal tool to store, preserve, publish, and share archaeological and cultural heritage data and content across the spectrum of archaeological settings. Ultimately, the goal of this talk is not to suggest KORA as the "one ring" of digital repository tools for archaeological content and data, quite the contrary. Instead, the goal is to place KORA within a rich diverse ecosystem of standards-based tools for archaeological data storage, preservation, sharing, and publication.

2.4 Archaeological Data and Data Standards

Respondents: Burke, H

FAIMS seeks to explore the possibilities of data federation in archaeology. Federation is defined as automated data interoperability requiring a minimum of manual intervention as opposed to, e.g., loose-coupling, which according to Kansa and Bissel's (2010) definition relies on extensive human interpretation to reconcile datasets. The degree of federation depends on the level to which semantic and technical interoperability can be achieved.

This panel is about semantic interoperability, while the next panel will be about technical interoperability. We wish to focus on the problem of producing semantically comparable data. For example, are shared record-keeping standards, as well as technical data standards, necessary for the success of this project? If so, can we agree on minimum record-keeping protocols?

The argument has been made that every project has its own research agenda, requiring a unique record-keeping strategy (vocabulary and taxonomy) customised to its goals. If that is the case, is it counterproductive to develop a master record-keeping protocol ?

If minimum record-keeping protocols are desirable, do you think that we can agree on, e.g., 10 things that need to be recorded about any excavation stratigraphic unit, surface survey unit, artifact, etc.

Is federation possible without shared record-keeping protocols? Is federation a worthwhile goal? Is the best achievable goal the loose coupling described by Kansa and Bissel (2010, 44)?

To what extent is it possible and desirable to separate quantifiable ("objective") and interpretive ("subjective") data? Is there a fundamental difference between a date of a lithic and a length of a lithic?

If we do have standards, are they regional, chronological, global? Can there be a shared set of standards between cultural heritage and research, or archaeology and related disciplines?

2.4.1 Artefacts and social identity

Allison, P. (University of Leicester)

Abstract: This presentation will outline my application of interpretative categories to artefacts in on-line databases. It will involve the presentation of the different processes I have used for my Pompeian data (on the Stoa Consortium) and also my `Engendering Roman Military Space' project (on ADS). It will also discuss the user-friendliness of these processes and questions of standardization.

2.4.2 Comparing artefact catalogues in Historical Archaeology

Crook, P. (La Trobe University)

Abstract: Artefact catalogues and data sets in historical archaeology are renowned for their inconsistency. Every excavation site seems to present a new group of artefact types, patterns or manufacturers to be classified, defined and organised and schemas are adapted to suit. Despite this, even highly variable datasets have a good degree of structural consistency; the critical differences arise in their

companion vocabularies and data management. This will be discussed in reference to a direct comparison of records of 'Willow' plates from available 20 available catalogues.

2.4.3 Reflections on Archaeological Data and Data Standards in Australia, Melanesia, the Pacific and the Mesoamerica

Ulm, S. (James Cook University)

Abstract: Drawing on fieldwork in Australia, Papua New Guinea, the Pacific and Mesoamerica I draw attention to similarities and differences in the ways that archaeology is conducted to identify obstacles and opportunities for federating archaeological data across subdisciplines and regional archaeological traditions. I will also use examples from the 'Index of Dates from Archaeological Sites in Queensland' and 'Queensland Historical Atlas' to discuss compromises between ideal and practical outcomes for recommending data standards.

2.4.4 Build it and they might come.

Johnson, I. (University of Sydney)

Abstract: I will explore the idea that standards can be encouraged by providing a path of least resistance, and that advances in technology are instrumental in the building of such paths. Standards often get adopted because there is a good tool which encapsulates them (thanks to digital cameras we all use EXIF and JPG) or because a standard becomes well known and saves time and effort over reinvention. Increasingly standards can be adopted without much effort because they are built-in to tools or available as plug-ins (sensu lato); where ad hoc solutions used to be the path of least resistance, they are increasingly looking like the hard option.

Whether one veers towards lockdown standards or loose coupling, the fragmentation of standards complicates the work of integration. I think we can be optimistic that by building tools which encapsulate good data structures and standards, and by promulgating wider awareness of those tools and standards, we can go a long way towards facilitating data federation.

2.4.5 What is Data: some Different Philosophies of Data

Ballsun-Stanton, B. (UNSW)

Abstract: So many technical fields presume that their "definition" of data is the only one that exists. In my research, both within industry and the realms of archaeological data, I have found significant philosophical differences in the way data is concieved and used.

Much of the argument about data standards becomes more luminous when we realise that the very thing we are debating about has different representations in the debaters' minds.

This presentation will explore the three philosophies of data I have found in the course of my research, and some thoughts of how they are represented in Archaeological data.

2.5 Opportunities and Issues of Data Federation

Respondents: Ballsun-Stanton, B.

There are many sources of archaeological data. There are dozens of online repositories (mostly "silos" - isolated datasets difficult to reuse or repurpose). Very few of these sources and repositories are able to communicate with each other. A primary goal of the FAIMS Project is to provide federation capabilities (data interoperability) to these various projects across different standards. We do not aim to create our own "standard." (XKCD #927). Instead, we are interested in supporting a range of standards widely used by the archaeological community.

