FAIRer Data through Digital Recording: The FAIMS Mobile Experience

SHAWN ROSS (iD)

BRIAN BALLSUN-STANTON (iD)

STEVE CASSIDY (iD)

PENNY CROOK (iD)

JENS KLUMP (iD)

ADELA SOBOTKOVA (iD)

*Author affiliations can be found in the back matter of this article

]u[ubiquity press

ABSTRACT

Critical data and metadata must be captured or created in the field, or shortly thereafter, to avoid loss. For the past 10 years, the Field Acquired Information Management Systems (FAIMS) project has developed and operated a customisable field data capture platform. Over time, we built features and approaches that incorporated the Findable, Accessible, Interoperable, and Reusable (FAIR) principles into born-digital datasets created during fieldwork. This paper synthesises our experience helping more than 40 projects adapt the FAIMS platform to nearly 70 research workflows in archaeology and other fieldwork domains. We review what elements of the FAIR Data Principles FAIMS was able to build into our software, how users received these capabilities, and what sociotechnical challenges impeded creation of FAIRer field data. Based on our experience, we argue that field data capture software can facilitate the production of FAIRer data, making those data much more Findable and Reusable, and somewhat more Accessible and Interoperable. Any such improvements, however, depend upon (1) making FAIR-data features an integral part of field data collection systems, minimising the burden imposed on researchers, and (2) researchers' willingness to spend time and resources implementing FAIR Data Principles that do not provide immediate benefits to their research.

1 INTRODUCTION

To address the grand challenges of archaeology, archaeologists need to amalgamate data from many projects (Kintigh 2006; Kintigh et al. 2014). These data need to be available, understandable, and amenable to reuse by those not involved in its creation. This need is not unique to the field of archaeology.

The Findable, Accessible, Interoperable, and Reusable (FAIR) Data Principles provide the foundation for publishing portable datasets that can be repurposed (Wilkinson et al. 2016). Briefly, data are Findable when they have a unique and persistent identifier, have been described with metadata, and the (meta)data are indexed and searchable. Data are Accessible when they can be retrieved by identifiers using standard protocols (such as via the web), and Interoperable when (meta)data conform to a shared language of knowledge representation, including use of published vocabularies and an explicit data model. Finally, data are Reusable when described with metadata sufficient for understanding and conforming to domain-relevant standards, explicitly licensed, and associated with detailed provenance (GO-FAIR 2017). The key elements of the FAIR Data Principles related to fieldwork focus on metadata, persistent identifiers (PIDs), provenance, and standardised and shared approaches to (meta)data representation.

If archaeological data are going to become FAIRer, the principles need to be implemented during fieldwork. The Digital Curation Centre observes that 'much of the most crucial information required for effective long-term curation and reuse must be captured at the conceptualisation and collection stages' (Digital Curation Centre 2010). Researchers in other fieldwork domains, such as ecology, have long recognised that 'important information about [data's] origins, context, and provenance may be lost' if long-term archival considerations are not considered early in the research lifecycle (Wallis et al. 2008). The FAIR Data Principles have now begun to shape discussion of field data capture in archaeology (Lindsay and Kong 2020; Sobotkova et al. 2021). Creating FAIR data early in the data lifecycle, however, requires planning from the beginning of a project (Niven 2011a; Whitmore and Dennis 2019), as well as a willingness on the part of researchers to invest time and resources to implement these plans.

This paper presents a case study exploring how the Field Acquired Information Management Systems (FAIMS) project was able to retrofit support for FAIR data onto an existing platform. It synthesises our experience co-developing nearly 70 customisations of the platform for more than 40 projects, discusses which capabilities were adopted, and considers the sociotechnical challenges to the creation of FAIRer field data—a labour-intensive undertaking that may benefit the archaeological community but can burden individual researchers. More broadly, we investigate the extent to which digital field recording systems can profitably incorporate the FAIR Data Principles, considering technical and user-driven constraints and opportunities.

2 THE FAIMS PROJECT AND FAIMS MOBILE PLATFORM

The FAIMS Project is a university-based research infrastructure project, hosted first by UNSW Sydney and later by Macquarie University, Sydney Australia. Its principal output has been FAIMS Mobile, an open-source platform used to create custom field data collection applications for Android devices. This platform was built between 2012 and 2014. Originally designed for archaeologists, it has now been used in disciplines including environmental geochemistry, avian ecology, linguistics, and oral history.

The architecture and functionality of FAIMS Mobile have been presented elsewhere (Ballsun-Stanton et al. 2018; Ross et al. 2015, 2013; Sobotkova et al. 2015), as have case studies about its deployment in archaeology (Sobotkova et al. 2016, 2021) and geochemistry (Noble et al. 2018, 2020; Thorne et al. 2018). Key features and capabilities of FAIMS Mobile are summarised in Supplemental Material.

This software is not a preconfigured data logger, it is a platform for the minting of custom Android applications. As such, every deployment implements its own data schema and user interface (UI), instantiating a particular research workflow in a custom FAIMS Module. Most projects employed FAIMS Project staff to assist with data and workflow modelling, and to produce the code needed for customisation. This co-development provided insights into field data collection approaches and how they could be embedded into a digital recording system. It also let us explore with researchers how data discoverability and reusability might be improved.

At the time of publication, FAIMS Mobile is undergoing a comprehensive rebuild. The new version will be available in 2023. This paper refers to the earlier software platform.

FAIMS Mobile was originally designed before Wilkinson et al., published the FAIR Data Principles (2016). The system did, however, incorporate contemporary linked open data (LOD) approaches, laying the groundwork for later support of the FAIR Data Principles. We have worked to incorporate these principles into FAIMS Mobile customisations at project, dataset, record, and attribute level. The sections below explain our approaches, characterise how they were received by users of the platform, and provide brief commentary on potential improvement.

3 MAKING DATA FINDABLE

As specified at GO-FAIR (GO-FAIR 2017: 'Findable'), Findability requires the application of a globally unique and persistent identifier (a 'PID'; F1), the description of data with rich metadata (F2, further defined by R1), incorporation of

the PID allocated at F1 into the metadata at F2 to ensure that data and metadata are well connected (F3), and registration of meta(data) in a searchable resource such as a register or repository (F4). Allocation of PIDs in the field and the production of dataset-level metadata offer the most scope for improvement during data creation.

3.1 IDENTIFIERS (F1)

The FAIMS Mobile Platform has the capacity to allocate two types of identifiers: those for records and custom identifiers, such as those for physical samples. The system generates a long, numeric identifier (called a Universally Unique Identifier (UUID), though only guaranteeing local uniqueness in our particular case, see 3.1.1) for all records; and human-readable, record-level identifiers are optional. In deployments to date, neither of these identifiers anticipated or facilitated later application of a PID such as a DOI. Customisations for the Commonwealth Science Industry and Research Organisation (CSIRO) Mineral Resources Unit in Western Australia, however, did assign International Generic Sample Numbers (IGSN) in the field (see below). FAIMS could also incorporate existing PIDs into the recording system, a capability discussed under '5.1.2 Shared vocabularies' below.

3.1.1 Record identifiers

FAIMS Mobile incorporates a primary key into each record to identify it in the database. Although we investigated the use of formal UUIDv4s as per RFC 4122 (Network Working Group 2005), interactions between our database (SQLite) and the Spatialite geospatial extensions limited us to a short integer identifier (bit length of 2^{63}, whereas a UUIDv4 has 2^{128}). Given the need to generate identifiers offline on multiple unconnected devices simultaneously, we produce locally unique identifiers. This identifier was formed by concatenating a '1', the local user ID (five digits), and the UNIX epoch (to milliseconds). For example, the identifier '10000216357639912721' was created by user ID 2, at 'Mon, 01 Nov 2021 10:51:52 GMT'. The use of long integers as identifiers hindered data export to Excel, which converted them into exponentials in a manner that truncated them.

