

# iDAI.field 2: A Modern Approach to Distributed Fieldwork Documentation

SEBASTIAN CUY, JULIANE WATSON, THOMAS KLEINKE, DANIEL DE OLIVEIRA, PHILIPP GERTH, JULIAN SCHIERENBECK, Deutsches Archäologisches Institut Berlin, Germany

---

An archaeological field research database that can be used for different projects poses an advanced technical problem. It does not only have to deal with different needs of a variety of disciplines and methods like excavation and survey but also be usable for architectural or object studies. Therefore, a generic data model is required that can deal with most circumstances while also trying to maintain standardization where possible. Another requirement is the ability to support distributed work and data entry in the field even in places with no internet coverage. This creates the need for a very robust synchronization between the clients and the server. As the use of GIS is still gaining importance in archaeological fieldwork, an additional focus lies on the integration of spatial data and descriptive documentation. This paper presents iDAI.field 2, the field research documentation system currently being developed by the German Archaeological Institute. This solution will be Open Source and relies heavily on web technologies. It makes use of CouchDB/PouchDB for data storage and synchronization. The cross-platform client application is realized with the Electron and Angular frameworks using TypeScript as the main programming language. The client also provides an open interface for import and export. This allows the integration of other fieldwork software like for example QGIS, iDIG and survey2gis into the toolchain.

---

## Key words:

Database, GIS, documentation, field recording, synchronization.

## CHNT Reference:

Sebastian Cuy et al. 2019. iDAI.field 2: A modern approach to distributed fieldwork documentation.

## INTRODUCTION

It comes as no surprise that archaeological excavations, surveys, building studies and similar projects that require field research produce a large amount of data. Also, different requirements for the terminology and the structure of the recorded data arise in the various archaeological disciplines. Consequently, there has been a strong tendency in the archaeological community to develop project-specific and institutional systems for recording data, which vary substantially with regards to their accessibility and can range from unique databases to handwritten documentation. The resulting heterogeneity of the data produced on the one hand is inevitable and even fruitful for an evolving scientific field. However, it presents a difficult challenge, especially for the sustainable storage and for ensuring availability and comparability of the data in the long-run.

This paper introduces iDAI.field 2 – the field research documentation system produced by the German Archaeological Institute (DAI). iDAI.field aims to present one solution to these problems by creating a holistic system that all researchers can use to document their fieldwork. It aims to strike a balance between standardization and customizability and provides support for all relevant aspects of the data life cycle. This does not only include the creation of the data – its capturing and processing – but the final steps of publishing and archiving as well. The design of the system also takes different scales of data collection into consideration, from the relative simplicity of one person entering and editing data to the complexities of interdisciplinary and inter-institutional collaboration. The latter point is particularly important, as most projects today are not carried out by a single discipline. Therefore, field research teams are often in different locations, a typical example being a field research team that is split between

---

□

Authors address: Deutsches Archäologisches Institut, Podbielskiallee 69-71, 14195 Berlin.

excavation and finds processing during the campaign, while the conservation of finds is done at a later time. The chances of both teams having adequate (if any) internet coverage during a field campaign cannot be guaranteed and based on the feedback that was given by various projects working with the DAI during the pre-development phase of iDAI.field 2 it was determined that the synchronization capabilities of the new system should not require a stable internet connection. Furthermore, the finds are rarely processed in the same place where the database records are being edited. It is this reality of archaeological documentation that creates the need for a robust synchronization between the various parties without the constant fear of data loss. At the same time, data consistency needs to be ensured; not only in a technical, but also in a conceptual way, in order to make project independent querying or data mining applications possible and pave the way to interoperability with other systems. Since the use of spatial data and photographic documentation are continuing to gain importance, it became obvious that a database alone would not be sufficient. Thus, features of a geographic information system and photo management software were included in the system as well.