This panel will discuss the following issues: Which standards should FAIMS support? What shouldn't it support?

First, should we support formal ontologies like CIDOC-CRM? If so, how? To what extent are comprehensive standards such as CIDOC-CRM useful for federation? Should we support OCHRE's more generalised core ontology? If so, how? What level of abstraction is the most useful?

Second, let's consider approaches to interoperability. Kansa and Bissel (2010) have argued for "simple web" strategies. What is the minimum level of simplicity possible which still supports federation?

What data cannot be federated and why?

2.5.1 Data Federation and Interoperability in Europe

Richards, J. (Archaeology Data Service)

Abstract: This presentation will outline the development of interoperability in Europe. It will describe three approaches. Each depends upon well defined ontologies and controlled vocabulary lists, but varies according to the degree to which individual research projects must follow them. The first approach depends on a close degree of semantic interoperability and adherence to core data standards. An example is English Heritage's "Heritage Gateway" which provides rich interoperability between heritage data providers, who all adhere to the core UK MIDAS data standard. The second approach provides a looser level of interoperability, which may be useful for resource discovery, but is less effective for analysis. Data providers manually map their own ontologies and vocabularies to a common standard. The ADS ARENA and ARENA2 portals, as well as the ADS/tDAR Transatlantic Gateway provide examples of this approach. The third approach uses data mining and Natural Language Processing to map and link data sets via common vocabularies and ontologies, following a semantic web architecture. The ADS Archaeotools project, as well as the STAR and STELLAR projects provide examples of this approach, which has the potential for loose federation but deep analysis. The presentation will be complemented by a presentation in Plenary Session 4 on some of the technical approaches underpinning interoperability.

2.5.2 Bridging the gaps: Assessments and plans for American federation strategies and data standards to unite state-level archaeological databases

Wells, J. (Indiana University South Bend)
Yerka, S. (University of Tennessee)

Abstract: An American research group has just been awarded a grant from National Science Foundation to create interoperability models for archaeological site databases in the eastern United States (NSF #1216810 & #1217240). The core team consists of researchers from the University of Tennessee, the Alexandria Archive Institute, and Indiana University. Open Context will be used as the primary platform for data dissemination for this project. The project aims to work with the databases held by State Historic Preservation offices and allied federal and tribal agencies in Eastern North America, with the goal of developing protocols for their linkage across state lines for research and management purposes. Data from some 15 to 20 states (more than a half million sites) will be integrated and linked to promote extension and reuse by government personnel in state and federal agencies, and domestic and international researchers. This project is designed to involve datasets from numerous organizations, and testers from the professional archaeological community, in order to generate data products in the form of maps, tables, and analyses useful for primary research, cultural resources management, higher education, and public outreach.

2.5.3 Library interoperability and lessons learned

Brin, A. (Arizona State University)

Abstract: A short discussion about Library interoperability and lessons learned from 20+ years of libraries, networks, and modern industry.

2.5.4 Interoperability, Integration, and Distributed Data

Kansa, E. (UC Berkeley)

Abstract: Because archaeology is inherently interdisciplinary, data sources relevant to archaeological investigations will be distributed across the Web. Many data sources will be oriented toward different needs and applications. Some will be developed mainly for administrative purposes and others for exhibition and publication. Still others will be disciplinary repositories primarily meeting the needs of data preservation and research analysis. Researchers will also need to tap into the data resources curated by other disciplines.

Cyberinfrastructure efforts must accommodate needs around efficient discovery and use of distributed data. Pragmatic, Web-oriented approaches can better support researcher needs in this widely distributed landscape. Good practices in Web architecture, including use of simple and widely supported standards such as Atom can facilitate simple, yet useful, services for aggregation and discovery. "Linked Data" data approaches also have great promise to enable easier data discovery and easier aggregation of multiple data sources. However, the complexity and conceptual challenges inherent

in data integration make certain goals of the "Semantic Web", especially automated reasoning over large graphs of data, infeasible and probably ill-advised. Rather than expecting total semantic harmonization of all archaeological data, interoperability approaches should emphasize data discovery, portability, and more incremental and contestable approaches to data integration that are based on particular research needs and applications.

2.5.5 The fault is not in our databases, but in ourselves: messy data, metadata, and interoperability

Rabinowitz, A. (The University of Texas at Austin)

Abstract: Discussions of interoperability and data federation generally begin with the assumption that data are structured, and that the challenge in the creation of a new system or the adaptation of an old one involves the selection of standards and the attempt to integrate those standards with the standards used by other data-management systems. In a perfect world, perhaps, all archaeological data would be collected within such systems, and would be born fully equipped with metadata ready for federation. As we all recognize, however, the world is far from perfect, and there are a great number of datasets from long-running or small-scale projects that are organized in idiosyncratic ways and either lack metadata or do not conform to common metadata standards. Given will and money, some of these will eventually be ingested into one of the big archaeological repositories, but many of them will not. At the same time, however, some of these projects will still seek to make their datasets available online, or will find local institutional solutions for long-term preservation. There will always be such individual solutions, either because a project lacks the time and money to arrange for ingest into a central repository, or because the project directors are attached to a particular customized system. It is important, therefore, to consider interoperability not only in terms of centralized repositories, but also in terms of isolated and even disorganized individual datasets. This is a question that has already been addressed in several venues by Eric Kansa and his collaborators, and the solutions they have proposed offer new ways for repositories to deal with those datasets (e.g. the use of Atom feeds to automatically create new metadata from researcher queries of online datasets).