Most projects also wanted meaningful, human-readable identifiers. At first, projects often requested automatic generation of such identifiers based on existing naming or numbering conventions. Such conventions could be complex and costly to implement; see Figure 1 for an example of an artefact group numbering scheme (Ballsun-Stanton and Estephan 2014: 1058–1087).
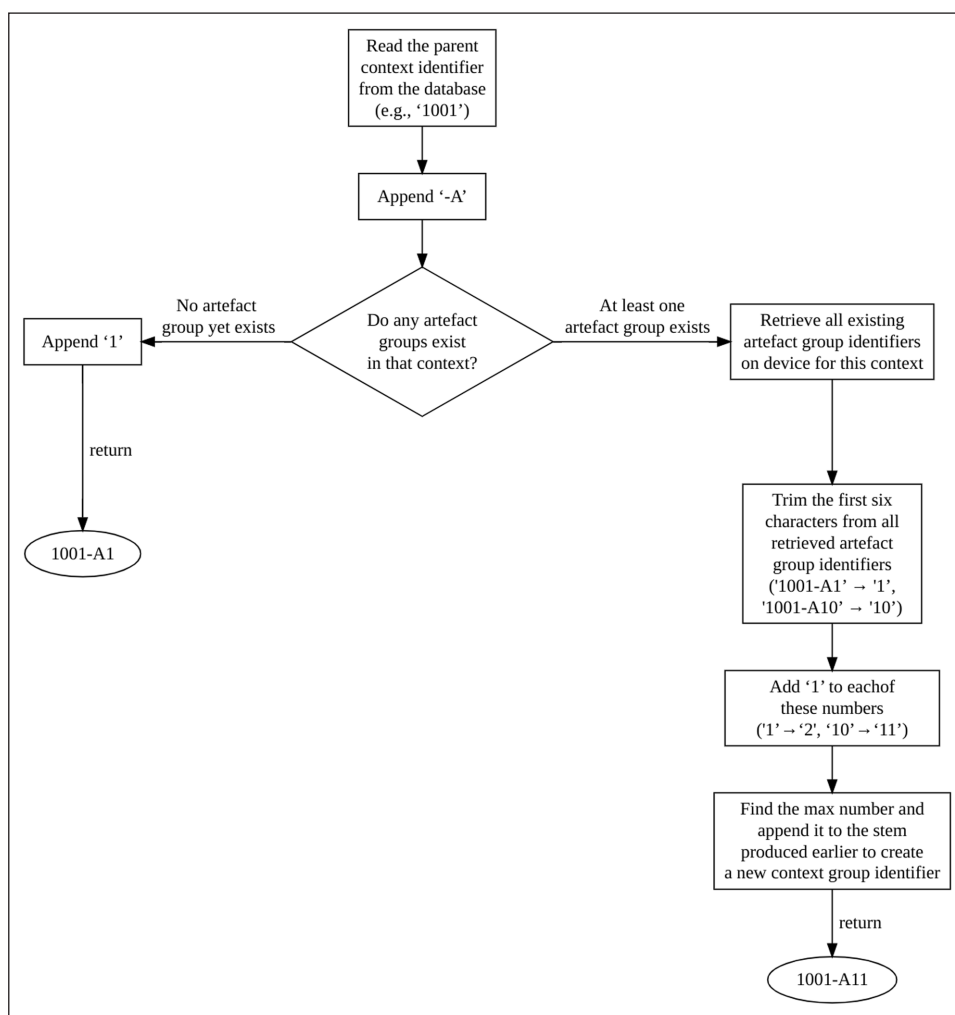

Figure 1 The process of generating human-readable identifiers for artefact groups from excavation contexts. Brian Ballsun-Stanton, CC-BY.

Such complex identifiers should be avoided. Generating them threatens to introduce race conditions, a hazard arising when a system's behaviour is dependent on the sequence of events in distributed locations. It was also possible to create duplicate identifiers on different devices when working offline. Complex lookups were, moreover, computationally expensive. To allow arbitrary customisation of the data schema, FAIMS Mobile uses a domain-key normal form (DKNF), append-only datastore. This approach allows fast writes but slows down during search and read operations as the size of the database increases. Since FAIMS is offline-first, moreover, all data interactions must be carried out locally on the device rather than on more powerful remote servers. Finally, the archaeological tradition of encoding references to the Site or Project into identifiers is unnecessary during digital recording, which offers other ways to produce meaningful relationships and preserve essential data.

Recognising these limitations and opportunities, later projects adopted short, auto-incrementing numbers as human-readable identifiers, assigning different ranges to different devices to avoid duplication. To make these identifiers more meaningful, projects prepended or appended information from the record itself to identifiers. One survey project, for example, concatenated the feature type and the auto-incrementing number into a human-readable identifier (e.g., 'Isolated find 1013'). Such identifiers, drawn from any information contained in a record, could be generated on the device or upon data export.

Owing to the frequency of formatted identifiers, we added a grammar parser (Antlr3) and a domain-specific language to define syntax (FAIMS Project 2015). We investigated minting record-level DOIs at the time, but DOI providers expressed concerns about allocating them in such large numbers—something unlikely to pose a problem today. As a result, FAIMS Mobile identifiers did not conform to DOI syntax, nor did they prefigure later allocation of a DOI. Minting DOIs (or 'pre-DOI' local identifiers) should be possible using an approach like the one FAIMS used for IGSNs (Klump et al. 2021).

3.1.2 Physical sample identifiers: In-field allocation of IGSNs

The CSIRO Mineral Resources Unit in Western Australia customised FAIMS Mobile for the rapid and accurate collection of environmental samples. This customisation allocated IGSNs to samples during fieldwork when they were collected (Noble et al. 2020). Allocation involved inserting the IGSN into the sample's digital record and attaching a physical label to the sample bag displaying human- and machine-readable IGSNs. Machine-readable IGSNs used a barcode to represent the IGSN, or a QR code that included the IGSN URI to make it universally readable. The entire allocation process was conducted offline, in remote settings, and under stringent time

constraints. Two approaches were used. The first involved generating an IGSN in FAIMS Mobile and printing a label using an attached Bluetooth thermal transfer printer. After field trials, however, the preferred method was to print a 'raffle book' of IGSNs on durable adhesive labels before fieldwork, attach a label to the sample bag when the sample was collected, and then to scan the IGSN from the QR or barcode into the active record.

3.2 PROJECT- AND DATASET-LEVEL METADATA (F2)

We have discussed our approach to metadata elsewhere (Sobotkova et al. 2021: 3, 19–20); here we extend that discussion, focusing on project- and dataset-level metadata creation related to data discoverability. Other aspects of metadata creation in FAIMS Mobile customisations are presented below (see '6.0 Making Data Reusable').

Different field workflows, such as surface survey, feature registration, or excavation used different FAIMS Mobile modules, particularly as the cost of customisation declined and the performance penalty and user-facing complexity of large systems covering multiple workflows became apparent. Each module, therefore, produces a coherent and meaningful dataset (e.g., of survey units, features, excavation contexts, etc.).

Dataset- and project-level metadata can be attached to a customisation via the 'Module Settings' page. Any information entered on this page can be exported along with module data. Fields include a general description of the project, a Spatial Reference System Identifier ID (SRID) to facilitate geospatial work with the data, permit information, project contact information, participants, copyright holder, client or sponsor, landowner, and a flag indicating whether any of the collected data are sensitive (see Figure 2).

