

# Building the DEFC App

---

Edeltraud Aspöck, Anja Masur, Seta Štuhec

Digitizing Early Farming Cultures aims to create standardized and integrated research data of Neolithic and Chalcolithic sites and finds of Greece and Anatolia (c. 7000–3000 BC according to Greek terminology). The bases for the project are non-digital and digital sources from a series of research projects from the new ÖAW research group Anatolian Aegean Prehistoric Phenomena (AAPP) at the OREA Institute. AAPP studies archaeological sites and finds from the Neolithic to the Bronze Age in Greece and Western Anatolia, two neighbouring and archaeologically closely related regions usually studied in isolation of each other. The group focusses on research questions concerning technological and social changes, settlement patterns, exchange and sourcing of raw materials.

Research across the area currently suffers from fragmented data organized according to differing knowledge schemes, which developed independently in the two different research traditions. Independent terminologies and chronologies have developed, hindering collaborative research. To provide a basis for studying archaeological phenomena collaboratively across the whole region, standardization of research data is required.

The aim of this project is to harmonize existing datasets, digitize analogue data from publications and integrate metadata for easy access and data reuse. The data will be made available in an open access archaeological data management system, which complies with standards of data sharing and interoperability with related initiatives. The present text describes how the project evolved.

## 1. Our resources

The resources of the DEFC project include digital and analogue/non-digital resources from different research projects lead by the members of AAPP research group.

[PREHISTORIC ANATOLIA](#) is an ERC Starting Grant project (263339) led by Dr. Barbara Horejs. The main focus of the project are the period of the first permanent settlements (sedentism) from Neolithic to Early Chalcolithic times, and the development of proto-urban centers from Late Chalcolithic to Early Bronze Age period in Western Anatolia. To achieve an integrated picture and to concentrate the broad spectrum of studies the focus is set on the following three research topics:

- Archaeological periods and definitions of cultures
- Societies in changing environments
- Communication, exchange and interregional Relationships

The project included new excavations at Çukuriçi Höyük, archaeological and environmental surveys in Kaykos and Kaystros valleys (micro-regions of Pergamon and Ephesos) that produced modern documented data. Two databases produced in the course of the project were used as a data resource in DEFC project:

- The Microsoft Access database „Sites and Modelling“(by Christoph Schwall) containing data about Neolithic and Chalcolithic Anatolian sites as well as an overview about the most important finds and their context.
- The Microsoft Access database „Çukuriçi Höyük“ with data about all finds and features from the Anatolian Neolithic/Bronze Age tell Çukuriçi Höyük.

The research in Neolithic settlement [AIGEIRA IN ACHAIA](#) conducted by Dr. Eva Aram-Stern resulted in a Microsoft Access database “Aegeira”, which contains information about Neolithic pottery from

Aegeira, Greece. Furthermore, research in Neolithic settlement [VISVIKI MAGULA](#) also conducted by Dr. Eva Alram-Stern resulted in another Microsoft Access database “Visviki”, which contains data about Neolithic pottery from Visviki, Greece.

[ÄGÄISCHE FRÜHZEIT: DAS NEOLITHIKUM IN GRIECHENLAND mit Ausnahme von Kreta und Zypern](#) is the first of two volumes of “Ägäische Frühzeit, 2. Serie” publication by Eva Alram-Stern. The publication deals with the Neolithic period in Greece including the research reports from 1975 to 1993. It includes an analysis and a site gazetteer of important information about Neolithic Aegean sites and has been as such used as a resource in the DEFC project.

[ÄGÄISCHE FRÜHZEIT: DAS NEOLITHISCHE UND VORPALATIALE KRETA](#) is the second of two volumes of „Ägäische Frühzeit, 2. Serie“ publication by Eva Alram-Stern. It deals with new research conducted on Neolithic, early Bronze Age and beginning of middle Bronze Age (from ca. 7000 to 2000 BC) sites in Crete.