These solutions make it less burdensome for the creators of archaeological datasets to share their information, but they still rely on the intervention of a repository or programmers. Kansa and Bissell (2010) note that datasets without URIs for "individual units of observation" will be less suitable for integration. There is room, therefore, for tools that will allow some automatic creation of metadata to take place at the source – that is, at the level of an individual dataset as it is being collected – so that it is easier to provide URIs for individual resources at the repository level. Not only would this provide a richer field for querying for a dataset intended for ingest into and publication through an online repository, but it would also enhance the long-term reusability of digital datasets that are likely to end up warehoused on university servers for the foreseeable future. This raises two questions: first, is there a way to make it easy and cheap for these projects to create metadata and document their legacy datasets; and second, if this is possible, are there a few lowest-common-denominator metadata items that are likely to be common across most archaeological datasets, and therefore useful for aggregation, that could be captured for individual items or files with minimal effort and without explicit mapping to existing ontologies?

These questions in turn lead to another, larger question: what is it that we want from federated data, either as data producers or data consumers? The answer to the last question conditions the answers to the first two, and thus I will begin with some very subjective observations about the ways in which Classical archaeologists tend to use aggregated data, and the natural limits of that aggregation. I will suggest that most of our basic queries of archaeological documentation (either digital or paper) concern information about spatial and temporal coordinates and type of material documented, with a relatively small set of sub-queries within the latter, and that beyond this level it is very difficult to look at data across different datasets without a significant amount of manual alignment (if not fudging) to take into account different data-collection circumstances, even when the datasets themselves adhere to standard ontologies and metadata schemata. Furthermore, I suspect that most of the archaeologists searching for such information query text records most often to get to images or data tables, which allow reuse in ways less mediated by the terminology or interpretations of the databases involved.

If this is a valid premise, then it may be possible to create federation strategies that facilitate the documentation and reuse of idiosyncratic datasets even in the absence of a central repository or a custom API. The critical elements in the incorporation of such strategies in a new federated data-management system are, to my mind, agreement on a few basic types of information shared across ontologies and on their description (e.g. spatial location: cf. the use of Pleiades URIs in the aggregation of resources by the Pelagios project), and the creation of tools that make it cheap and easy for the managers of individual datasets to provide item-level metadata that includes such information – especially for files that are often associated with records rather than treated as items in their own right, such as images. A system that provides for such metadata creation could also help with the problem of the long tail of documentation that is produced after the conclusion of field research, which is often the most difficult to manage and the least likely to have extensive metadata, since it is organized primarily within the personal research frameworks of individual and dispersed specialists. I will conclude by discussing some preliminary work on some of these issues that has been carried out by the Institute of Classical Archaeology and the Texas Advanced Computing Center, using as its starting point a very large, idiosyncratic, and internally heterogeneous legacy dataset.

2.6 Sensitive Data

Respondents: Ross, S.

In your field of research what data is sensitive and why and how do you protect it? Sensitivity can cover a wide span of attributes from spatial to cultural, including notions of ownership and IP, authorship and publication.

What approaches and management strategies do you propose for sensitive data?

2.6.1 Managing Indigenous IP in the ARAM system

Ridges, M. (Office of Environment & Heritage (NSW) & University of New England)

Abstract: This talk will outline how Aboriginal IP was protected in the Aboriginal Regional Assessment module (ARAM) of the Aboriginal Heritage Information Management System (AHIMS), which is the

statutory site register in NSW maintained by the Office of Environment and Heritage. A key part of managing IP in this system was to develop a new mapping methodology called values-interests-priorities (VIP). This talk will briefly outline the VIP method and how it circumvents many IP issues that occur with Indigenous knowledge recording.

2.6.2 Rights Management and Confidential Data in AHAD/tDAR

Crook, P. (La Trobe University)

Abstract: I will briefly outline the management of confidential resources and metadata in the open-editing environment of Australian Historical Archaeology Database (AHAD) which is powered by Digital Archaeological Record (tDAR). I will touch on the following topics: tDAR's management of sensitive data and work in progress; the introduction of copyright and other rights attributions to AHAD; and the unintended consequences of site location obfuscation with respect to data exchange between AHAD and Research Data Australia (RDA).

2.6.3 Are digital technologies a good or bad thing for archaeology and cultural heritage practice?

Colley, S. (University of Sydney)

Abstract: The paper will briefly outline key principles enshrined in codes of ethics of professional archaeological organisations in light of ever changing impacts of digital technologies. Digital technologies extend long-standing ethical questions including, for example, maintenance of professional standards and how to balance intellectual, cultural property and other rights against the public 'right to know'. New ethical issues include e.g. sustainability; the role of public and private interests in producing, promoting, funding and maintaining widely used digital technologies and platforms; convergence of professional and 'community' practices in the digital sphere; economics and access to digital content and resources and digital literacy.