In practice, use of these fields by projects was inconsistent. The settings page could have allowed project- and dataset-level metadata to be linked to all data produced by the customisation, improving discoverability. Module-level metadata was often incomplete, however, and included only human-readable metadata the project found useful in the field (e.g., permit numbers), or that FAIMS staff added to facilitate user support (e.g., SRIDs or contact information).

Module metadata arose from requirements gathered between 2012 and 2014; at that time, little consideration was given to the creation of metadata solely to aid dataset discoverability. Alignment of the metadata with generic standards such as Dublin Core or the DataCite Metadata Schema (DataCite 2021; DCMI Usage Board 2020) and domain-specific standards such as those promoted by the UK Archaeology Data Service (ADS) or Digital Antiquity's tDAR repository (Archaeology Data Service 2022; Niven 2011b) could improve early creation of project- and dataset-level metadata contributing to data findability. No users requested improvement,

Figure 2 Project- and dataset-level metadata captured in the 'Module Settings' page.

extension, or standardisation of module metadata. To the best of our knowledge, published data captured using FAIMS Mobile (e.g., Noble et al. 2017) had project-level metadata added later in the data lifecycle.

3.3 DATA–METADATA LINKING VIA IDENTIFIERS (F3)

FAIMS Mobile binds together all types of data captured using the application, including structured, text, multimedia, geospatial, and instrument data. Available instrument metadata, such as internal or external GPS accuracy, is also captured. Since all (meta)data are contained within the record, they are intrinsically associated with the record's identifier. Cross-references and relationships between

(meta)data entities or records are discussed below under '5.3 Relationships and cross-references'.

3.4 (META)DATA REGISTRATION (F4)

Repository ingest of (meta)data is facilitated by the flexibility of FAIMS Mobile exports. Data can be exported in common formats like CSV, KML, XML, PDF, JSON, or shapefile (see, e.g., Ballsun-Stanton and Heřmánková 2021). Custom exports can also be written once and then applied to all FAIMS Mobile customisations. A GeoJSON exporter, for example, was built for transferring data to Open Context. Since users always processed data locally before submitting it to repositories, this direct export was never used. Export to common formats for ingest into a

variety of desktop software for processing and analysis proved most useful. Exporters, like data collection customisations, are available on GitHub.

3.5 SUMMARY OF THE FAIMS EXPERIENCE MAKING DATA MORE FINDABLE

FAIMS Mobile supports the production of locally unique machine- and human-readable **identifiers** (F1), encouraging researchers to rely on simple numbering systems plus record contents to generate human-readable identifiers, rather than complex, automated naming schemes. These local identifiers are not, however, articulated with PID allocation, leaving room for improvement. IGSNs, by contrast, have been assigned to physical samples in the field. Project- and dataset-level **metadata** (F2) can be captured for each FAIMS Mobile customisation, but completion of this metadata by users has been inconsistent, and was not aligned with generic or domain-specific metadata standards. FAIMS Mobile binds all (meta)data (of any type) into a single dataset, allowing it to be **linked** (F3) via an identifier. The system exports data in a variety of standard formats, facilitating eventual **registration** (F4) in a repository. No projects used direct-to-repository export as all subjected their data to further processing.

In short, most of the Findability elements of FAIMS Mobile are being added later in the data lifecycle at present, usually when data are ingested into a repository, but F1 and F2 (especially) could be profitably incorporated into future field data capture systems.

4 MAKING DATA ACCESSIBLE

As per GO-FAIR (2017: 'Accessible'), data accessibility entails the ability to retrieve data using their identifier over a standardised protocol (A1), preferably using open, free, and universal communications protocols (A1.1) that support authentication and authorisation where necessary (A1.2), and ensuring that metadata are accessible even when the data are no longer available (A2). Enabling accessibility mostly lies in the realm of the data publication service or repository, but one aspect of this principle could be facilitated during fieldwork: flagging sensitive data to indicate when it requires restricted access (A1.2).

4.1 SENSITIVITY FLAGGING (A1.2)

Many projects in oral history, linguistics, archaeology, and ecology used FAIMS Mobile to collect sensitive data that required access controls. While any such controls would be implemented by the repository housing the data, the Module Settings page (discussed above) includes a 'Has Sensitive Data' field that could be set to 'true' or 'false'. The flag applied to all data collected by the associated customisation and included no explanation.

No users requested more granular flagging of sensitivity or the ability to elaborate why the data was sensitive. Refinement of the dataset-level sensitivity metadata and incorporation of record- and attribute-level sensitivity markers would improve this capacity.

5 MAKING DATA INTEROPERABLE

Interoperability arises from using 'formal, accessible, shared, and broadly applicable language for knowledge representation' (GO-FAIR 2017: 'Interoperable'). It includes articulation of an explicit data model and the use of shared vocabularies, thesauri, or ontologies for (meta)data terms (I1). Vocabularies, moreover, should be both human and machine readable, reference terms using PIDs, and be defined using common knowledge organisation systems like RDF (I2). Finally, references between (meta)data, whether internal or external to the dataset, should be possible and qualified (e.g., 'A is part of B'; 'B contains A'; I3). FAIMS Mobile supports consistent use of local controlled vocabularies, can link them to shared vocabularies and provides qualified cross-references. Customisation, importantly, requires specification of a data model.

5.1 CONTROLLED VOCABULARIES (I1; I2)

5.1.1 Local vocabularies

FAIMS Mobile promoted the use of local controlled vocabularies (Ballsun-Stanton et al. 2018; Sobotkova et al. 2021). The system included the usual drop-down lists, radio buttons, and checkboxes. In addition, FAIMS implemented more advanced features like 'picture dictionaries' that allowed users to select vocabulary terms using images, and 'hierarchical dropdowns' that allowed them to step through a classification hierarchy to find a term or to follow a procedure leading to a classification (e.g., stepping through taxonomic levels to select a species or applying the USGS soil characteristics workflow to arrive at a classification of 'sandy loam'; see Figures 3–4). Use of controlled vocabularies is a crucial step towards making datasets more Interoperable and Reusable. Employment of such vocabularies was, moreover, embraced by users since it improved the consistency of their data and sped up the data-entry process. FAIMS controlled vocabularies were defined using machine-readable XML (see '5.2 Data modelling') but were not described using any formal knowledge representation system.

5.1.2 Shared vocabularies

All entities, attributes, and vocabulary terms in FAIMS Mobile could embed PIDs by associating a URI with the desired element in the XML definition file (FAIMS Project 2016: 56, 77, 164). The URI could then be exported alongside or instead of the local term. Local terms could thus be linked to terms in shared vocabularies, thesauri,
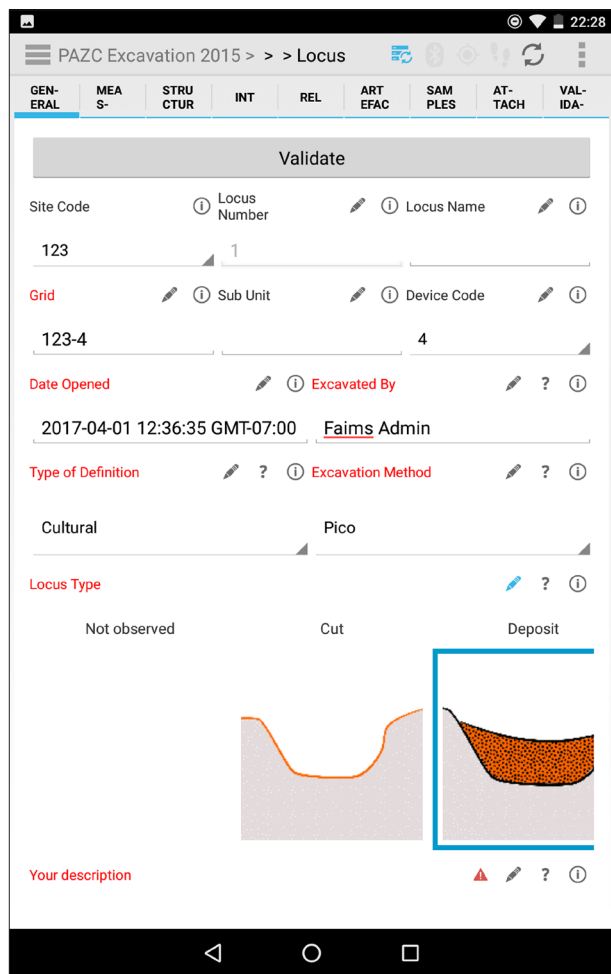
Figure 3 Picture gallery controlled vocabulary in the Proyecto Arqueológico Zaña Colonial Excavation module (Van Valkenburgh et al. 2015). Photo Credit: Petra Hermankova, CC-BY.