[SCHACHERMEYR POTTERY COLLECTION](#) is a pottery assemblage consisting of all relevant types of Neolithic pottery that was collected by Fritz Schachermeyr (†) who left it to the Austrian Academy of Sciences so that it could serve researchers, students and all parties interested in studying the Neolithic pottery of Greece. The collection was published by Dr. Eva-Alram Stern in the publication [SCHACHERMEYR, F. \(†\), 1991](#). Sammlung Fritz Schachermeyr: Die neolithische Keramik Thessaliens. Aus dem Nachlass bearbeitet von Eva Alram-Stern. Veröffentlichungen der Mykenischen Kommission 13. “Die neolithische Keramik Thessaliens” by Fritz Schachermeyr (†). Additionally, most representative sherds were selected to be digitized in 3D, which would finally allow them to become available to the broadest possible audience.

## 2. Resources analysis

The first step to achieve standardized and integrated research data of Neolithic and Chalcolithic sites and finds of Greece and Anatolia was to analyze the data resources.

The result of the analysis: weight of digital datasets is on pottery while most of the site information is contained in publications, in particular the ‘Ägäische Frühzeit’. Therefore, we will need to digitize the information on sites – create a database of archaeological sites in the Aegean and Anatolia which we can link to our information on finds and the digital information on pottery and excavations.

## 3. User Requirements for a site database

In regular focus group meetings with the AAPP research group we discussed requirements for both the new data structure and the technical implementation:

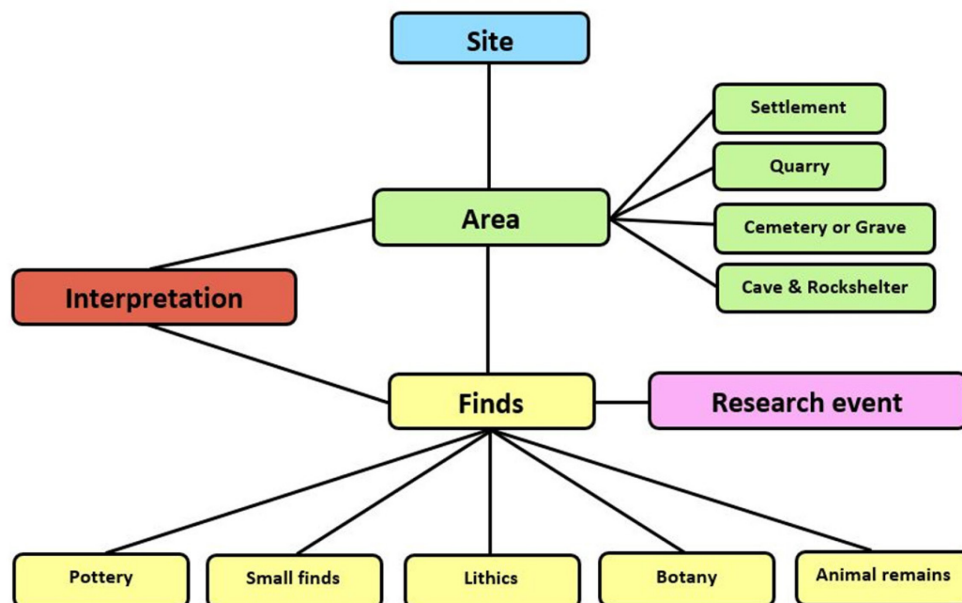
- a site database that provides more detail on finds and archaeological evidence than existing online open access site databases for the region;
- an online research resource for students and researchers that shows *typical finds*, in particular pottery characteristic for the period and region
- a map interface that allows searching for both Greek and Anatolian sites;
- 3D models of selected pottery sherds of the Schachermeyr collection;
- enable extending this database or linking it with other relevant databases;
- online and open access publication of all data.

## 4. Conceptual data model

The mindmap with all existing database fields, as well as information from some analogue resources (see data analysis above) formed the basis for the creation a conceptual data model.