2.6.4 Indigenous Data and Digital Archives

Nakata, M. (UNSW)
Thorpe, K. (University of Technology Sydney)

Abstract: Prof Nakata and Ms Thorpe will provide a brief insight into some of the complexities around Indigenous knowledge, digital archives and online repositories, and will draw on experiences from their work in developing, maintaining, and resourcing an online data source for social scientists in Australia called the Aboriginal and Torres Strait Islander Digital Archives (ATSIDA).

2.7 Sustainability Strategies

Respondents: Johnson, I.

How can we make this project self-supporting in the long term? For those of you who currently manage repositories, what solutions do you have to this problem?

Based on your experience, which of the following revenue-raising strategies work and is generally acceptable?

- data curation fees (paid by researcher generating data)
- Subscription or access fee (user pays)
- customization fees (free open source base product, customization extra)
- budget line on grant applications
- others?

2.7.1 Funding the ADS

Richards, J. (Archaeology Data Service)

Abstract: The ADS has successfully undergone the transition from annual core grant from research councils to a business model based on a one-off deposit fee levied at the point of deposit. The ADS charging policy had evolved over the last 12 years and this presentation will describe that evolution and the key elements of the policy, including the establishment of an endowment fund. In the last two years the ADS has moved to scale up its level of activities in order to take on the long term preservation and dissemination of a greater proportion of commercially-funded developer-led fieldwork. This has involved the introduction of a semi-automated ingest tool which will allow ADS to offer reduced charges to those depositors who choose to provide well-formed archives and comprehensive metadata.

2.7.2 Sustainability and Provisioning Public Information Goods: "There's Got to be a Profit Around Here Somewhere!"

Kansa, E. (UC Berkeley)

Abstract: Financial sustainability is a paramount concern for archaeology's digital resources. However, it cannot be divorced from financing the discipline in general. Archaeology, as a scholarly or heritage management practice, is manifestly not financially sustainable. It requires continued public support. Because archaeological knowledge dissemination and preservation are integral aspects of the practice of archeology, information dissemination and preservation should not be held out separately in terms of financing. Data preservation (most importantly) and continued experimentation and development of new models for communicating and making sense of archaeological data, will need continued public financing.

The policy focus for sustainability should look beyond individual projects and organizations. Being too narrowly focused on select projects or organizations can motivate behaviors and orientations that put

public interest in a distant second place to the parochial interests of a given organization. The recent battle in the US over the Research Works Act (RWA) illustrates this danger. Some scholarly societies, including the American Anthropological Association and the American Institute of Archaeology actively supported the RWA, despite widespread opposition from researchers, universities, and libraries. These scholarly societies had a difficult time looking beyond the narrow interests of their fee-based publication arms. An information ecosystem with multiple participants will better align sustainability needs with the public interest.

2.7.3 Four core tenets of long-term sustainability

Brin, A. (Arizona State University)

Abstract: A short discussion about the four core tenets of long-term sustainability for digital repositories.

2.7.4 Getting your fingers burnt with on-line interactive publishing

Allison, P. (University of Leicester)

Abstract: This presentation will outline technical and communication processes involved in getting my Pompeian households data published on-line, changes over time and the on-going maintenance of these data. I will discuss two different websites, 'Pompeian Households' and 'Insula of the Menander' in Pompeii (on the Stoa Consortium), between 1995 and 2008 (and beyond), highlighting some pitfalls, and solutions, which can lead to discussion about continuing relevance of such problems.

2.8 Analysis

Respondents: Wilson, A.

We would like to have you present your experience with analytical and visualisation tools. What can we learn from other disciplines about what archaeologists can do with their datasets?

Give us in 5 minutes a vision of your ideal toolset/toolbox. What processing analysis and visualisation is needed in the field to help guide your work? What components are needed for post-processing as you analyse your data as the basis for interpretation.

2.8.1 The Aboriginal Sites Decision Support Tool (ASDST)

Ridges, M. (Office of Environment & Heritage (NSW) & University of New England)

Abstract: The ASDST is an Office of Environment and Heritage (OEH) system that provides a regional scale visualisation of Aboriginal site data held in the Aboriginal Heritage Information Management System (AHIMS). The ASDST uses archaeological predictive modelling to project site distribution across the whole landscape of NSW. This is modelled separately for different site feature types to enable

better visualisation of distribution patterns of each feature type. The system also models each feature against pre settlement and current predicted likleihood to enable visualisation of accumulated impact in the landscape. The ASDST products can be visualised through an on-line mapping system, see: http://www.environment.nsw.gov.au/licences/AboriginalSitesDecisionSupportTool.htm

This talk will demonstrate the tool and discuss the potential of automated modelling from a live archaeological database.

2.8.2 Using digital video to record, present and share information about archaeology and heritage

Colley, S. (University of Sydney)

Abstract: I will screen a short digital video about the Kentwell Cottage Heritage Conservation Project in western Sydney which was jointly produced with Denis Gojak. This will be used to discuss the current potential and challenges of using and archiving digital video to record and present information about Australian archaeology and heritage management.

2.8.3 Heurist: a rich collaborative database

Johnson, I. (Arts eResearch,

University of Sydney)
Wilson, A. (University of Sydney)

Abstract: In this talk we will demonstrate some of the unique functions of Heurist, a flexible online multi-user database for heterogeneous research data.