or ontologies, such as the Pleiades Gazetteer or the Getty Art and Architecture thesaurus (Ballsun-Stanton et al. 2018; e.g., Sobotkova et al. 2021: 21). Although FAIMS Mobile supports the use of published vocabularies, researchers did not use this term-mapping capability, even when encouraged to do so by FAIMS staff during data modelling. On one hand, the underutilisation of this feature can be seen as a clear example of researchers' reluctance to spend time improving the FAIRness of data when the effort did not directly support their own objectives. On the other hand, it may simply reflect the poor suitability of published vocabularies in specialised fields such as archaeology—a situation that we hope will improve in time.

5.2 DATA MODELLING (I1)

Digital data collection systems all require the development of a data model—a beneficial consequence of 'going digital'. Some systems (like FileMaker) create this model implicitly from forms made by users, while others (like FAIMS Mobile) do so explicitly. Customising FAIMS Mobile required articulation of the entities, attributes, and values that researchers want to collect and any relationships between them (the 'data schema'). The data schema
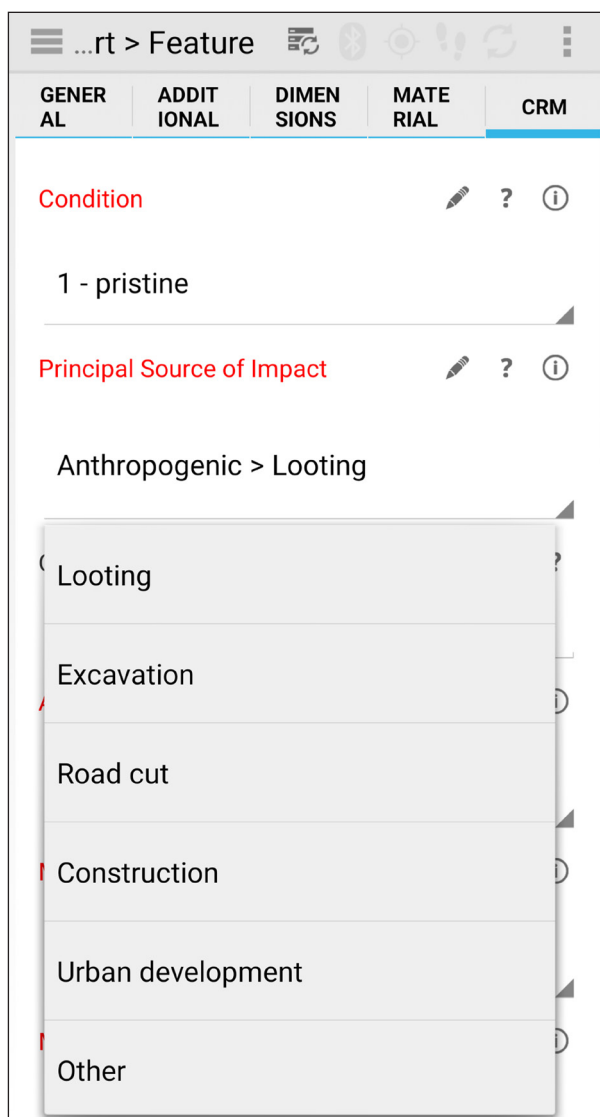

Figure 4 Level 2 of a hierarchical vocabulary in the Perachora Peninsula Archaeological Project Legacy Data Verification module (Sobotkova, Hermankova, and Nassif-Haynes 2020b). Photo Credit: Brian Ballsun-Stanton, CC-BY.

was defined by a machine-readable XML file (see also '6.2.2 Metadata arising from customisation' below for UI definition). Definition files used a custom-built, domain-specific language, and thus required 'specialised… algorithms, translators, or mappings' to render them, a practice discouraged by I1. Use of, or translation into, a standard format like OWL or RDF-Schema would bring it into alignment with the principle.

5.3 RELATIONSHIPS AND CROSS-REFERENCES (I3)

The FAIMS Mobile datastore did not intrinsically create relationships between entities like a third normal form (3NF) relational database does. Instead, each relationship had to be declared explicitly. A parent–child relationship, for example, needed code to copy identifiers to child records, simulating a foreign key in a 3NF database. In addition, relationships could be qualified. They could be bidirectional ('Context A adjoins Context B' and vice

```
<rels>
  <RelationshipElement name="Survey Unit - Walker" type="hierarchical">
    <description>A 1-to-n relationship between the parent and its children,
respectively.</description>
    <parent>Parent Of</parent>
    <child>Child Of</child>
    <property type="measure" name="Order" isIdentifier="true"/>
  </RelationshipElement>
</rels>
```

Figure 5 Code defining a hierarchical relationship between 'Survey Unit' and 'Walker Cell' for surface survey as part of the Perachora Peninsula Archaeological Project Gridded Pedestrian Survey module (Sobotkova, Hermankova, and Nassif-Haynes 2020a).

versa) or hierarchical ('Context A is parent of Feature C' while 'Feature C is child of Context A'); see Figure 5 for a hierarchical example. Relationships between entities could be declared once in the data schema, in which case they applied to all records, or they could be optional and invoked by the user as needed when records were created. Taken together, qualified cross-references within a dataset described how the different entities and their records related to each other, supporting comprehensive, qualified cross-referencing as required by Interoperable element I3.

Since these relationships were machine-readable, they could be manipulated programmatically. We created an exporter, for example, to generate a Harris Matrix from excavation records that employed chronologically qualified relationships (Ballsun-Stanton 2018). Machine readability also facilitated interoperability, especially if the verbs qualifying relationships were aligned with a shared vocabulary such as that provided in the DataCite Metadata Schema (2021: 12.b).

Mediocre performance limited the utility of ad hoc cross-references. Most projects wanted to select the entities involved in such a relationship from a drop-down that was generated on demand. The read penalty imposed by the DKNF, append-only datastore made on-the-fly lookups slow once the dataset reached a certain size (a few hundred records for complex customisations). Relationships hard-coded for an entire dataset did not suffer from this slowdown.

Sociotechnical problems were more challenging. Although qualifying relationships and exporting data as RDF triples would have been technically straightforward, it was never requested. Most projects requested only 'flat' data in the form of CSVs and shapefiles, for processing in spreadsheet and desktop GIS software. While many projects used simple, fixed parent–child relationships, the implementation of other types, or definition of relationship types, was seen as adding unnecessary complexity for little benefit (with the notable exception of excavation projects wanting to generate Harris Matrices).

5.4 SUMMARY OF THE FAIMS EXPERIENCE MAKING DATA MORE INTEROPERABLE

FAIMS Mobile implemented various approaches to **shared knowledge representations** (I1): use of local controlled vocabularies (including hierarchical classifications) and specification of a data model using a machine-readable XML file. Both were expressed using a custom domain-specific language rather than a common knowledge representation language like OWL or RDF-Schema. Controlled vocabularies improved data quality and were enthusiastically adopted. Although these local vocabularies could be linked to published, **FAIR vocabularies** (I2) via embedded URIs, researchers did not see the advantage of spending time and effort to make such mappings. As a result, improvements to semantic interoperability were limited. FAIMS Mobile could create arbitrary **qualified relationships** (I3) between entities, but only the simplest such relationships were commonly used.