The DEFC App data model meets the particular requirements of dealing with data of different granularity. This means that some data sets are very detailed containing many sub-levels (such as the documentation of modern archaeological excavations and pottery databases), whereas other data sets comprise of broad archaeological data (e.g. old excavation reports and other legacy data). The data model therefore connects different levels of information about excavation processes, finds and chronological periodization. To meet the needs of all possible users, a clear conceptual data model that is based on the most probable relationships between archaeological objects has been defined with the following main model classes:

- Site (location where archaeological research took place/observations were made);
- Research event (project and type of archaeological research that was carried out);
- Area (particular part of the site within a certain time span – period, e.g. excavation trench, a part of a settlement in a certain period);
- Finds (artefacts, human-, animal- and plant remains found in a certain area);
- Interpretation (archaeologist's interpretation of areas/finds etc.).



## 5. Implementation

### 5.1. Arches

In spring 2015, the analysis of different data management systems resulted in choosing [Arches](#), an open source software system designed for the heritage sector to inventorize and manage all types of immovable cultural heritage. A new version of *Arches* was just being launched (April 2015). The new *Arches* version (v3.0) appeared to be a graph database with the ontology CIDOC CRM integrated. It was possible to customize the *Arches* application 'Heritage Inventory Package' (HIP) according to user needs and *Arches* also includes a tool for creating thesauri, mapping features and a timeline. This allows publishing data online according to internationally accepted standards. Hence, *Arches*

provided most features that we needed for our system to achieve DEFC project objectives. We therefore decided to customize *Arches* for our project. For this we got technical support by the [ACDH](#).

At this stage (when) we submitted a paper to the Digital Heritage conference, where we introduced the project and announced to use *Arches* for implementation.

[Aspöck, Edeltraud; Masur, Anja \(2015\). Digitizing Early Farming Cultures. Customizing the Arches Heritage Inventory and Management System. Proceedings of Digital Heritage International Congress 2015, 28. Sept. - 2. Oct., Granada, Spain. DOI: 10.1109/DigitalHeritage.2015.7419549.](#)

However, setting up *Arches*, we experienced problems which showed that *Arches* version (v3.0) was not a practical solution for our project requirements

- the documentation was not finished at this time
- SKOS import of thesauri was not working properly
- we missed semantic richness and the use of the CIDOC CRM extensions CRMarchaeo and CRMsci
- the mapping of the *Arches* data model did not always fit our requirements
- our data model was too specific and not all concepts were covered by *Arches*
- one of the most problematic issues was that the user interface did not adapt to modifications in the data model, and we would have had to create a new user interface.

For these reasons we decided not to use *Arches*, but to find another solution.

## 5.2. DEFC App (site database)

In summer 2015 after experiencing all this problems it has been decided to develop a new site database from scratch. Peter Andorfer and Ksenia Zaytseva (ACDH) created a Django-based database which we now call DEFC app. The database set up took around 6 months. In this time many meetings between Digital Archaeology research group and ACDH technical developers took place to make sure all requirements will be implemented. AAPP research group has been involved in the process through several workshops and has given regular feedback after testing each version of the app.

The data model of DEFC app consists of all our previously defined components: Sites, Research Event, Area, Finds, and Interpretation. Data is entered into each component individually, while the components and the data are linked to each other. Most of the entry fields consist of (interdependent) dropdown lists filled with controlled vocabulary to minimize data entry mistakes and ensure better querying conditions.

The database features a map interface with all entered sites (where coordinates are available) and views of the 3D scanned potsherds of the Schachermeyr collection. Additionally, DEFC App is linked to the [Zotero online bibliography database](#) that is combines several bibliographic databases of AAPP research group.

## 6. Data entry

The database allows different levels of access. Everyone with internet access can browse the published data, but to insure data credibility a login account is necessary to add new data to the database.

