



MACQUARIE
University
SYDNEY · AUSTRALIA

Seven years of FAIMS Mobile

Shawn A Ross

Office of the Deputy Vice-Chancellor (Research)

August 23, 2019



‘Small data’ infrastructure across the data lifecycle

Lessons from FAIMS: summary

Lessons from FAIMS: in-depth

From current practice to better practice

References



MACQUARIE
University
SYDNEY · AUSTRALIA

‘Small data’ infrastructure across the data lifecycle



Figure 1: Archaeologists contemplate data standards (FAIMS Stocktaking, 2012)

'Long tail' research: most field data is small data [Borgman, 2015]

- Smaller scale; smaller communities; local control.
- Diverse questions, approaches, and methods.
- Heterogeneous data; variety of content, structure.
- Data and infrastructure emerge from fieldwork.
- Relative lack of standards.
- Limited infrastructure and funding.
- Challenges associated with big(ger) data from photogrammetry, SfM, video, geophysics, etc., will exacerbate these problems.

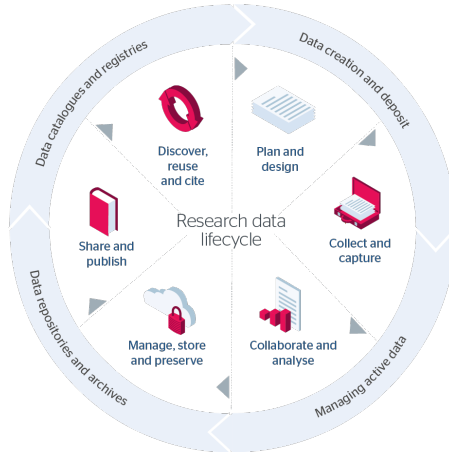


Figure 2: [JISC, 2018] Image CC-BY-ND

Consider the infrastructure needed to manage the three main phases of the data lifecycle

- Publication (most mature): domain-specific repositories.
- Processing and analysis (less mature): project-level code [Stewart Lowndes et al., 2017], then Virtual Labs / Science Gateways, like [Alveo, 2019] in language analysis.
- Capture (least mature): most varied, needs to work offline under difficult conditions. Commercial solutions insufficient [Bureau of Reclamation, 2017].



MACQUARIE
University
SYDNEY • AUSTRALIA

Lessons from FAIMS: summary

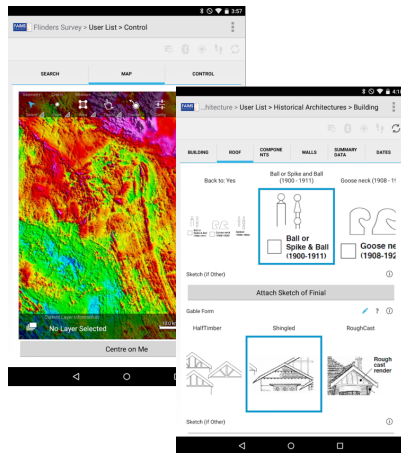


Figure 3: FAIMS Mobile: GIS and 'picture dictionaries'