Digital field recording systems like FAIMS Mobile can facilitate the use of FAIR vocabularies, explicit data models, and cross-references between (meta)data elements—if researchers can be convinced of the value offered by these approaches.

6 MAKING DATA REUSABLE

Reusability relies on describing data with 'rich' and 'plural' metadata (R1) so that it can be understood, reused, and combined with other data in a variety of settings (GO-FAIR 2017: 'Reusable'; parallel to F2 but focused on understanding data for reuse rather than discovery). Reusability also depends upon clear licensing (R1.1), description of the origin and history of the data, including the workflow that produced it ('provenance'; R1.2), and the application of domain-relevant (meta) data standards (R1.3). All aspects of this principle can be incorporated into field data recording and benefit from attention early in the data lifecycle.

6.1 RICH AND PLURAL METADATA (R1)

FAIMS Mobile allowed creation of metadata at three levels:

1. Project and dataset (tied to a given customisation, which usually produced coherent dataset).
2. Individual records in a dataset.
3. Individual values in a record.

Furthermore, FAIMS Mobile enabled production of metadata in three ways:

1. Manually by users.
2. Automatically by the system (including reuse of information captured previously).
3. As a by-product of customising FAIMS (i.e., writing the XML files used to define the data schema and UI).

This section focuses on producing record- and value-level metadata, whether manually or automatically (see '3.3 Project- and dataset-level metadata' for higher-level metadata).

6.1.1 Record-level metadata

Record-level metadata need to be created when metadata vary from one record to the next. So, for example, while the year of fieldwork can be defined once for the entire dataset, data creation time must be specified for each record. Such record-level metadata are crucial for selecting, understanding, and reusing data once a dataset has been found.

Manual entry of record-level metadata. Manual entry of record-level metadata was accomplished, like data creation, through the simple expedient of providing form fields to capture appropriate attribute–value pairs. Metadata fields could use any available form element (see '5.1.1 Local vocabularies' above), sensors could be called via 'action buttons', and validation applied, either on the device at the time of data capture or on the server after synchronisation. Help text and images could be attached to any form field to guide (meta)data entry. These features supported precise and consistent entry of both data and metadata.

Location was the most common manually created record-level metadata. Locations were captured from internal or external GPS sensors via an action button. They were recorded as longitude–latitude, northing–easting (with projection metadata stored at the dataset level), or both (conversion was done on the fly). Position accuracy was also recorded using the error reported by the GPS receiver.

Most customisations included 'notes' or 'description' fields, where free text could be entered describing the record, or a part of it. One project, for example, separated description of features from observations about associated materials to make it clearer what was being described or qualified. Descriptive metadata could also be segregated into different kinds of notes. The same

project divided feature 'description' from 'interpretation' (and provided a third 'comments' field for text that did not fit either category), thus separating objective observations from subjective interpretations. Metadata at the record- or sub-record level and divided by the type of commentary helped researchers find the specific contextual or clarifying information they needed during data selection or analysis, and could serve a similar purpose during data reuse.

Automated creation of record-level metadata. Since automated generation of (meta)data improves quality without burdening the fieldworker, field data capture systems should automate (meta)data creation whenever possible (Pascoe, Morse and Ryan 1998; Ryan, Pascoe and Morse 1999). All FAIMS Mobile records included, for example, the automatic generation of 'Author' and 'Timestamp', documenting who created or edited a record and when. The Author field stored the name of whomever was logged in. Timestamp recorded the date, time, and time zone in a standard format captured from the device clock (checked against GPS time when the application was launched with a warning issued if the two did not match). Author and timestamp could be displayed read-only in the mobile application.

Automation could also be applied elsewhere in a customisation. We have discussed the use of automation to improve data collection during kinetic fieldwork elsewhere (Sobotkova et al. 2021: 14–15, 19). It can also be employed for metadata creation. In our experience, for example, users are reluctant to keep entering the same (meta)data repeatedly, especially if values do not change much from one record to the next. During pedestrian surface survey, for example, agricultural conditions, surface visibility, slope, ease of passability, and other environmental information might repeat many times when a team surveyed a large field. On paper forms, a 'same as previous' checkbox might indicate that values repeat. FAIMS Mobile could automatically populate fields from data entered into the previous record or insert data from defaults entered elsewhere in the application. Users deployed such automation enthusiastically where values tended to repeat.

Record-level metadata and multimedia files. Any (meta)data acquired as part of a record could also be written to the file name and/or metadata of an associated file. Photos taken of a feature during archaeological survey, for example, could have the record identifier, creator, time, date, location, feature type, associated materials, and notes written to the image EXIF metadata or an XMP sidecar file when the photos were exported from FAIMS. Writing record identifiers to filenames or file metadata helped ensure appropriate associations. This feature was very popular since it mitigated the need for manual file renaming or metadata creation and achieved FAIR goals by ensuring that metadata was preserved along with the digital object.

6.1.2 Field- or value-level metadata

FAIMS Mobile could also attach metadata to individual form fields, thus tying it to a particular value within a record. Two types of field-level metadata could be added: a numeric estimate of certainty (expressed on a 0–1 scale set by a slider) and a free-text annotation. Certainty was provided to make it easier to flag doubts about observations or interpretations. We intended Annotation to mimic the 'margins of the page' or the 'back of the form', carrying over some of the freedom of paper into digital recording, which often trades flexibility for consistency. Certainty and Annotation were usually exported as additional values in a single tabular cell separated by pipes, so that processing was straightforward using common spreadsheet software or programs like OpenRefine.

6.1.3 Uptake of metadata-related features

Manual, record-level metadata entry was widely used. Simpler forms of automation and validation were also popular, such as copying values from the previous record or validating for completion of a field. Users directly benefited from these features, whether in data consistency and completeness or speed and efficiency in the field.

Field-level Certainty and Annotations were used unevenly from one project to the next. Annotations were put to a wide variety of uses, including asking project staff to add missing vocabulary terms, describing problems encountered during data collection, or adding information to contextualise the data collected. During observations subject to uncertainty, such as the preliminary assignment of chronological periods to artefacts examined in the field, the uncertainty slider was popular, replacing the typical question mark seen in many forms and datasets. Separating uncertainty from the data (e.g., '47 | Certainty = 50%' vs '47?') made later processing easier, especially for numeric or categorical fields. Feedback from users indicated that a simple certain/uncertain flag, however, would be more useful than the sliding scale.

The provision of help text (in the 'Infobox' associated with each field) and associated training proved important, especially for students or other novices. Without it, untrained users often recorded certainty or annotations inside data fields (using a nearby field if data entry was constrained).

The combination of record- and field-level metadata, particularly the ability to annotate records or individual values with unconstrained text, helped some projects to overcome the rote aspects of form-filling, especially for students during field schools. When thoughtfully deployed, promoted, and used, they provided space for reflective practice, including consideration of the archaeological record and the experience of being in the field (contra Caraher 2016; see also Sobotkova et al. 2021).

6.2 DATA PROVENANCE (R1.2)
6.2.1 Data history

A FAIMS Mobile capability that sets it apart from other field data collectors is its maintenance of data history, even when collecting or editing data offline. The system's append-only datastore ensured that no data could be overwritten; any edits or deletions became new records that superseded previous versions (which were hidden from view unless a user examined a record's history using the server's web interface). This data history could be reviewed, displaying who made edits and when. Questionable edits could then be re-evaluated with their authors during fieldwork to resolve any errors or misunderstandings, while memories were still fresh. Changes could be selectively undone and, as with other edits, this action created another new version of the record, leaving evidence of the reversion.