As first data to be entered in DEFC app (spring 2015) the following publications and other resources have been selected:

- ALRAM-STERN, E., 1996. Die Ägäische Frühzeit“. Band 1: Das Neolithikum in Griechenland. Veröffentlichungen der Mykenischen Kommission 16.
- SCHACHERMEYR, F. (†),1991. Sammlung Fritz Schachermeyr: Die neolithische Keramik Thessaliens. Aus dem Nachlaß bearbeitet von Eva Alram-Stern. Veröffentlichungen der Mykenischen Kommission 13.
- ÖZDOĞAN, M.; BAŞGELEN, N.; KUNIHOLM, P. (ed.), The Neolithic in Turkey. New Excavations & New Research. Volume 1 - The Tigris Basin, 2011, Istanbul.
- ÖZDOĞAN, M.; BAŞGELEN, N.; KUNIHOLM, P. (ed.), The Neolithic in Turkey. New Excavations & New Research. Volume 2 - The Euphrates Basin, 2011, Istanbul.

As help for the data entry the following tools have been created:

- Pottery image gallery that provides visual examples of different types of pottery form, decoration and detail from different regions in Greece and Anatolia (due to copyright issues this tool is only visible to users who register an account)
- A map of regions and districts of Turkey and Greece differentiated by the AAPP research group and DEFC app. The regions of Turkey are a result of research of AAPP research group, whereas the districts correspond to the official districts of Turkey. The map of Greece and its regions has been created based on the resources used by the British School of Athens.

## 7. 3D models

90 most representative sherds were selected from the Schachermeyr pottery collection and 3D digitized using Breuckmann smart Scan HE 5 Megapixel Color 3D Scanner. The 3D models were added to the DEFC app homepage using [3DHOP](#) (3D Heritage Online Presenter). The 3D models are not going to be presented as a plain collection of separate 3D models, but will be incorporated within the DEFC App database as a part of a rich dataset providing the archaeological information.

In order to ensure a certain technical reliability the 3D model provenance metadata were added to the 3D models. When deciding on which data to include with the digital Schachermeyr models, we considered the standards recommended by [3D ICONS](#), the [Archaeological Data Service \(ADS\) Guides to Good Practice](#) and the [IANUS project](#). Additionally, the [CIDOC CRMdig](#) extension has been taken into account, since it is planned to map all datasets to CIDOC CRM and its extensions. The provenance metadata are divided into three groups:

- Administrative metadata
  - Project name: the name of the scanning project;
  - Actor: the person who captured 3D data and the person who did the post-processing;
  - Time of capture: when were the data captured and when were they processed;
  - Copyright;
  - Digital asset ID: the file name;
  - File format;
  - URI/ID.
- Activity (capture)
  - Type of activity: which 3D acquisition method was applied (e.g. 3D structure light scanning) ;
  - Digital device: the hardware for the 3D capturing. In the case of 3D scanning, the metadata that should be recorded are:
    - Scanning device;
    - Texturing device: scanner camera or external camera;
    - Computer device;
    - Accompanying software (and pre-set parameters);

- Capture conditions record of any external activities that might have influenced the 3D capture procedure (e.g. weather).
- Activity (processing)
  - Software: the post-processing software and its version;
  - Action: the processing steps that have been carried out (e. g. hole filling, polishing, texture-colour conversion, decimation, compression)
  - Import/export file format.

## 8. Public view

At this time all published data is accessible for viewing and downloading. Additionally, geo-visualization of sites is available as well as customized filtering and ordering of separate entries. In the future more advanced query interface will be set up.

The code of the database is available via API on the project website (application program interface) and [GitHub page](#).

## 9. Linked open data

DEFC app is an open access database. Everyone can browse and query the data. In order to enable integrating other databases or linking to those, we map our database to the ontology CIDOC CRM. This work will be completed in 2016

In 2016, we will integrate several databases containing data about Greek pottery from different sites by mapping those to CIDOC CRM and storing these in a triple store. This use case will show how integrating of different data resources into our site database can work.