We will focus on Heurist's ability to inherit database structure from other instances, allowing the creation of structural templates which can be mixed and matched to define a database for a particular project, while encouraging good structure and interoperability. We will demonstrate the way Heurist builds and displays relationships between entities, links to external data on the web and harvests metadata created in the field using the FieldHelper software.

We will also demonstrate Heurist's ability to represent archaeological features through maps and timelines, and its ability to serve data in a variety of formats, including the publication of data as XML feeds, to web pages or within a CMS. These capabilities have been used to deliver situated mobile data on smartphones and pads, as well as building public web sites such as The Dictionary of Sydney.

3 Work Groups

The following pages were given to individual work groups.

4 Artefact Working Group

4.1 Purpose

Formulate the minimum requirements for:

a. Artefact data attributes to be deployed across the system (captured through mobile device applications and used to facilitate the production of comparable datasets).

b. Functionality of the capture of artefact data on the mobile device application.

c. eResearch tools required for searching artefact data.

d. eResearch tools required for the analysis of artefact data.

The working group should also discuss and provide direction on any other aspect of the FAIMS project relevant to the capture, management or analysis of artefact data.

4.2 Scope

For pragmatic purposes, this phase of the FAIMS project has restricted the development of digital artefact recording to two classes of artefacts: ceramics and lithics. While the requirements for data attributes may vary from project to project, we hope that this working group may agree on some minimum requirements for these two classes that will be universal across all assemblages.

FAIMS data capture application development is focused on the digital creation of data in the field on mobile devices (e.g., Andriod phones and tablets), but the project also envisions the development of companion desktop software. Applications will be modular and extensible, but the project is unlikely to initially support the full gamut of attributes required for detailed laboratory recording. As a result (and to encourage the creation of compatible datasets across as wide a range of projects as possible), we ask this group to focus on a minimum threshold of attributes that should be recorded under all (or almost all) circumstances.

At this stage, it is most important to establish a minimum list of attributes (e.g., dimensions, shape, fabric, etc.), rather than the specific vocabularies or taxonomies to populate them (although some discussion on the latter may not be avoidable). We recognise that such vocabularies and taxonomies may vary from one region or time period or project to another, and we expect their details to be developed in other fora.

4.3 Topics to Consider

a. Are there universal attributes (dimensions, shape, fabric, etc.) for lithics and ceramics that applies to all types of archaeology (indigenous, classical, historical, maritime)? Or do these categories

need to be customised by region and chronological period? Are there high-level vocabularies or taxa that might also be widely applicable (e.g., functional categories for ceramics, such as storage, transport, or serving vessels), or will all such vocabularies / taxa need to be customised by region and chronological period?

b. What attributes do you always record? What attributes do you sometimes omit?

c. Are there significant differences in the way you record an artefact and the way you record archaeological features? How might that affect the development of a tool for digital capture?

d. Do you regularly do some sort of group / assemblage record, and then a more detailed inventory record for selected artefacts?

e. Do you photograph or sketch artefacts in the field? What percentage of artefacts is ultimately photographed or drawn (in the field or in the laboratory)?

f. How precise are measurements of artefacts taken in the field? In the laboratory?

g. How do you access reference data needed to identify and date objects in the field or laboratory? Should these be integrated in data capture?

h. How often do you need to search through artefact data collected in the field (digitally or manually)? Is it critical, or is it really only accessed in the lab?

i. Is there more consistency in stratigraphic recording in comparison to artefact recording?

j. What visualisation or statistical tools do you rely on to analyse your artefact data? What visualisation and statistical tools would you like access to in a perfect world?

k. How do you find artefact data created by other researchers? Is searching for artefact data more complex than searching for stratigraphic or site data?

4.4 Expected Outcomes

a. List of essential attributes required for recording a ceramic artefact in the field.

b. List of essential attributes required for recording a lithic artefact in the field.

c. List of functionality requirements for recording artefact data on a mobile device application.

d. List of functionality requirements for searching artefact data on the FAIMS portal.

e. Suggestions regarding workflow or approach (e.g., Do we need separate group and individual records? Should group records be handled with excavation / survey units and individual inventory here?).

f. Suggestions for analytical tools to be included within or developed by the FAIMS portal.

g. Should controlled vocabularies be used throughout? If so, global or local?

5 Contract Archaeology Working Group

5.1 Purpose

Formulate the minimum requirements and parameters for:

a. Functionality of mobile device applications required by contract archaeologists.

b. Functionality of processing, analysis, and visualisation tools required by contract archaeologists.

c. Submission of archaeological data to government agencies.

d. Federation of government cultural heritage registries.

e. Collection of, and access to, sensitive data in contract archaeology.

This working group should also discuss and provide direction on any other aspect of the FAIMS project arising during the Workshop relevant to the capture, management or analysis of archaeological data in the realm of contract archaeology.

5.2 Scope

FAIMS has been funded to develop infrastructure to enhance archaeological research. NeCTAR, the funding program, focuses on university-based research. In Australia, however, private-sector contract archaeologists produce the majority of historical and Aboriginal archaeological data. It is essential that FAIMS develops a system that meets the requirements of archaeologist working in both the academic and contract spheres. There is likely to be considerable overlap in most areas, and these areas of congruence will be discussed in other work groups. The purpose of this working group is to highlight matters which predominantly affect contract archaeologists carrying out their work for the purpose of cultural heritage management.