This data history also revealed the progress of recording as it played out in the field: how long it took to create records, when hesitations occurred, how the pace of fieldwork changed, or which fieldworkers recorded the most quickly and efficiently. This information helped with the interpretation of data, and with analyses of fieldwork itself.

Since data versioning imposed no burden on users, and it offered practical benefits related to data review and revision, it was well received. No one, however, requested that exports include data history, limiting its utility to practical application during the project rather than the exposure of data provenance.

6.2.2 Metadata arising from customisation

FAIMS Mobile is a generalised system that requires customisation. Since it is customised using code, this code itself becomes an important metadata artefact. Above, we discussed how the data schema definition file represented a machine-readable data model (see '5.2 Data Modelling'). Customisation also required definition of the UI, including what fields would be displayed and how they would be grouped and ordered. This definition encapsulated a field workflow and represented important process metadata (or 'paradata'; Börjesson et al. 2022). Like the data schema, machine-readable XML defined a FAIMS Mobile UI customisation. We were therefore able to create a renderer that produced a wireframe of the UI, demonstrating how machine-readable metadata can be transformed into human-readable form. We used this wireframe during the design process, and projects appreciated it as a representation of the data-capture workflow.

6.2.3 FAIMS Mobile as FAIR software

Finally, the FAIMS Project curates a module library on GitHub, including information pages for customisations that follow open-source good practice with a LICENCE and README file (e.g., the repository containing

Sobotkova, Janouchova, and Nassif-Hayes 2018). These READMEs hold software metadata, including a text description, authorship, source of funding, release date, hardware and software requirements, licence, key features, reuse potential, developer contact information, and a bank of screenshots. Intended to foster software reuse, these repositories expose project data models, workflows, and approaches. Combined with published documentation of the FAIMS Mobile platform (e.g., Ballsun-Stanton et al. 2018), such module documentation represents a step towards software FAIRification (see also UNESCO 2021; van Werkhoven et al. 2019; Whitmore and Dennis 2019), an importance component of a dataset's origin story. Having a well-documented repository of modules also allowed later adopters to find and reuse customisations through a simplified 'copy and adjust' process to suit their own workflows.

6.3 DOMAIN-RELEVANT (META)DATA STANDARDS (R1.3)

FAIMS Mobile did not integrate generic or domain metadata standards by default, although some of the manual and automated metadata creation described above produced metadata aligning with them (e.g., recording author/contributor and time of creation). In our experience, moreover, researchers are unfamiliar with metadata standards, and need both software features and expert guidance to support its implementation. None of the 40-odd projects who worked with us to customise FAIMS Mobile systematically implemented such a standard.

As in the case of project- and dataset-level metadata (see 3.3 above), standards can be applied to records, parts of records, or individual fields where possible. The ADS metadata standards (Archaeology Data Service 2022), for example, implement the Getty Research Institute approach to heritage metadata (Baca 2016). They include five types of metadata: administrative, descriptive, preservation, technical, and use, associated with either 'Objects' (usually files) or 'Archives' containing multiple Objects (see Camidge 2020 for an example of Object metadata). This standard assumes that many data are presented as files, e.g., documents or databases, but some of the proposed metadata elements could be applied at a more granular level (e.g., location, subject, and period). If datasets, subsets, or records are to receive DOIs, mandatory and recommended (at least) elements of the DataCite Metadata Schema should also be incorporated (DataCite 2021). Application of relatively accessible metadata standards like those used by the ADS and DataCite has the potential to improve data reusability in return for modest effort.

6.4 SUMMARY OF THE FAIMS EXPERIENCE MAKING DATA MORE REUSABLE

FAIMS Mobile can record **metadata** (R1) at the level of the customisation (project/dataset), record, or value in a record. The system automatically generated metadata where possible and supported its manual creation. These features were widely, if not consistently, used. **(Meta) data provenance** (R1.2) was recorded in three ways: utilisation of an append-only datastore maintained a comprehensive data history, use of XML files for customisation provided essential information about data capture processes, while the open source nature of the FAIMS Mobile platform ensured transparency. Metadata that researchers found useful was generated, but none ever systematically implemented a **domain-relevant metadata standard** (R1.3).

Metadata quality is highest when it is created alongside data, and systems can be designed to automatically retain a data history and other aspects of provenance. Metadata standards can also be implemented—if data capture systems make doing so as easy as possible, and researchers are persuaded that the additional effort is worthwhile.

7 DISCUSSION

The key areas where use of FAIMS Mobile was able to incrementally improve the FAIRness of field data include assigned and inherited identifiers (F1; I1–2), metadata (F2; R1), maintenance of a data history (R1.2), and explicit articulation of data schemas (I1) and field workflows (R1.2). Most important was the creation of rich and varied metadata at all levels: project/dataset, record, and value. Creation of metadata utilised features to improve data quality (controlled entry; validation) and efficient data entry (automation). These metadata improvements fostered both findability (especially dataset-level metadata) and reusability (especially more granular metadata). Assignment of PIDs like IGSNs in the field, mapping local controlled vocabularies to shared vocabularies, and maintaining data history are FAIR-related features that are absent or difficult to implement in other systems.

Technical implementation, however, is only part of the solution to making field data FAIRer. Having capable infrastructure can make a practice possible, and well-designed UI can make it easier, but that practice is unlikely to become widespread unless it is also normative, rewarding, and (potentially) required (Cuevas Shaw, Errington and Mellor 2022; Nosek 2019).

FAIMS Mobile was designed as a pragmatic response to the needs of field researchers wanting to create well-structured, born-digital data while offline.

Researchers wanted cleaner and more comprehensive data. They wanted their photos linked to geospatial data and structured data. They wanted to be able to see who changed data when so that errors could be corrected. They wanted automation and validation so that recording kept up with fieldwork and data were complete. They wanted exports in common formats. They wanted these things to support their own research outputs. Our efforts to implement and promote features that produced FAIRer data, like the ability to embed URIs in controlled vocabularies, saw little uptake. Researchers used our system not to produce FAIR data for others to find and reuse, but to facilitate efficient collection of quality data for their own benefit. Indeed, to the best of our knowledge, only two comprehensive datasets arising from the use of the FAIMS Mobile have been published so far (Lupack et al. 2022; Noble et al. 2018 Appendix A).

Barriers to the adoption of the FAIMS Mobile platform and production of FAIRer data arose from system complexity, time and expertise required for customisation, lack of expertise, and misaligned incentives. As researchers moved from paper to digital, they had to model data and workflows, write or (more frequently) commission the writing of definition files, and then test the resulting application—essentially a small software development project. This up-front investment was considerable, but it produced high-quality, analysis-ready data, saving even more time later in the project. Nevertheless, many prospective users found it difficult to dedicate weeks at the beginning of a project to save months at the end, a major socio-technical barrier to adoption of the platform (Sobotkova et al. 2016). Those researchers who saw this process through focused on utilitarian outcomes relevant to their own projects, particularly traditional publications. Anything that added complexity or took extra time, such as planning for or implementing the FAIR data principles, was seen as an unnecessary burden. We also found little dialogue between field archaeologists collecting data and data specialists working with Linked Open Data or FAIR data. In Nozek's (2019) culture-change terminology, FAIR data practices are not yet sufficiently normative or rewarding. In some cases they are not even possible due to a lack of expertise, or easy enough to implement in available software (including FAIMS Mobile).

For FAIR data to become common in archaeology, the software that supports its production must implement features that make its production as easy as possible for the end user. Metadata, PID, and vocabulary expertise must be available and utilised by the project. Efforts to normalise and reward the publication of FAIR data must continue, so that researchers are willing to expend the time and resources to produce it.