## RELATED WORK

Software for the digital documentation of archaeological field work has been developed since the late 1980s. Jensen [2017] gives an overview over different solutions developed in the UK and throughout Europe. Most current tools, e.g. ArK [Eve and Hunt 2008], Inari AIS [INARI] or Spacialist [SPACIALIST], take a web-first approach that requires an internet connection for data entry. Others, like achaeoDOX [Räther and Sieglhoff 2015] or the ArcGIS-based IntraSiS [INTRASIS], implement a client-server model that uses a central database for data synchronization. What all these systems have in common is a relatively high barrier of entry due to the technical complexities involved in setting up web servers and databases and configuring the respective software stack. One particular application that stands out in terms of the usability is the mobile application iDig [IDIG] which unfortunately is only available for the iOS platform. This is why one explicit goal when developing iDAI.field 2 was establishing a low technical barrier and general ease-of-use. Another important point that sets iDAI.field 2 apart from most of the aforementioned systems, which are commercial solutions often requiring costly license subscriptions, is that our software is developed under an open source license and its constantly evolving source code has been available online<sup>1</sup> from the beginning in the hope to encourage the formation of a community of interested archaeologists and developers.

## SOFTWARE ARCHITECTURE

To cover the above-mentioned requirements, the German Archaeological Institute started the development of iDAI.field 2 in 2016. As the name suggests, it was meant as a replacement for the Filemaker based field research software iDAI.field [Schäfer 2011]. The inflexibility of the relational data model and the need for parallel editing without internet connection led to the development of different independent and incompatible versions that implemented individual project requirements. Also, this predecessor did not provide the possibilities of including interfaces to widespread or emerging technologies like GIS or 3D recording, despite its substantial license cost. These limitations, in addition to the time spent training new students to use this complicated database system, made the utilization difficult for several projects. It is for these reasons that synchronization, customization, open interfaces and ease of use were the cornerstones for the design of the new system.

iDAI.field 2 consists of three parts: The core of the system is the cross-platform desktop application iDAI.field 2 client built with the framework Electron<sup>2</sup>, which allows creating native applications with web technologies like Javascript, HTML and CSS. Further, Angular<sup>3</sup> is used as a client-side web application framework for building the application using Typescript<sup>4</sup> as the programming language. The data is saved within the NoSQL database PouchDB<sup>5</sup>, which allows the data to be stored locally, without the need of being online (Fig. 1). This allows the app to be run as a stand-alone application without requiring an internet connection or the installation of separate software packages. Besides that, being able to rely on the CouchDB replication protocol<sup>6</sup> greatly reduced the cost to implement robust synchronization and conflict-detection mechanisms. The main benefit of using a NoSQL datastore

<sup>1</sup> <https://github.com/dainst/idai-field>.

<sup>2</sup> <https://electronjs.org>.

<sup>3</sup> <https://angular.io>.

<sup>4</sup> <https://www.typescriptlang.org>.

<sup>5</sup> <https://pouchdb.com>.

<sup>6</sup> <https://docs.couchdb.org/en/stable/replication/>.

however is that there is no need for a fixed schema that has to be defined globally and a-priori. Technically this allows a highly customizable data model that can evolve along individual projects and new methodologies.

Consequently, the iDAI.field server provides online synchronization capabilities by using CouchDB<sup>7</sup> as the underlying database system. The server further provides search capabilities, making use of Logstash<sup>8</sup> and the search engine Elasticsearch<sup>9</sup>, which are made available to other applications through a RESTful API implemented as a Java Servlet. While the data of DAI projects is hosted on the iDAI.field server, it is important to note that client-to-client synchronization is always possible without further infrastructure requirements and that the client-server architecture is easily applicable to any other server that meets the requirements to install CouchDB.