It is necessary to consider the requirements of government compliance for data capture and publication, but it will not be possible to cover all jurisdictions. FAIMS will endeavour to provide a useful portal, but some government agencies may not have the resources, capacity, or inclination to share their data in the long run.

5.3 Topics to Consider

a. What are the fundamental differences between the practices of contract and academic archaeologists, in the field and in the lab?

b. How different are the legislative frameworks in each state, and across each subfield? How does this affect the way you collect and manage archaeological data?

c. How do government agencies influence the quality of data recording? Are standards and guidelines meaningful? Are they enforced?

d. How do contract archaeologists develop and improve approaches to data capture?

e. How do you share or publish your data? Who owns the data? Do IP concerns restrict data sharing? Who wants this stuff anyway?

f. How useful and accessible do you find datasets and reports created by other contract archaeologists?

5.4 Expected Outcomes

a. List of functionality requirements and desires for archaeological data management, including mobile device applications, data processing and visualisation, and submission to repositories / registries that are specific to contract archaeology (if any).

b. List of requirements for archaeological data or metadata standards and expectations, specific to contract archaeology (if any).

c. Summary of IP concerns (if any) and suggestions for addressing them.

d. List of government registries to be included in the FAIMS network (e.g., for digital submission and search), and strategies for approaching the relevant agencies.

6 Excavation Working Group

6.1 Purpose

Formulate the minimum requirements for:

a. Excavation data attributes to be deployed across the system (captured through mobile device applications and used to facilitate the production of comparable datasets).

b. Functionality of the capture of excavation data on the mobile device application.

c. eResearch tools required for searching excavation data.

d. eResearch tools required for the analysis of excavation data.

The working group should also discuss and provide direction on any other aspect of the FAIMS project relevant to the capture, management or analysis of excavation data.

6.2 Scope

While the requirements for data attributes may vary from project to project, we hope that this working group may agree on some minimum requirements for excavation that may be widely applicable.

FAIMS data capture application development is focused on the digital creation of data in the field on mobile devices (e.g., Android phones and tablets), but the project also envisions the development of companion desktop software. Applications will be modular and extensible. To ensure core functionality and encourage the creation of compatible datasets across as wide a range of projects as possible, we ask this group to focus on a minimum threshold of attributes that should be recorded under all (or almost all) circumstances during excavation.

6.3 Topics to Consider

a. What are the 10 most important components of excavation record? Why?

b. What is the desired granularity of your data? What is the achievable granularity of your data? What is the smallest spatial or conceptual entity that you currently record? That you would like to be able to record digitally?

c. What data do you need to quantify (e.g., deposit composition is 30% gravel and 30% sand)?

d. What in your data is a qualitative observation (e.g., the matrix is highly compacted)?

e. What in your data is a measured, objective observation (qualitative or quantitative, e.g., "this stratigraphic unit yielded 10kg of pottery", "the color of the soil in the matrix is Munsel 7YR")?

f. What in your data is interpretive (e.g., "this wall is late prehistoric")?

g. Is there any "data" that doesn't neatly fit in the above categories?

h. When discussing the above questions, did the group use the same vocabulary to refer to the same concepts? What similar concepts were mapped to different words? What different concepts were mapped to the same words?

i. Is there a significant difference in recording between "simple" single-phase excavations and "complex" multi-layer excavations, enough so that we need separate systems for each?

sectionExpected Outcomes

a. List of essential attributes required for recording excavation data in the field.

b. List of functionality requirements for recording excavation data on a mobile device application.

c. List of functionality requirements for searching excavation data on the FAIMS portal.

d. Suggestions regarding workflow or approach - what would the process of recording a stratigraphic unit in the field look like?

e. Suggestions for analytical tools to be included within or developed by the FAIMS portal.

f. Should controlled vocabularies be used throughout (and if so, global or local)?

7 Federation Working Group

7.1 Purpose and Scope

Formulate the minimum requirements for:

FAIMS seeks to explore the possibilities of data federation in archaeology. Federation is defined as automated data interoperability requiring a minimum of manual intervention as opposed to, e.g., loose-coupling, which according to Kansa and Bissel's (2010) definition relies on extensive human interpretation to reconcile datasets. The degree of federation depends on the level to which semantic and technical interoperability can be achieved.

We have extremely ambitious goals for our federation methods, providing a system which can not only federate disparate syntaxes, but perform limited epistemological translation as well. However, in order to do so, we must identify the most critical data warehouses and mobile data sources producing and containing this data.

This panel is about semantic interoperability and resources that should be made interoperable. We wish to focus on defining the problem of transforming semantically comparable data. What aspects of metadata must be recorded to translate various data source's syntax and meaning to another format? Which aspects should be recorded? Is there a common ground across, e.g., stratigraphic units, surface-survey units, ceramic or lithic records, of at least (say) 10 data points that can serve as the kernel of this federation? If so, how much common ground is more or less universal, and how much varies by region, chronological era, or sub-discipline? Is this common ground too abstract to contribute meaning to the items it stores?

