TECHNICAL NOTE

A new digital field data collection system for dendrochronology

Peter W. Brewer*, Christopher H. Guiterman

Laboratory of Tree-Ring Research, University of Arizona, 1215 E. Lowell Street, Tucson, AZ 85721-0045, USA

A R T I C L E   I N F O

A B S T R A C T

A wide variety of information or 'metadata' is required when undertaking dendrochronological sampling. Traditionally, researchers record observations and measurements on field notebooks and/or paper recording forms, and use digital cameras and hand-held GPS devices to capture images and record locations. In the lab, field notes are often manually entered into spreadsheets or personal databases, which are then sometimes linked to images and GPS waypoints. This process is both time consuming and prone to human and instrument error. Specialised hardware technology exists to marry these data sources, but costs can be prohibitive for small scale operations (>$2000 USD). Such systems often include proprietary software that is tailored to very specific needs and might require a high level of expertise to use. We report on the successful testing and deployment of a dendrochronological field data collection system utilising affordable off-the-shelf devices ($100–300 USD). The method builds upon established open source software that has been widely used in developing countries for public health projects as well as to assist in disaster recovery operations. It includes customisable forms for digital data entry in the field, and a marrying of accurate GPS location with geotagged photographs (with possible extensions to other measuring devices via Bluetooth) into structured data fields that are easy to learn and operate. Digital data collection is less prone to human error and efficiently captures a range of important metadata. In our experience, the hardware proved field worthy in terms of size, ruggedness, and dependability (e.g., battery life). The system integrates directly with the Tellervo software to both create forms and populate the database, providing end users with the ability to tailor the solution to their particular field data collection needs.

© 2016 Elsevier GmbH. All rights reserved.

1. Introduction

When samples are collected for dendrochronological research it is necessary to record associated metadata. At the very least this will include identifiers for the sample (typically site, tree, and sample codes) but to adequately address most research questions there will be many others.

Traditionally, data collected in the field are recorded via hand-written notes in field notebooks or paper recording forms. Inevitably in such a digitally-enabled world, this information is then transferred to computers on return to the lab, with field notes typically entered into spreadsheets, databases and or text documents. This procedure is time-consuming, prone to human error and not conducive to collecting any more field data than is strictly necessary to answer the research questions being directly asked by the researcher. Researchers will also often use digital cameras and GPS devices to capture images and location coordinates. The separation of camera, GPS, and field notes means work is required to connect these separate pieces of information; for example, by downloading and renaming image files and waypoints and then manually entering these data into a database.

The desire for an effective, integrated digital field data collection system is shared by researchers working in many disciplines. While there has been high-end field data collection hardware available for some time, both the high cost and the high-level of expertise required to customise the software for a specific need has meant that data collection systems have rarely been deployed across a research discipline, such as for dendrochronology. The introduction of consumer-grade smart-phones and tablets in recent years has provided an alternative. Rapid technological advances have led to mobile data collection being easier to access, more efficient, and less expensive. These devices are nearly ubiquitous in developed countries, with user interfaces that are extremely intuitive to use. Researchers and organisations in many fields including public health (Morris, 2009), seismology (Kwok, 2009), epidemiology (Aanensen et al., 2009), and psychology (Rachuri et al., 2010) have successfully developed and deployed easy to learn mobile applications (apps).

The system described here combines a range of open source technologies including Open Data Kit (Hartung et al., 2010), the

* Corresponding author. Tel.: +1 5206210753.
E-mail address: p.brewer@ltrr.arizona.edu (P.W. Brewer).

Tree Ring Data Standard [TRiDaS; Jansma et al., 2010] and Tellervo (Brewer, 2016, 2014) to provide a customisable and intuitive method for dendrochronological field data collection. The system is available in Tellervo from version 1.2.1 and includes the following: (1) customisable field data form creation for objects (e.g., site-level attributes) and elements (e.g., tree and sample level attributes); (2) use of the Open Data Kit 'ODK Collect' app on Android devices; (3) upload of completed data forms and media files to Tellervo server from ODK Collect; and (4) metadata error checking in Tellervo, with extensions for compiling and renaming media files and output of a comma-separated (.csv) file that includes all field data and links to images.