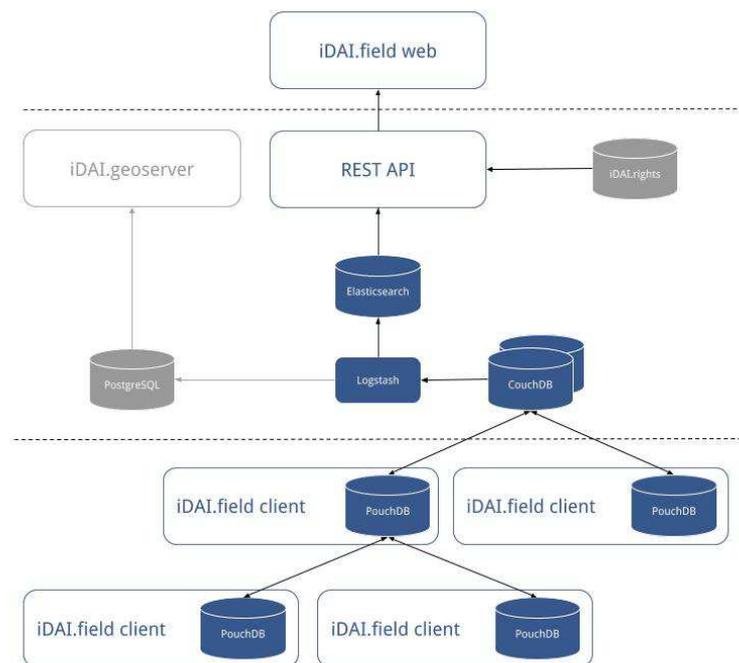


Fig. 1. Schematic overview of software architecture.

iDAI.field web acts as a frontend to the server and allows for all the data stored in iDAI.field server to be presented on the web<sup>10</sup>. It currently offers project specific and password protected entry points for the visualization and future publication of the research data. It is in an early beta stage and will be extended further to also feature a meta-search that will enable cross-project queries.

## DATA-FLOW

The typical data workflow in field projects using iDAI.field 2 starts with the recording of survey data, with the help of a total station or a GPS system (see Fig. 2). Further processing, such as format conversion, georeferencing and automatic or manual topological cleanup can be done with survey2gis [SURVEY2GIS] or any GIS software. At this stage the survey data is also provided with a unique identifier, which is used to join geometry and descriptive

<sup>7</sup> <http://couchdb.apache.org>.

<sup>8</sup> <https://www.elastic.co/de/products/logstash>.

<sup>9</sup> <https://www.elastic.co/de/products/elasticsearch>.

<sup>10</sup> <http://field.dainst.org>.

metadata recorded in the client. This workflow uses GeoJSON<sup>11</sup> as a standard format for export and import. Depending on the collected data and the exporting tool, 2D and also 3D coordinates are supported. Another alternative way of recording survey data in the field is available. By using the iOS application iDig it is possible to directly connect to a total station via Bluetooth. The benefit of this method is the ability to already record some descriptive data in the app at this stage. While it is beneficial that the data can be directly transferred to the iDAI.field client via WiFi, this does require the use of a proprietary API based on HTTP and CSV. Additionally, this rules out intermediate processing, which is possible when using established GIS tools and formats. Therefore, it should be noted that the use of survey2gis or any GIS software seems to be the preferred tool for connecting the collected geodata with iDAI.field 2.

Further, more elaborate descriptions of the documented contexts and/or artifacts can then be accomplished on a computer running the iDAI.field client. This also includes the possibility of importing images and connecting them to the corresponding documented objects and activities. Additionally, the client supports the importing of georeferencing information in the world file format. This allows for the visualization of raster image material, such as aerial photographs or digitally/manually drawn plans, as background layers in the iDAI.field client.

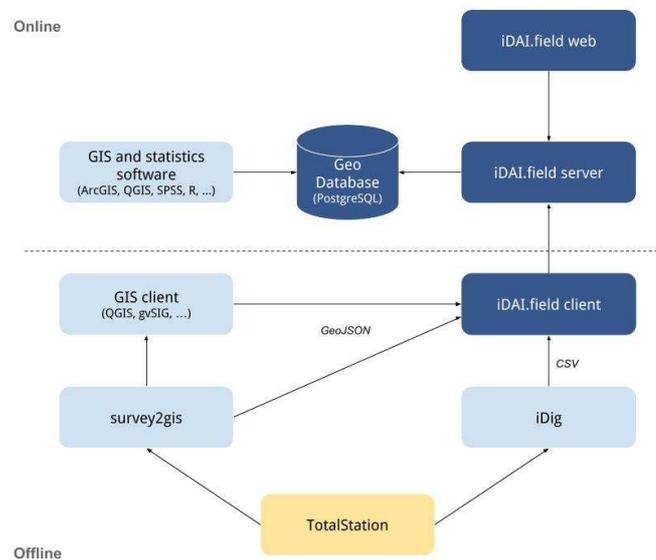


Fig. 2. Schematic overview of data flow.