The argument has been made that every project has its own research agenda, requiring a unique data storage "solution" customised to its goals. While much of this argument lies in questions of "solution" to the problem (which we must avoid like the plague), does the problem itself presume unique records? What aspects of commonality exist? What aspects are ontologically and epistemologically unique? Must they always be so? Why?

We have a secondary goal of making mobile data-collection applications. We wish the core, at least, of these applications to be as universally applicable as possible. The data they create, therefore, must be flexible but usefully self-described, without requiring human intervention to parse and interpret. What data and metadata must this mobile device collect or generate? Does this mobile device need a traditional OLTP database? Can the idea of a "fact table" from a data warehouse be applied profitably? Can any aspects of DARCS patch theory (see technical glossary) be applied to the collected observations of these devices?

7.2 Topics to Consider

a. Are shared recording standards, as well as technical data standards, necessary for the success of FAIMS project?

b. Is federation possible without shared recording protocols?

c. Is federation a worthwhile goal?

d. Can FAIMS stakeholders in this room agree on minimum recording protocols for the creation of new data?

e. Is the best achievable goal the loose coupling described by Kansa and Bissel (2010, 44)?

f. To what extent is it possible and desirable to separate quantifiable ("objective") and interpretive ("subjective") data (Ballsun-Stanton 2011)? For example, is there a fundamental difference between the length of a lithic artefact and the date of a lithic artefact?

g. How rigid should be the line between objective measurement versus subjective interpretation? Does this need encoding in metadata?

h. How can we ensure the capture of data that allows re-interpretation?

i. If we do have standards, are they regional, chronological, sub-disciplinary, or global? Should we create our own standard and/or support translation between competing standards?

j. Can there be a shared set of standards between cultural heritage and research, or archaeology and related disciplines, that does not discard all meaning of the relationships within the structure of the database?

k. What existing resources need to be supported? Prioritize these resources:

    a. Online repositories.

    b. Field Databases.

    c. Reference collections.

    d. Other tools or resources (specify).

l. How do we get these resources (such as state registries) on board?

7.3 Expected Outcomes

FAIMS asks this group to produce answers to all the questions above, and in particular to produce a definitive statement on:

a. What level of federation is desirable by FAIMS stakeholders?

b. What (if any) core, minimum data and metadata standards (attributes, vocabularies, ontologies, etc.) can most archaeologists agree upon, and how do these standards vary (e.g., are they universal, regional, sub-disciplinary, etc.)?

c. What type of semantic standards should FAIMS embrace or create?

d. What existing resources need to be supported by FAIMS, listed by type and priority.

8 Sensitive Data Working Group

8.1 Purpose

Describe the problem and formulate testable requirements for the various issues surrounding "sensitive" data. What legal, moral, ethical, confidential, political, and social problems can occur with the collection, aggregation, display, and export of the data collected by mobile devices that FAIMS is developing? What do bad actors look for in archaeological data? How can we protect against them?

The working group should also discuss and provide direction on any other aspect of the FAIMS project relevant to the capture, management, or publication of sensitive data.

8.2 Scope

FAIMS is aware that many categories of archaeological data collected in the field may be sensitive, personal, or confidential in nature. FAIMS is asking this group to define the kinds of data should be considered sensitive, and to articulate approaches to recording such data. FAIMS also requires guidelines and strategies for managing, archiving, sharing, and digitally publishing sensitive data.

8.3 Topics to Consider

a. What kind of sensitive data will we be capturing with mobile devices?

    a. What specific components of the data are sensitive? Who should it be protected from?

    b. What component of the data should be shared? What level of sharing control should it have in the best of all possible worlds? What could go wrong with that level of sharing control?

    c. What makes this data sensitive beyond its individual components?

    d. What sensitive data is routinely recorded?

    e. What edge cases should we cover?

    f. How have previous systems failed you?

b. How will people want to interact with this system?

    a. Do we want to provide all these methods?

    b. What other methods do we want to provide?

c. What are the different use cases for adding data, editing data, accessing data, and searching data?

d. What preparatory actions does the system need to support?

e. How can the system promote and require external analysts to maintain adequate protection of the data?

    a. Will people actually *use* this feature, or will they work around it?

    b. How would an analyst actually interact with the data?

c. How will bad actors want to interact with the system?

    a. Do we want to presume good faith?

        a. Are all actions allowed unless forbidden, or are all actions forbidden unless allowed?

        b. What general uses cases for classification are there?

    b. What audit capabilities are necessary?

d. What aggregations are necessary to prepare data for bulk export?

    a. Do these aggregations protect the sensitive data?

        a. Why?

    b. How could a bad actor de-anonymize these aggreations?

        a. Are the detail-tradeoffs worth the level of protection afforded?

        b. Is this just complying with statutory regulations without providing true anonymity?

e. How would sensitive data be useful to other archaeologists in the field?

    a. How would they authenticate themselves?

    b. How would they access it?

    c. How would bad actors use this as an attack vector?

    d. How have you used other sensitive data? How did you access it? What sucked about the access method?

f. Is there any consistency in rules about sensitive data? What's the most ironic rule "trap" you've experienced in this regard?

8.4 Expected Outcomes

a. List of (say) 10 types of sensitive data, with a brief statement about what makes each sensitive.

b. List of requirements for the recording of sensitive data in the field.

c. List of minimum security and publication requirements for archiving and publishing sensitive data.

d. List of functionality requirements for searching, publishing and sharing sensitive data on the FAIMS portal.

e. Suggestions for tools to be included or developed for the FAIMS portal useful for analysing sensitive data.