8 CONCLUSION

Archaeological data should be made as FAIR as possible when it is created in the field before contextual information and implicit knowledge is lost. Field data collection software can facilitate the production of FAIR data. FAIMS Mobile, an example of such software, made data more Findable (F1) by supporting allocation of unique, persistent identifiers like IGSNs and capturing key project- and dataset-level metadata (F2). It made data more Interoperable by providing a means to connect local controlled vocabularies to published vocabularies, thesauri, and ontologies through their URIs (I1, I2) and supporting qualified cross-references within the system (I3). Finally, FAIMS Mobile made data more Reusable by facilitating the creation of rich and varied record- and value-level metadata (R1; R1.3), including documentation of data provenance (R1.2).

FAIMS Mobile offers a technical foundation for producing FAIR data during field research. Experience customising nearly 70 customisations of the software on more than 40 projects, however, indicates that utilisation of this capacity was incomplete. Features that helped researchers accomplish their own goals around timely analysis and publication of traditional outputs were widely used, while features that focused on making data FAIRer to benefit others were often neglected. While improvements to field data capture software and more collaboration between archaeologists and data specialists can make FAIRification easier, the major barriers are sociotechnical. Until a disciplinary culture emerges in archaeology where production and publication of FAIR data are normal and rewarded, these activities are unlikely to happen at the necessary scale, limiting the amount of genuinely reusable data available to address our discipline's grand challenges.

ADDITIONAL FILE

The additional file for this article can be found as follows:

- **Supplemental Materials.** Key research-specific FAIMS Mobile features, capabilities, and approaches. DOI: https://doi.org/10.5334/jcaa.96.s1

FUNDING INFORMATION

COMPETING INTERESTS

The authors report that they have managed or been employed by the Field Acquired Information Management Systems (FAIMS) project, a university-based research infrastructure project. FAIMS was primarily grant-funded, but also offered customisation, feature development, and support on a fee-for-service basis through a research consultancy arrangement.

AUTHOR AFFILIATIONS

Shawn Ross orcid.org/0000-0002-6492-9025
Macquarie University, AU

Brian Ballsun-Stanton orcid.org/0000-0003-4932-7912
Macquarie University, AU

Steve Cassidy orcid.org/0000-0003-2357-9652
Macquarie University, AU

Penny Crook orcid.org/0000-0001-8509-7865
Macquarie University, AU

Jens Klump orcid.org/0000-0001-5911-6022
CSIRO, AU

Adela Sobotkova orcid.org/0000-0002-4541-3963
Aarhus University, DK

REFERENCES

Archaeology Data Service. 2022. *Guide to ADS Metadata*. Archaeology Data Service, 2022. Available at https://archaeologydataservice.ac.uk/about/ourMetadataOverview.xhtml [Last accessed 25 April 2022].

Baca, M. 2016. *Introduction to Metadata: Third Edition*. Getty Publications.

Ballsun-Stanton, B. 2018. *Manual exporter for harris matrix*. Github. https://github.com/FAIMS/harris-matrix-manual-exporter [Last accessed 24 October 2022].

Ballsun-Stanton, B and **Estephan, P.** 2014. *ui_logic.bsh, FAIMS/Boncuklu*. Github. https://github.com/FAIMS/Boncuklu/blob/960b294ccc7f15ee6274f0ab430fcbb69bdd2828/ui_logic.bsh [Last accessed 24 October 2022].

Ballsun-Stanton, B and **Heřmánková, P.** 2021. *KML Exporter, FAIMS/kmlExporter*. Github. https://github.com/FAIMS/kmlExporter [Last accessed 24 October 2022].

Ballsun-Stanton, B, Ross, SA, Sobotkova, A and **Crook, P.** 2018. FAIMS Mobile: Flexible, open-source software for field research. *SoftwareX*, 7: 47–52. DOI: https://doi.org/10.1016/j.softx.2017.12.006

Börjesson, L, Sköld, O, Friberg, Z, Löwenborg, D, Pálsson, G and **Huvila, I.** 2022. Re-purposing Excavation Database Content as Paradata. *KULA: Knowledge Creation, Dissemination, and Preservation Studies*, 6(3): 1–18. DOI: https://doi.org/10.18357/kula.221

Camidge, K. 2020. *Colossus dive trail maintenance and Wheel Wreck dating 2019*. Archaeology Data Service, 2020. Available at https://archaeologydataservice.ac.uk/archives/view/colossusww_he_2020/ [Last accessed 7 August 2022].

Caraher, W. 2016. Slow Archaeology: Technology, Efficiency, and Archaeological Work. In: Averett, EWA, Gordon, JM and Counts, DB (eds.). *Mobilizing the Past for a Digital Future: The Future of Digital Archaeology*. Grand Forks, ND: The Digital Press at the University of North Dakota. pp. 421–441. DOI: https://doi.org/10.31356/dpb008

Cuevas Shaw, L, Errington, TM and **Mellor, DT.** 2022. Toward open science: Contributing to research culture change. *Science Editor*, 14–17. DOI: https://doi.org/10.36591/SE-D-4501-14

DataCite. 2021. *DataCite Metadata Schema 4.4*. DataCite Schema, 30 March 2021. Available at http://schema.datacite.org/meta/kernel-4.4/ [Last accessed 21 April 2022].

DCMI Usage Board. 2020. *DCMI Metadata Terms*. Dublin Core: Metadata Innovation, 20 January 2020. Available at http://dublincore.org/specifications/dublin-core/dcmi-terms/2020-01-20/ [Last accessed 19 April 2022].

Digital Curation Centre. 2010. *Lifecycle Model FAQ|Digital Curation Centre (DCC)*. Continuing Education on New Data Standards & Technologies, 2010. Available at https://www.dcc.ac.uk/faq/dcc-curation-lifecycle-model [Last accessed 16 March 2022].

FAIMS Project. 2015. *Attribute Format String*. FAIMS Data, UI and Logic Cook-Book – FAIMS 2.6 User and Module Documentation, 2015. Available at https://faims2-documentation.readthedocs.io/en/latest/FAIMS+Data%2C+UI+and+Logic+Cook-Book/ [Last accessed 21 March 2022].

FAIMS Project. 2016. *init.sql*. Github. https://github.com/FAIMS/faims-web/blob/master/lib/assets/init.sql [Last accessed 24 October 2022].

GO-FAIR. 2017. *FAIR Principles – GO FAIR*. GO FAIR, 2017. Available at https://www.go-fair.org/fair-principles/ [Last accessed 29 March 2019].

Kintigh, K. 2006. The Promise and Challenge of Archaeological Data Integration. *American Antiquity*, 71(3): 567–578. DOI: https://doi.org/10.2307/40035365

Kintigh, KW, Altschul, JH, Beaudry, MC, Drennan, RD, Kinzig, AP, Kohler, TA, Limp, WF, Maschner, HDG, Michener, WK, Pauketat, TR, Peregrine, P, Sabloff, JA, Wilkinson, TJ, Wright, HT and **Zeder, MA.** 2014. Grand Challenges for Archaeology. *American Antiquity*, 79(1): 5–24. DOI: https://doi.org/10.7183/0002-7316.79.1.5

Klump, J, Lehnert, K, Ulbricht, D, Devaraju, A, Elger, K, Fleischer, D, Ramdeen, S and **Wyborn, L.** 2021. Towards Globally Unique Identification of Physical Samples: Governance and Technical Implementation of the IGSN Global Sample Number. *Data Science Journal*, 20(1): 33. DOI: https://doi.org/10.5334/dsj-2021-033

Lindsay, I and **Kong, NN.** 2020. Using the ArcGIS Collector Mobile App for Settlement Survey Data Collection in Armenia. *Advances in Archaeological Practice*, 1–15. DOI: https://doi.org/10.1017/aap.2020.26

Lupack, S, Ross, SA, Sobotkova, A, Heřmánková, P and **Kasimi, P.** 2022. *Surface Survey and Legacy Data in the Upper Plain of the Heraion of Perachora: The Perachora Peninsula Archaeological Project 2020 [Digital Supplement]*. DOI: https://doi.org/10.5281/zenodo.6856524

Network Working Group. 2005. *RFC 4122 – A Universally Unique IDentifier (UUID) URN Namespace*.