If an internet connection is available, the data can then be synchronized with the server (see below). On the one hand, this establishes a central repository for all projects which in turn simplifies data management and preservation. On the other hand, the server offers interfaces in different formats and standards that can again be used by other applications for further data analysis (such as GIS applications or tools for statistical analysis). One such interface will be achieved by making a synchronized PostgreSQL database with the PostGIS extension available, as this constitutes a standard interface for many GIS tools. Another API to the iDAI.field data that is provided on the server side is the Representational State Transfer Application Programming Interface (REST-API) that is already used by iDAI.field web. This provides access to the data as linked data serialized in JSON-LD<sup>12</sup>.

## DATA MODEL

The process of creating an application that can be used for different projects and scientific questions and is supposed to give research teams the possibility to share and collaborate on the data poses advanced technical problems especially with regards to the underlying data model. It not only has to cope with the different needs from a variety

<sup>11</sup> <https://tools.ietf.org/html/rfc7946>.

<sup>12</sup> <https://www.w3.org/TR/json-ld>.

of disciplines and scientific methods, but it also has to walk the fine line between data consistency and project-specific needs. Therefore, a flexible data model is required that is customizable while still providing enough common ground to ensure comparability and data exchange and to enable the development of a user-friendly application.

Making archaeological data that has been documented in various projects comparable has been an ongoing effort in the last decades. There have been many attempts, efforts and approaches to create standardized terminologies, overarching data models and ontologies for cultural heritage data. Concerning the latter, the CIDOC CRM<sup>13</sup> has been among the most popular approaches. iDAI.field 1 was built many years with the intention of making project data comparable. Since 2006, nearly 50 projects of the German Archaeological Institute have carried out field campaigns and this has resulted in a large amount of acquired data. Interestingly, a trend seems to have emerged, as

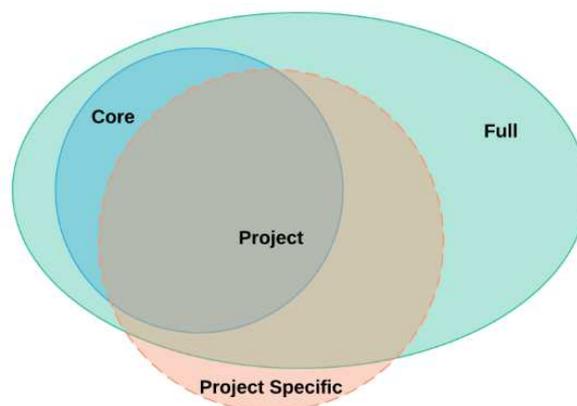


Fig. 3. Schematic overview of data model, available fields from different pools at a given point in iDAI.field 2.

each field campaign continues to collect more data than the last. The standard has been a proprietary database software, which has a core list of attributes, to which other fields were added over the years. After extracting this tabular data, different usage-statistics were applied in order to determine which attributes were actually needed by every single project. This approach was taken to verify or disclaim whether or not the projects were indeed comparable. Briefly summarized, the outcome was as expected: Most projects seemed to only use a few distinctive attributes, which were largely determined by the type of project, as well as by the preferences of the project supervisor. Still, a nucleus of fields and values could be extracted that was essential to all projects. Furthermore, it became apparent that while the use of some fields seemed heterogeneous, others were only homogenous among projects with similar needs. By default, a survey and a research project focusing on specific buildings will have different needs.