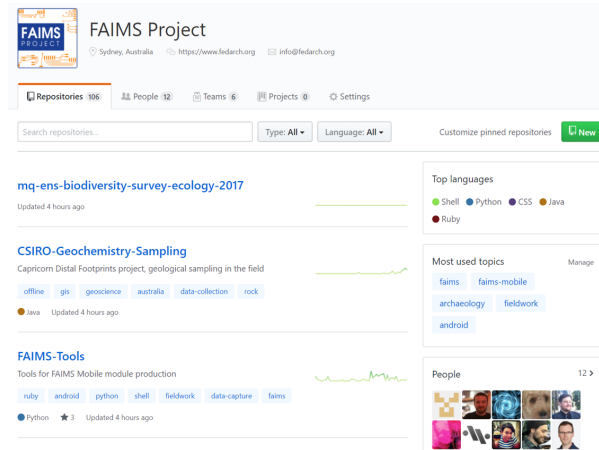


Figure 4: FAIMS Mobile customisations on GitHub

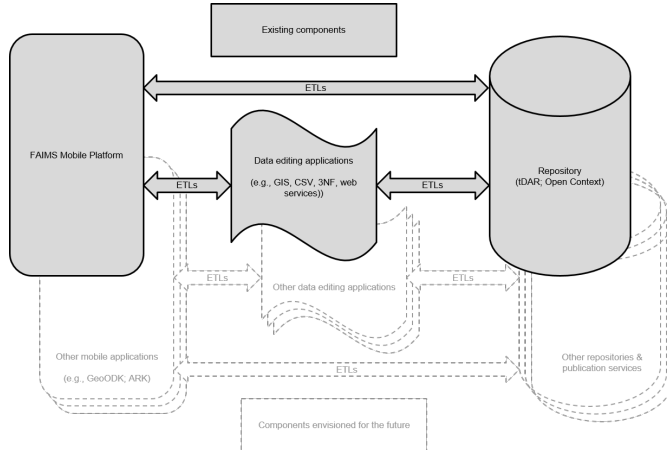


Figure 5: FAIMS Mobile federation



Figure 6: FAIMS Mobile 'core' code is GPLv3



- We deserve research-specific software.
- Diverse practices and limited resources require generalised software.
- Do one thing well with modular and federated software (but slice the pie thoughtfully).
- Open-source software has advantages (but is difficult to sustain).
- Scope requirements carefully.
- Invest in outreach and engagement.

How do we get from where we are now to where we want to be?

- Understand the evolving expectations of transparent research.
 - Look past desktop software (Excel, ARCGIS, Filemaker, Access, etc.).
 - Rally around emerging research- and domain-specific solutions (even if imperfect).
 - Overcome ‘not invented here’; you don’t need a bespoke solution.
 - Budget for ‘ground-up’ transparency (data and code). Up-front costs will be high but offer longer-term payoffs (in costs, time, and quality).
 - Implement (and budget for) fundamental good practice in data and code management before other technologies.
 - Improve research design (prioritise approach over methods)
- [Muthukrishna and Henrich, 2019, Hole, 1973]



MACQUARIE
University
SYDNEY • AUSTRALIA

Lessons from FAIMS: in-depth

- The Field Acquired Information Management Systems (FAIMS) Project began in 2012 as a national Australian information infrastructure project in archaeology.
- Developed FAIMS Mobile for field data capture [Ballsun-Stanton et al., 2018].
- Use expanded beyond archaeology to geoscience, ecology, ethnography, linguistics, oral history.
- Has been customised for over 50 workflows at more than 30 projects.
- Data and workflow modelling for these customisations provided deep insights into field data capture and the infrastructure needed to support it.

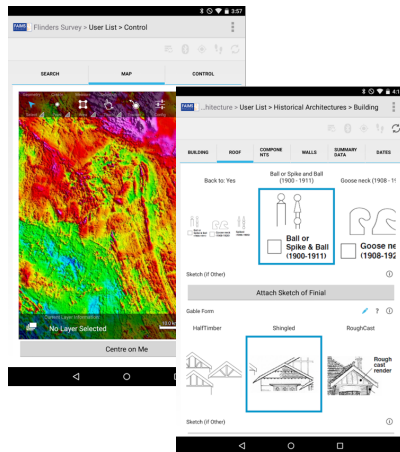


Figure 7: FAIMS Mobile: GIS and 'picture dictionaries'

- Fundamentally customisable.
- Tightly binds structured, geospatial, multimedia, and free text data.
- Works offline.
- Automated bi-directional synchronisation using local or online server
- Record history: append-only datastore, versioning, rollback.
- Mobile GIS.
- Connects to internal and external sensors, Bluetooth / USB devices.
- Multilingual.
- Granular help.
- Granular metadata / uncertainty.
- Generalised export.
- 'Hooks' for data interoperability, Open Linked Data approaches.



- We deserve research-specific software.
- Diverse practices and limited resources require generalised software.
- Do one thing well with modular and federated software (but slice the pie thoughtfully).
- Open-source software has advantages (but is difficult to sustain).
- Scope requirements carefully.
- Invest in outreach and engagement.

Archaeology needs (and deserves) research-specific software, contra [Roosevelt et al., 2015].

- Most commercial / mass-market software does not meet research needs.
- Risk of lock-in, unwelcome changes to features or business models, and product discontinuation.

Compare ecology in Australia: TERN, ALA, Biocollect, and associated research clouds [TERN, 2019, ALA, 2019a, ALA, 2019b].

Commercial software doesn't meet our needs, and bespoke development is too expensive and usually unsustainable.