1.1. Open Data Kit

ODK is an ecosystem of tools designed to enable data collection via mobile devices. The project was initiated by a team from the University of Washington and was initially aimed at supporting socio-economic and health surveys in developing regions (Hartung et al., 2010). It has since been used by a wide variety of projects all over the world. ODK is open source and based on established open data standards such as XForms (specifically the OpenRosa subset) and HTTP protocols for the transmission of data. A similar project called KoBoToolbox, based on the same standards (therefore compatible with ODK) has also been used successfully by many projects.

ODK is comprised of: ODK Build (a tool for designing data collection forms), ODK Aggregate (a server for managing forms and data), ODK Briefcase (a standalone offline application alternative to ODK Aggregate), and ODK Collect (an application for collecting field data on a phone or tablet). The open standards and the separation into distinct tools mean that users can select the tools that suit their needs, use compatible products from other developers, or develop their own compatible tools.

The focus of ODK development has been to enable diverse teams of data collectors to contribute to central databases, from where data can be analysed and reports generated. The generic nature of the system has been successful in enabling a diverse range of surveys and research to be completed. This flexibility can be leveraged in the "dendro community" to generate field data collection forms tailored to specific studies. At the same time, this flexibility also comes with the drawback of lack of standardisation of data. Different users collecting similar data in different ways produce a significant challenge in combining data from multiple studies: a problem that the Tree Ring Data Standard (TRiDaS) was designed to address.

1.2. Tree Ring Data Standard (TRiDaS)

TRiDaS is an open data standard designed to describe dendrochronological data and metadata from all sub-disciplines of dendrochronology. By standardising the way data and metadata are collected, it becomes easier to compare and combine data from different studies. This is useful not only when combining data collected for similar purposes by different research teams, but also for utilising data collected for quite different research. For example, although there are many differences in the data-collection needs of researchers using dendrochronology for archaeology, ecology, climatology and art history, there are also many similarities. By ensuring data fields that are common to multiple sub-disciplines are stored in the same way, data become far more reusable.

1.3. Tellervo

The TRiDaS data model is used within Tellervo. Tellervo is an open source dendrochronological system for the measuring and curation of dendrochronological samples. It is an enterprise-style tool designed for medium to large laboratories, enabling multiple users to run the Tellervo desktop application to store and share data in a shared Tellervo server database. Tellervo includes features such as: support for a variety of measuring platform hardware; barcode-based curation of samples; integrated 3D mapping; and webservice communications enabling users to securely access data across the Internet, enabling real-time collaboration between users at different institutions. The newest feature in Tellervo, and the focus of this paper, is the implementation of ODK support. This provides users with an efficient, integrated method for collecting and storing metadata.

2. Tellervo-ODK workflow

While it is possible to use the ODK Build, Aggregate, Briefcase, and Collect tools together unaltered, transferring data from either ODK Aggregate or Briefcase to the Tellervo server proves time consuming and requires technical expertise. When generating new forms, users would need to take care to design fields that follow the TRiDaS nomenclature and structure, and then manually import to the Tellervo database. The workflow proposed here is to use the ODK Collect tool but replace the functionality of ODK Build, Aggregate and Briefcase with the newly implemented equivalents directly in Tellervo. A summary of the Tellervo-ODK workflow is outlined below:

1. Design an ODK data collection form using the Tellervo ODK form builder.
2. Submit the finished form to the user's own Tellervo server.
3. On the user's Android-powered tablet or phone, install ODK Collect and configure it to access the user's Tellervo server.
4. Within ODK Collect, download the form definition created in step 1.
5. Fill in the form for each study site, and/or dendro sample collected.
6. At a convenient time upload the collected field data (either via the mobile data network or a standard Wifi connection) to the user's Tellervo server.
7. In Tellervo desktop, access the field data stored on the Tellervo server. Check and edit for accuracy and confirm import into the Tellervo database. Optionally export data to CSV files.
8. Automatically generate barcodes for new samples and proceed with measuring and analysis as normal.

3. Designing data collection forms

The Tellervo ODK form builder (Fig. 1) is accessed from the file menu in Tellervo and replaces the functionality of the ODK Build application. It is configured to generate two types of forms: one for recording information about sites (or 'objects' in TRiDaS terminology); the other for recording information about trees and samples (or 'elements' and 'samples' in TRiDaS terminology). There are only a few required fields for each form that identify the object, element, and sample codes, and the user can then select from a variety of predefined fields to add.