Thus, the data model of iDAI.field 2 contains three different parts that overlap, complement and complete each other (Fig. 3). The 'core' section contains the attributes that were identified as being the absolutely necessary attributes for the documentation of archaeological fieldwork. These were determined based on the usage statistics acquired from the various DAI projects. Furthermore, this is the basic configuration that is distributed with the software. Additionally, there are attributes that are more detailed and important for specific types of projects. This is the extended 'full' configuration of the iDAI.field client. Furthermore, every project will have the opportunity to create attributes that are unique to their specific needs. These very unique attributes, despite offering projects their necessary flexibility, will clearly not be comparable to other projects. Therefore, the maximum consistency of acquired data is maintained. To guarantee that the existing fields are filled with comparable data, the available value lists are continuing to be enlarged. The vocabulary employed by iDAI.field 2 has been largely checked and approved by specialists from their respective fields of archaeology, anthropology and archaeobotany. Projects can start with a large number of options and then sort out what they do not need. Additional values can also be made available, after extensive consideration. The pool of available fields is openly accessible and those that are deemed unnecessary can be hidden at any stage of the project. The measurements of finds can be used as an example. They are an essential element of most excavations and are, therefore, part of the core section. However, a project may decide that they are

<sup>13</sup> <http://cidoc-crm.org>.

not necessary for their initial campaign or that they do not have time to enter these values while in the field. In order to simplify the process these fields can be rendered invisible but be made easily accessible at any later stage. The development team is currently evaluating the best ways to structure and streamline this process further.

The system is set up in a modular manner. The overview section shows the main resources, which are separated into trenches, buildings, survey areas and places. The excavation section allows for the creation of stratigraphic units, operations and finds, which are automatically provided with the spatial relation for the resource they were created in (Fig. 4). Additionally, designated modules are designed for the separation of the scientific documentation of architecture and survey fieldwork. All properties of a resource inherit the attributes from their main type (stratigraphic unit, find etc.), so a generic feature can be documented and later assigned e.g. as a floor or grave, pottery or terracotta. This setup allows for the maximum amount of flexibility regarding the defining categories of material culture. Furthermore, through the past problems of documenting archaeological data in the somewhat strict, predefined margins of a database, two main points were identified where additional adaptability for the recording of archaeological findings was needed: measurements and datings. This issue was addressed by giving the user the

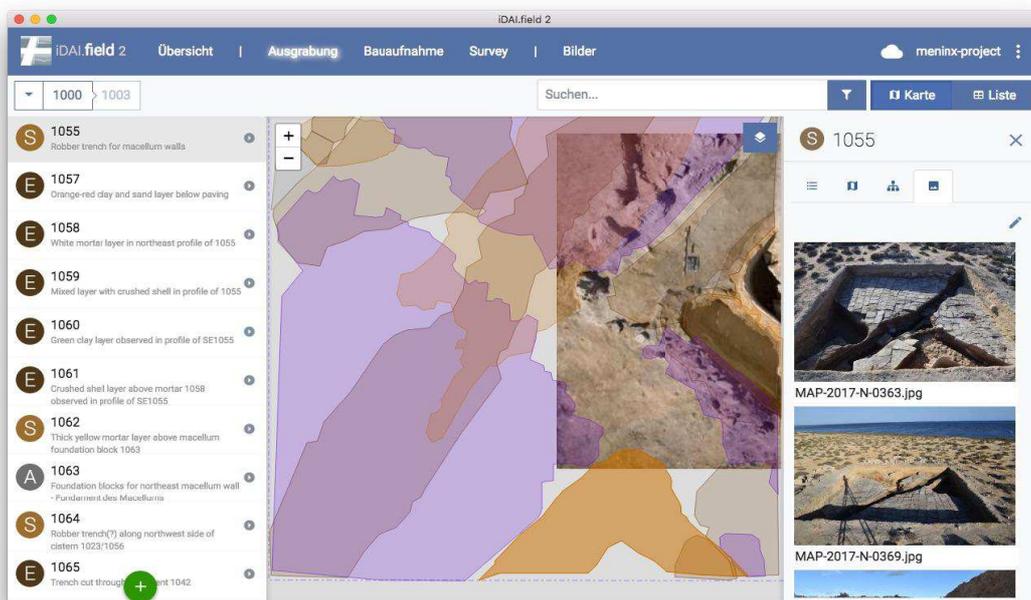


Fig. 4 Excavation view while synchronizing with other client.