- Generalised software can be deeply customised to accommodate our diverse data types, data models, workflows, etc.
- The code used to customise it describes the data model and workflow.
- Customisations can be published and re-deployed trivially.
- Can deliver research-grade software affordably.

FAIMS Mobile cost perhaps 3x a single bespoke application, but has been customised 50x. Customisation cost is 1/10th bespoke, and still $<1/2$ even if 'core' platform development costs are amortised across projects.

Generalised: customise using code

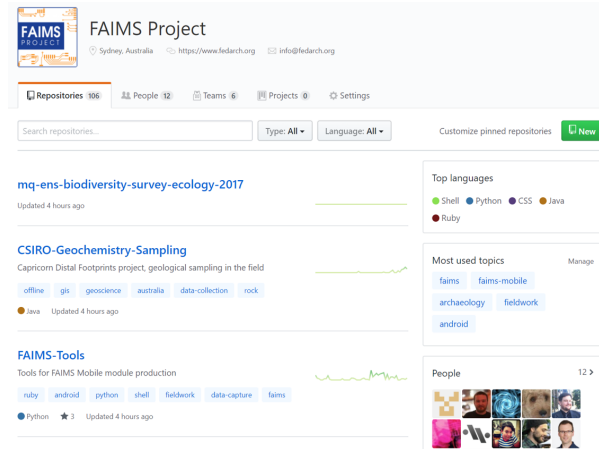


Figure 8: FAIMS Mobile customisations (XML files, mostly) on GitHub

Do one thing well.

- Identify other infrastructure in the domain and interoperate with it (via ETLs or APIs).
- It is better to divide by data-lifecycle phase rather than data type, since (1) our data is so integrated and (2) field data capture poses unique challenges.

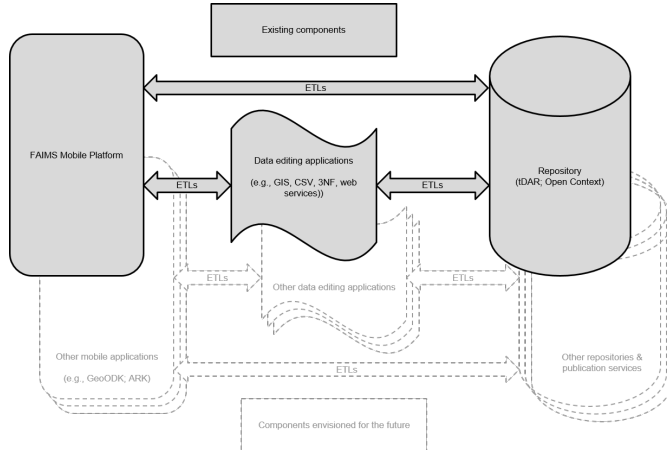


Figure 9: FAIMS Mobile federation strategy

Open source has advantages but is difficult to sustain.

- Emerging open research principles strongly prefer OSS as opposed to proprietary 'black boxes'.
- Transparency and reusability (esp. customisation code).
- Ability to hand off from one organisation to another (esp. 'core' platform code).
- Ability to fork code prevents lock-in and mitigates unwelcome decisions by software developers.
- BUT OSS business models are hard to scale and rely on occasional injections of grant or institutional funding.

Talk to a wide range of potential users, seeking facts not opinions.

- Don't ask researchers what they think, ask them what they have done
- what software they have adopted and why, and what problems they have expended resources to solve.
- 'Lean startup' methodology very useful, based around testing of ideas through interviews with potential users [Strategyzer AG, 2019].
- In our case, we over-invested in mobile GIS and under-invested in usability (especially a GUI for customisation).

If you build it they will not come; people can't use technologies they don't know about.

- As per industry standards, dedicate at least 30% of any information infrastructure budget to outreach and engagement (sales and marketing).
- Typical academic outreach (journal articles, conference presentations, workshops, even booths at major conferences) are not enough.

- [Sobotkova, 2018]
- [Ballsun-Stanton et al., 2018]
- [VanValkenburgh et al., 2018]
- [Sobotkova et al., 2016]
- [Ross et al., 2015]
- [Sobotkova et al., 2015]
- [Ross et al., 2013]



MACQUARIE
University
SYDNEY · AUSTRALIA

From current practice to better practice

How do we get from where we are now to where we want to be?