Niven, K. 2011a. *Planning for the Creation of Digital Data*. Archaeology Data Service/Digital Antiquity Guides to Good Practice, 2011. Available at https://guides.archaeologydataservice.ac.uk/g2gp/CreateData_1-0 [Last accessed 21 March 2022].

Niven, K. 2011b. *Project Metadata*. Archaeology Data Service/Digital Antiquity Guides to Good Practice, 2011. Available at https://guides.archaeologydataservice.ac.uk/g2gp/CreateData_1-2 [Last accessed 14 June 2021].

Noble, R, González-Álvarez, I, Reid, N, Krapf, C, Pinchand, T, Cole, D, Lau, I, Fox, D, Petts, A, Klump, J, White, A and **Brant, F.** 2018. *Regional Geochemistry of the Coompana Area. Department for Energy and Mining, South Australia/ CSIRO* Report Book 2018/00036. DOI: https://doi.org/10.25919/5c59cf28d6fd2.

Noble, R, Reid, N, Klump, J, Robertson, J, Cole, D, Fox, D, Pinchand, T, González-Álvarez, I, Krapf, C and **Lau, I.** 2020. Testing a rapid sampling and analysis workflow in the remote Nullarbor Plain, Australia. *Explore: Newsletter for the Association of Applied Geochemists*, 2020(186): March, 1; 6–18. HDL: http://hdl.handle.net/102.100.100/342645.

Noble, R, Reid, N, Klump, J, White, A and **Cole, D.** 2017. *Coompana field sampling data*. DOI: https://doi.org/10.4225/08/5A1E01E1796DD

Nosek, B. 2019. *Strategy for Culture Change*. Center for Open Science, 11 June 2019. Available at https://www.cos.io/blog/strategy-for-culture-change [Last accessed 3 May 2022].

Pascoe, J, Morse, D and **Ryan, N.** 1998. Developing personal technology for the field. *Personal Technologies*, 2(1): 28–36. DOI: https://doi.org/10.1007/BF01581844

Ross, SA, Ballsun-Stanton, B, Sobotkova, A and **Crook, P.** 2015. Building the Bazaar: Enhancing Archaeological Field Recording Through an Open Source Approach. In: Wilson, AT and Edwards, B (eds.). *Open Source Archaeology: Ethics and Practice*. Warsaw, Poland: De Gruyter Open. pp. 111–129.

Ross, SA, Sobotkova, A, Ballsun-Stanton, B and **Crook, P.** 2013. Creating eresearch Tools For Archaeologists: The Federated Archaeological Information Management Systems Project. *Australian Archaeology*, 77(1): 107–119. DOI: https://doi.org/10.1080/03122417.2013.11681983

Ryan, N, Pascoe, J and **Morse, D.** 1999. Fieldnote: A handheld information system for the field. In: *1st International Workshop on TeloGeoProcessing*. 1999. p.

Sobotkova, A, Ballsun-Stanton, B, Ross, S and **Crook, P.** 2015. Arbitrary Offline Data Capture on All of Your Androids: The FAIMS Mobile Platform. In: Traviglia, A (ed.). *Across Space and Time. Papers from the 41st Annual Conference of Computer Applications and Quantitative Methods in Archaeology (CAA)*. 2015. Amsterdam University Press. pp. 80–88.

Sobotkova, A, Hermankova, P and **Nassif-Haynes, C.** 2018. *Tundzha Regional Archaeology Project (TRAP) Map Digitisation* (version fc0b21c). V2.5 (Android 6+). Github. https://github.com/FAIMS/map-digitisation/releases/tag/v1.1 [Last accessed 24 October 2022].

Sobotkova, A, Hermankova, P and **Nassif-Haynes, C.** 2020a. *Perachora Peninsula Archaeological Project (PPAP) Gridded Pedestrian Survey* (version 0e39009). FAIMS v2.6 (Android 9+). Github. https://github.com/FAIMS/trap-gridded-survey/releases/tag/v1.1 [Last accessed 24 October 2022].

Sobotkova, A, Hermankova, P and **Nassif-Haynes, C.** 2020b. *Perachora Peninsula Archaeological Project (PPAP) Legacy Data Verification* (version cced016). FAIMS v2.6 (Android 9+). Github. https://github.com/FAIMS/Perachora-2020/releases/tag/v1.1 [Last accessed 24 October 2022].

Sobotkova, A, Ross, SA, Ballsun-Stanton, B, Fairbairn, A, Thompson, J and **VanValkenburgh, P.** 2016. Measure Twice, Cut Once: Cooperative Deployment of a Generalized, Archaeology-Specific Field Data Collection System. In: Averett, EW, Gordon, JM and Counts, DB (eds.). *Mobilizing the Past for a Digital Future: The Potential of Digital Archaeology*. Grand Forks, ND: The Digital Press @ University of North Dakota. pp. 337–371. DOI: https://doi.org/10.31356/dpb008

Sobotkova, A, Ross, SA, Hermankova, P, Lupak, S, Nassif-Haynes, C, Ballsun-Stanton, B and **Kasimi, P.** 2021. Deploying an offline, multi-user mobile system for digital recording of landscape archaeology in the Perachora Peninsula, Greece. *Journal of Field Archaeology*, 46(8): 571–594. DOI: https://doi.org/10.1080/00934690.2021.1969837

Thorne, R, Reid, N, Gray, D, Ballsun-Stanton, B, Bardwell, N, Klump, J, Davis, A, Ross, S and **Sobotkova, A.** 2018. *M436 Distal Footprints: UNCOVER Australia – Hydrogeochemistry of the Capricorn Orogen*. DOI: https://doi.org/10.25919/5cddb339279ec

UNESCO. 2021. *UNESCO Recommendation on Open Science*. UNESCO. ARK: https://unesdoc.unesco.org/ark:/48223/pf0000379949.

Van Valkenburgh, P, Sobotkova, A, Ross, SA and **Ballsun-Stanton, B.** 2015. *PAZC: Proyecto Arqueològico Zaña Colonial/Zaña Colonial Archaeology Project in Peru, excavation*. (version 3270444). FAIMS v2.0 (Android 5+). Github. https://github.com/FAIMS/PAZC/releases/tag/v1.0 [Last accessed 24 October 2022].

van Werkhoven, B, Meakin, J, Lamprecht, A-L and **Pablo, DSR-S.** 2019. *FAIR Software at the 2019 eScience Symposium.* Software Sustainability Institute, 5 December 2019. Available at https://www.software.ac.uk/blog/2019-12-05-fair-software-2019-escience-symposium [Last accessed 21 March 2022].

Wallis, JC, Borgman, CL, Mayernik, MS and **Pepe, A.** 2008. *Moving Archival Practices Upstream: An Exploration of the Life Cycle of Ecological Sensing Data in Collaborative Field Research*, 3(1): 114–126. DOI: https://doi.org/10.2218/ijdc.v3i1.46

Whitmore, D and **Dennis, T.** 2019. *Top 10 FAIR Data & Software Things: Archaeology*. Library Carpentry, 2019. DOI: https://doi.org/10.5281/zenodo.2555498

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