Fields can be of various types including: text, numeric, predefined choices, GPS locations, photographs, and videos. The default name of each field follows the naming conventions outlined by TRiDaS. However, these names can be overridden for improved clarity, which we feel is especially useful when technicians are unfamiliar with TRiDaS names. Each field also includes a description that is displayed on the mobile device. The descriptions can be altered (or even translated) to help technicians and students.
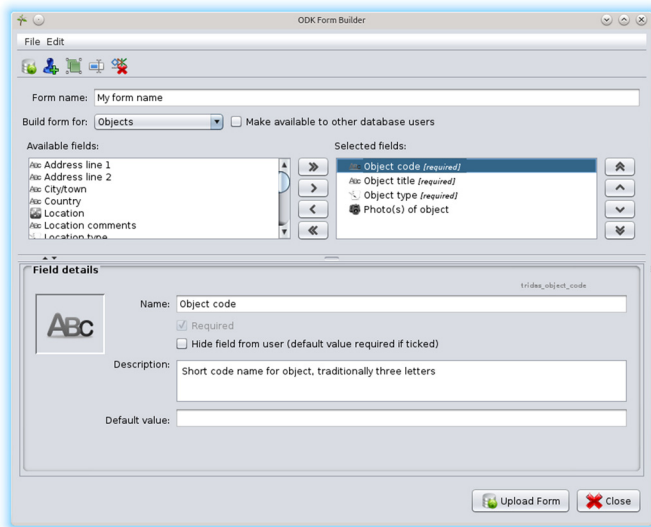
Fig. 1. Screenshot of the ODK form builder interface in Tellervo. The available fields are listed on the left and the selected fields on the right, while the bottom panel shows the details of the currently selected field.

Each field can be assigned a default value which will pre-populate the form. This feature can speed up data collection when certain items are always the same (e.g., when the element is always a tree or the object is always a site). A field can also be recorded but hidden from the user (e.g., the sampling date). These features allow for efficient and accurate collection of important metadata without the possibility of committing errors.

Some fields described by TRiDaS include 'controlled vocabularies'. These are fields with a defined list of options such as 'sample type' or 'taxon'. The Tellervo form builder automatically includes a list of possible options, but in the case of 'taxon' this is quite a long list that can be cumbersome to navigate on a mobile device. The form builder therefore provides the option to limit these lists to a subset and also to set a default value. For example, the taxon list can be reduced to a handful of species known to grow in the study area.

The Tellervo form builder also provides the ability to include additional user-defined fields. Although TRiDaS was designed to be inclusive, it is possible that some research questions may require collecting metadata not currently supported. Users should review the pre-defined fields before creating new ones, however, because duplicating already supported fields negates the purpose of data standardisation. Additionally, it should be noted that the data collected in these user-defined fields are not stored or managed by Tellervo, but are included in CSV exports for the user to manage elsewhere.

When a form definition is complete it is uploaded to the Tellervo server with the 'Upload form' button.

4. Collecting data in the field

Data collection is done using mobile devices such as a phone or tablet with the Android operating system. Other operating systems are not currently supported by ODK and thus we are limited in this regard. Forms can be downloaded to any number of devices, and the data will be integrated by Tellervo. Users should be aware, however, that element (tree) codes are not checked for duplicates by ODK and caution is urged to maintain integrity within and between devices. The freely available ODK Collect (or the compatible KoBoCollect) app should be installed from the standard
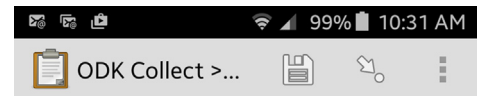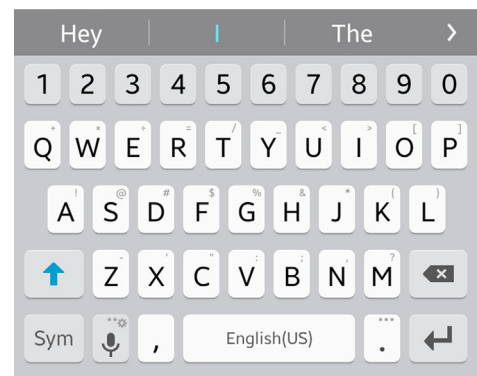
Fig. 2. Screenshot of the ODK Collect app running on an Android phone. The screen shows how the field highlighted in the form builder in Fig. 1 is displayed to the user.

marketplace. ODK Collect is designed to run on any Android-power hardware including entry-level phones running early versions of Android.