- Understand the evolving expectations of transparent research.
 - Look past desktop software (Excel, ARCGIS, Filemaker, Access, etc.).
 - Rally around emerging research- and domain-specific solutions (even if imperfect).
 - Overcome ‘not invented here’; you don’t need a bespoke solution.
 - Budget for ‘ground-up’ transparency (data and code). Up-front costs will be high but offer longer-term payoffs (in costs, time, and quality).
 - Implement (and budget for) fundamental good practice in data and code management before other technologies.
 - Improve research design (prioritise approach over methods)
- [Muthukrishna and Henrich, 2019, Hole, 1973]



MACQUARIE
University
SYDNEY • AUSTRALIA

References



ALA (2019a).
Atlas of living australia.
<https://www.ala.org.au/>.
Accessed: 2019-4-24.



ALA (2019b).
BioCollect: Advanced data collection for biodiversity projects.
<https://www.ala.org.au/biocollect/>.
Accessed: 2019-4-24.



Alveo (2019).
Alveo: A virtual lab for human communication science.
<http://alveo.edu.au/>.
Accessed: 2019-4-24.



Ballsun-Stanton, B., Ross, S. A., Sobotkova, A., and Crook, P. (2018).
FAIMS mobile: Flexible, open-source software for field research.
SoftwareX, 7C:47–52.



Borgman, C. L. (2015).
Big data, little data, no data: scholarship in the networked world.
MIT press.



Bureau of Reclamation (2017).
DataApp: A mobile app framework for field data capture.

<https://www.usbr.gov/research/challenges/dataapp.html>.
Accessed: 2018-3-27.



Hole, F. (1973).
Questions of theory in the explanation of culture change in prehistory.
In Renfrew, C., editor, *The Explanation of Culture Change: Models in Prehistory*, pages 19–34. Duckworth.



JISC (2018).
Research data management toolkit.
<https://rdmtoolkit.jisc.ac.uk/research-data-lifecycle/>.
Accessed: 2018-11-6.



Muthukrishna, M. and Henrich, J. (2019).
A problem in theory.
Nature human behaviour, 3(3):221–229.



Roosevelt, C. H., Cobb, P., Moss, E., Olson, B. R., and Ünlüsoy, S. (2015).
Excavation is destruction digitization: Advances in archaeological practice.
Journal of Field Archaeology, 40(3):325–346.



Ross, S. A., Ballsun-Stanton, B., Sobotkova, A., and Crook, P. (2015).

Building the bazaar: Enhancing archaeological field recording through an open source approach.

In Wilson, A. T. and Edwards, B., editors, *Open Source Archaeology: Ethics and Practice*, pages 111–129. De Gruyter Open, Warsaw, Poland.



Ross, S. A., 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.



Sobotkova, A. (2018).

Sociotechnical obstacles to archaeological data reuse.
Advances in Archaeological Practice, 6(2):117–124.



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., editor, *Across Space and Time. Papers from the 41st Annual Conference of Computer Applications and Quantitative Methods in Archaeology (CAA)*, pages 80–88. Amsterdam University Press.



Sobotkova, A., Ross, S. A., 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, E. W., Gordon, J. M., and Counts, D. B., editors, *Mobilizing the Past for a Digital Future: The Potential of Digital Archaeology*, pages 337–371. The Digital Press @ University of North Dakota, Grand Forks, ND.



Stewart Lowndes, J. S., Best, B. D., Scarborough, C., Afflerbach, J. C., Frazier, M. R., O'Hara, C. C., Jiang, N., and Halpern, B. S. (2017).

Our path to better science in less time using open data science tools.

Nature Ecology & Evolution, 1(6):s41559–017–0160.



Strategyzer AG (2019).

Strategyzer | trusted by over 5 million business practitioners.

<https://www.strategyzer.com/>.

Accessed: 2019-4-24.



TERN (2019).

TERN - terrestrial ecosystem research network.

<https://www.tern.org.au/>.

Accessed: 2019-4-24.



VanValkenburgh, P., Silva, L. O. G., Repetti-Ludlow, C., Gardner, J., Crook, J., and Ballsun-Stanton, B. (2018).

Mobilization as mediation: Implementing a Tablet-Based recording system for ceramic classification.

Advances in Archaeological Practice, pages 1–15.

Source code for this presentation is available at: <https://github.com/saross/Ross-FAIMS-current>.

FAIMS Project software and documentation can be found at:
<https://github.com/faims>.

This work is licensed under a Creative Commons Attribution 4.0 International License.