option to enter more than one value for a single object, while being able to define and customize the position of measurements (e.g. height measured twice: at the minimal and maximal extent of the figurine, different diameters of vessels etc.) and the type and origin of datings (e.g. BP, BCE) – including uncertainty. The entered values are normalized in the database itself. Datings are simply converted to plus and minus (BCE and CE) and measurements to micrometers so that the values are not only textual representations (e.g. 2 cm, 2 m) and remain comparable. Therefore, in the future it will be possible to search and compare objects, stratigraphic units, graves etc. that were given a similar dating. This could simply be items within a similar time period, but could also be used to compare certain entries, like coins that have an absolute date. Naturally, one could also filter for specific periods and datings.

## DATA SYNCHRONIZATION

Depending on the conditions and the workflows of the field research project, there are a lot of possible scenarios for allocating the data and how each client could access the internet. Archaeological work is generally split into numerous tasks that are being performed by different agents and the software needs to be able to deal with this. A

common scenario is an excavation leader who has the overall responsibility for a site, while several trench leaders take care of their individual trenches. The general find processing is then centralized in a depot. All of these actors need to have access to overlapping data despite having, perhaps, sparse internet connectivity.

A solid synchronization mechanism is, therefore, needed to deal with all these different circumstances. In general the client could be run as standalone software without any internet or local network connection. All the data is then stored locally in the PouchDB. To synchronize your local data you could either connect with another computer using iDAI.field by pointing to the IP address with the port used by PouchDB in the synchronization settings (fig. 5), or by connecting with the iDAI.field server available online. The necessary information on how to connect to another client is given to the user in the synchronization settings. Considering the very real possibility of a slow internet connection, one benefit of using CouchDB with a NoSQL data format is the small size of the database, which also streamlines this process (the largest test dataset with 3000 entries was under 35 MB – this excludes full-sized images as only thumbnails are stored in the database).

The replication feature of PouchDB and CouchDB is the major strength of the database, which ensures the data consistency in iDAI.field. It is not necessary to define one central master database – in fact, it could be any client. However, with every synchronization attempt conflicts may occur. Conflicted revisions are stored as part of the revision tree within the database. These can be handled in a conflict resolution screen where conflicting attributes are highlighted and can be adjusted manually or by selecting the winning value.

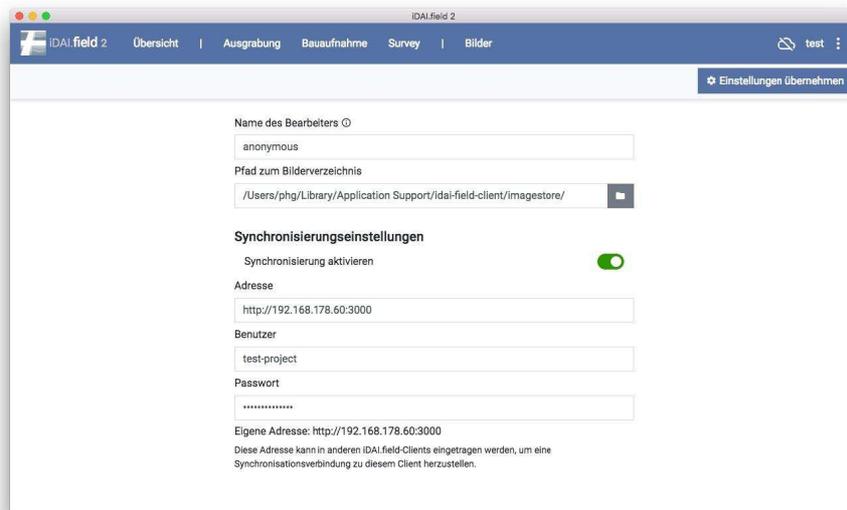


Fig. 5. Synchronization settings.

## CONCLUSION AND OUTLOOK

The developed application provides a generic and holistic approach for handling the data of field research projects. It supplies the capability to support different characteristics of projects within the same database schema and also provides a maximum amount of flexibility. This includes the setup of project specific needs for documentation, while also offering the means to compare projects with one another through the standardized objects used to document archaeological fieldwork in the core configuration. Furthermore, the process of setting up a project database is simplified, as nothing additional to the installation package<sup>14</sup> needs to be installed and no registration, in any form, is required.