After installing and launching, go to the 'general settings' page and 'configure platform settings'. Here you should enter the URL of your Tellervo server, along with your username and password. The URL is the standard Tellervo server URL you normally use but with '/odk' appended (e.g., https://mytellervoserver.com/odk). The username will be your normal Tellervo username; however, the password will be a special ODK password you have set in the administration page of Tellervo. ODK uses standard HTTP-digest authentication which requires passwords to be stored on the server in plain text, compared to the more complex and secure challenge-response mechanism used by Tellervo. Rather than downgrade the standard authentication method used by Tellervo, a second (slightly less secure) password is used for ODK. Note that the ODK login only provides access to ODK form definitions and data, not to the rest of the Tellervo database.

Once ODK Collect is configured, you can go to 'get blank form' to access the form(s) definitions created in the Tellervo form designer. This only needs to be done once, to get the form into the ODK app. Once it is there, you can select 'fill blank form' to create a form instance for each site or tree/sample. Fields are presented one at a time within the form and a simple swipe of the screen navigates back and forth between fields (see Fig. 2 for a screenshot of a Tellervo form in ODK Collect). With some practice, users can easily fill out multiple forms simultaneously, further increasing the efficiency of data collection.

When you are ready to upload data to the server you can click the 'send finalised form' button and all form data (and associated media files such as photos) will be uploaded to your Tellervo server.

5. Importing data to the Tellervo database

ODK Collect is designed to be used in the field where network connections are sporadic or non-existent. There is therefore no capability to validate data entry errors such as duplicate sample codes. Subsequently, data uploaded from ODK Collect does not get inserted directly to the Tellervo database, but instead stored separately and readied for validation.

Within the 'bulk data entry' screen in the Tellervo desktop application there is a button to populate the data table with the ODK Collect data from the staging area on the server. This is done through a wizard interface that enables the user to request the consolidation and renaming of associated media files. It also provides the option of exporting the field data (including any user defined non-TRiDaS fields) to a CSV file. Once the field data is extracted into the bulk data entry table, it can then be validated, edited and augmented if necessary before finally being committed to the database. For new sample records, the bulk data entry screen includes the functionality to automatically generate barcode labels. These can then be applied to samples when they are prepared.

6. Limitations

A common concern when migrating from paper-based to digital-based data collection is the robustness and reliability of the device. The addition of inexpensive ruggedised, waterproof casing can help ensure the device remains undamaged in difficult conditions. With regards data safety, ODK stores data on removable flash memory cards so in the event of device failure, data should be accessible. Data can also be backed up regularly to your Tellervo server, which should in turn be backed up using enterprise-level best practice procedures. The ODK instance forms can also be copied onto a laptop in the field, thereby duplicating the raw data before importing into Tellervo. In comparison, traditional handwritten field notes cannot be easily duplicated. It is typical for the notes for an entire field excursion to be contained in one field notebook, the loss or substantial damage to which would be disastrous. Although digital field data collection techniques are not immune to data loss, neither are traditional techniques. The implementation of best-practice backup procedures mitigates the risks.

The battery life of devices has also been raised as a concern. Field testing by the authors with multiple devices has shown they are more than capable of lasting for a long day in the field. Data entry typically takes a short time for each record, leaving the device in standby mode for most of the day. For extended backcountry field excursions, the addition of solar chargers and/or external battery-based charging devices may be necessary.

The capabilities of these mobile devices to provide accurate location data require careful consideration. Studies show the accuracy of mobile phone based GPS data is typically lower than dedicated handheld GPS units (Zandbergen, 2009; Garnett and Stewart, 2015; Jones et al., 2015), with devices typically accurate in the order of 5–10 m. Whether this level of accuracy is acceptable will depend on the individual researcher and their requirements. For example, while this may be deemed accurate enough to locate a site, it is unlikely to be accurate enough to locate individual trees. A solution to this is the addition of dedicated GPS receivers connected wirelessly to the mobile device via Bluetooth. Low cost (<$100 USD) receivers will increase accuracy to around 1–2 m, while sub-meter and even 1 cm accurate receivers are available, albeit at a much higher price. An additional benefit of these dedicated receivers is that they provide coordinates immediately on request, whereas the integrated sensors in the mobile devices typically take some time

to provide a fixed location. Whichever device or combination of devices is chosen, it is prudent to run thorough tests on the accuracy of data prior to embarking on a field study.