9 Survey Working Group

9.1 Purpose

Formulate the minimum requirements for:

a. Pedestrian surface survey data attributes to be deployed across the system (captured through mobile device applications and used to facilitate the production of comparable datasets).

b. Functionality of the capture of survey data on the mobile device application.

c. eResearch tools required for searching survey data.

d. eResearch tools required for the analysis of survey data.

The working group should also discuss and provide direction on any other aspect of the FAIMS project relevant to the capture, management or analysis of pedestrian surface survey data.

9.2 Scope

While the requirements for data attributes may vary from project to project, we hope that this working group may agree on some minimum requirements for pedestrian surface survey that may be widely applicable.

FAIMS data capture application development is focused on the digital creation of data in the field on mobile devices (e.g., Andriod phones and tablets), but the project also envisions the development of companion desktop software. Applications will be modular and extensible. To ensure core functionality and encourage the creation of compatible datasets across as wide a range of projects as possible, we ask this group to focus on a minimum threshold of attributes that should be recorded under all (or almost all) circumstances during surface survey.

9.3 Topics to Consider

a. What are the 10 most important components of survey record? Why?

b. What is the desired granularity of your data? What is the smallest spatial or conceptual entity that you would like to be able to record digitally?

c. What is the achievable granularity of your data? For example, do you only record site-level data, or do you record blanket artefact densities, including absences of artefacts? What is the smallest spatial or conceptual entity that you record?

d. What data do you need to quantify (e.g., "this unit has 70% visibility")?

e. What in your data is a qualitative observation (e.g., "this field is difficult to cross")?

f. What in your data is a measured, objective observation (qualitative or quantitative, e.g., "this total pickup yielded 10kg of pottery", "the color of this ceramic fabric is Munsel 7YR")?

g. What in your data is interpretive (e.g., "the pottery in this unit dates to the Roman era")?

h. Is there any "data" that doesn't neatly fit in the above categories?

i. When discussing the above questions, did the group use the same vocabulary to refer to the same concepts? What similar concepts were mapped to different words? What different concepts were mapped to the same words?

9.4 Expected Outcomes

a. List of essential attributes required for recording surface survey units in the field.

b. List of functionality requirements for recording surface survey unit data on a mobile device application.

c. List of functionality requirements for searching surface survey unit data on the FAIMS portal.

d. Suggestions regarding workflow or approach - what would the process of recording a survey unit in the field look like?

e. Suggestions for analytical tools to be included within or developed by the FAIMS portal.

f. Should controlled vocabularies be used throughout (and if so, global or local)?

10 Sustainability Working Group

10.1 Purpose and Scope

FAIMS seeks to develop an infrastructure that would allow archaeological data to be managed from its digital creation to processing, archiving and publication. We want to establish the core of a distributed, open-source ecosystem for archaeological data management. Such infrastructure will combine many tools and repositories, and exceed many existing digital resources in scope. Given the idiosyncratic data that archaeologists collect, the mobile tools will need to be continually developed and customised to fit the needs of new archaeological objectives. Repositories and the databases behind them will require ongoing maintenance. Maintaining the functionality of the entire infrastructure and sustaining the development of its individual components will be a major challenge. FAIMS would like this group to consider the challenges of long-term maintenance and development of the FAIMS infrastructure and its components. FAIMS asks this group to articulate approaches and models to sustain the project.

Topics to Consider

a. How do you usually acquire the tools and software required for your fieldwork?

b. How much does it cost your project to manage your data? Consider hardware, software, publication, and personnel costs.

c. Do you think you could invest more in software and tools, and less in personnel?

d. What are the main differences between academic and contract archaeologists in how they invest in data management?

e. How often do you encounter incompatible datasets that hinder your research?

f. Should individual components be funded separately and through different strategies?

g. Are you willing to pay for individual components separately? If so, which ones?

h. What models of funding are you familiar with or can you identify for an open source archaeological data management ecosystem?

10.2 Expected Outcomes

a. Identify possible funding models for FAIMS project.

b. Identify possible funding models for individual FAIMS components.

c. Identify "reasonable costs" for various possible fees.

d. Identify how best to approach different categories of stakeholders (e.g., university researchers vs. contract archaeologists) when funding the project.

e. Articulate a successful overall strategy for long-term sustainability for FAIMS digital tools.

10.3 Ten funding approaches for consideration

Consider this list a starting point - eliminate, add, and qualify as you see fit.

a. Software as a service - a recurring fee for mobile application and other software.

b. A customization fee for field recording systems.

c. A curation fee for online archiving and publication.

d. An access fee for use various FAIMS online components.

e. A subscription fee for the FAIMS portal (i.e., a fee charged to libraries, universities, or researchers to access the datasets).

f. Pay-per-view for downloads.

g. Subscription or hourly rate fees for help-desk support.

h. Grant support sought by FAIMS (i.e., applications for further infrastructure grants).

i. Expectation or requirement that individual (academic?) users of FAIMS apply for grant support (e.g., on ARC or NSF grants).

j. Corporate and/or other institutional sponsorship.