Before its release in January 2018, two different projects have tested the iDAI.field client during their fieldwork campaigns. Their willingness to share their expertise and real-life experience with the software was invaluable to the

<sup>14</sup> Available under <https://github.com/dainst/ida-field/releases>.

development team. Both the client-to-client and the client-to-server synchronization performed more than satisfactorily, while the import of geodata was flawless and the data model just had to be extended marginally. Although the field test was successful and the usability feedback was positive, it did become clear that certain features had to be implemented to complete the archaeological toolchain. As a goal for the near future, a view based on the Harris matrix (available as of version 2.2 released in April 2018) as well as new and extended possibilities for importing and exporting data in different formats will be realized. Additionally, a major goal is to further integrate iDAI.field 2 into the existing infrastructure. In particular, integration with iDAI.world will allow for the possibility to publish finds in iDAI.objects Arachne, link places to the iDAI.gazetteer, to keep track of bibliographical references with iDAI.bibliography, as well as providing interfaces to external portals. Another objective is a multilingual user interface and the further development of data publication via iDAI.field web. Data that was acquired with the FileMaker database will also be migrated into iDAI.field 2, while the productive system was initialized, and new projects will start to use the software in 2018.

The constant maintenance and continuous care of the repository has to be assured and is essential to the prospective long-term usage of the software as well as to the usability of the projects. At the same time, the DAI aspires to guarantee the stability and security of the project data.

## REFERENCES

- Felix Schäfer. 2011. iDAI.field – Ein modulares Dokumentationssystem für archäologische Forschungsprojekte, KuBA 1 (2000), 201 ff.
- P. Jensen. 2017. Where are we? Reviewing the Integration of Complex Spatial Data in Current Field Archaeology, *Internet Archaeology* 44. <https://doi.org/10.11141/ia.44.8>
- S. Eve and G. Hunt. 2008. ARK: A Development Framework for Archaeological Recording, in: Posluschny, A., K. Lambers and I. Herzog (eds.), *Layers of Perception. Proceedings of the 35th International Conference on Computer Applications and Quantitative Methods in Archaeology (CAA), Berlin, Germany, April 2–6, 2007* (Kolloquien zur Vor- und Frühgeschichte, Vol. 10). Dr. Rudolf Habelt GmbH, Bonn, pp. 10 + CD-ROM.
- Jörg Räther and Eicke Sieglhoff. 2015. archaeoDox – Information management from the dig into the archive in *Proceedings of the 19th International Conference on Cultural Heritage and New Technologies 2014 (CHNT 19, 2014)*.
- INARI. Inari Archaeological Information System. Website. Retrieved August 2018. <https://www.inari-software.com/en/archaeology/>
- INTRASIS. intraSiS - intrasite information system. Website. Retrieved August 2018. <http://www.intrasis.com/>
- IDIG. iDig - Recording Archaeology. Website. Retrieved August 2018. <http://idig.tips/>
- SPACIALIST. Spacialist – Eine virtuelle Forschungsumgebung für die Spatial Humanities. Website. Retrieved August 2018. <http://www.escience.uni-tuebingen.de/projekte/spacialist.html>
- SURVEY2GIS. Open-Source GIS-Tools für GIS, Vermessung und Datenakquise. Website. Retrieved August 2018. <https://www.survey-tools.org/>

### *Imprint:*

*Proceedings of the 22nd International Conference on Cultural Heritage and New Technologies 2017. CHNT 22, 2017 (Vienna 2019). <http://www.chnt.at/proceedings-chnt-22/> ISBN 978-3-200-06160-6*

*Editor/Publisher: Museen der Stadt Wien – Stadtarchäologie*

*Editorial Team: Wolfgang Börner, Susanne Uhlirz*

*The editor's office is not responsible for the linguistic correctness of the manuscripts.*

*Authors are responsible for the contents and copyrights of the illustrations/photographs.*