As previously highlighted, the standalone nature of the ODK Collect app does present limitations with regards data entry errors and duplications. With no access to existing database records in the field it is possible to duplicate codes. In circumstances where multiple field teams are collecting on the same expedition there is also potential for duplication of codes while sampling. In these circumstances practical procedures such as pre-assigning codes to groups (e.g., 1–10 and 11–20, or odds and evens) mitigates such problems. Data entry errors are always a possibility regardless of whether the data is entered in the field or in the laboratory. The data validation and import stage within Tellervo described in the workflow above does provide an additional opportunity to carefully check records prior to importing into the database. The removal of any manual data entry for GPS data, as well as linking of records to media files such as photos, removes the potential for errors at these steps.

7. Conclusion

The accurate collection of field information is essential for the success of any dendrochronological study. While traditional handwritten field notes have proven fit-for-purpose over the years, the availability of smart-phones and tablets can make the process more efficient and accurate.

The method described here builds upon established software and standards to provide users with a quick, easy and cost effective field data collection procedure. The speed at which data can be recorded encourages the routine collection of rich metadata beyond the essential information required for the task at hand. It also encourages the standardisation of metadata which will enhance the possibility of data reuse and collaboration. The simple interface for generating data collection forms means that users can quickly generate forms tailored to the particular field data collection needs.

Acknowledgements

References

Aanensen, D.M., Huntley, D.M., Feil, E.J., al Own, F., Spratt, B.G., 2009. Epicollect: linking smartphones to web applications for epidemiology, ecology and community data collection. PLoS ONE 4 (9), e6968, http://dx.doi.org/10.1371/journal.pone.0006968.

Brewer, P., 2014. Data management in dendroarchaeology using Tellervo. Radiocarbon 56 (4), S79–S83, http://dx.doi.org/10.2458/azu_rc.56.18320.

Brewer, P.W., 2016. Tellervo: A Guide for Users and Developers., http://dx.doi.org/10.5281/zenodo.51329, www.tellervo.org

Garnett, R., Stewart, R., 2015. Comparison of GPS units and mobile Apple GPS capabilities in an urban landscape. Cartogr. Geogr. Inf. Sci. 42 (1), 1–8.

Hartung, C., Lerer, A., Anokwa, Y., Tseng, C., Brunette, W., Borriello, G.,2010. Open Data Kit: tools to build information services for developing regions. In: Proceedings of the 4th ACM/IEEE International Conference on Information and

Communication Technologies and Development, ACM. ACM Press, p. 18, http://dx.doi.org/10.1145/2369220.2369236.

Jansma, E., Brewer, P., Zandhuis, I., 2010. TRiDaS 1.1: the tree-ring data standard. Dendrochronologia 28 (2), 99–130, http://dx.doi.org/10.1016/j.dendro.2009.06.009.

Jones, T.W., Marzen, L., Chappelka, A., 2015. Horizontal accuracy assessment of global positioning system data from common smartphones. Pap. Appl. Geogr. 1 (1), 59–64.

Kwok, R., 2009. Personal technology: phoning in data. Nature 458 (7241), 959–961, http://dx.doi.org/10.1038/458959a.

Morris, K., 2009. Mobile phones connecting efforts to tackle infectious disease. Lancet Infect. Dis. 9 (5), 274, http://dx.doi.org/10.1016/S1473-3099(09)70118-5.

Rachuri, K.K., Musolesi, M., Mascolo, C., Rentfrow, P.J., Longworth, C., Aucinas, A., 2010. Emotionsense. In: Proceedings of the 12th ACM International Conference on Ubiquitous Computing – Ubicomp'10, http://dx.doi.org/10.1145/1864349.1864393.

Zandbergen, P.A., 2009. Accuracy of iPhone locations: a comparison of assisted GPS, WiFi and cellular positioning. Trans. GIS 13, 5–25, http://dx.doi.org/10.1111/j.1467-9671.2009.01152.x.