Scientific Datasets SIG

Scientific Datasets in Archaeological Research

(Working Document)

Version: 0.5 **30/01/2017**

Authors: **Nikolaos Kazakis (ARC)**
Nestor Tsirliganis (ARC)

Contributing partners: **Attila Kreiter (MNM NÖK)**

This report includes results of work conducted in the ARIADNE-Advanced Research Infrastructure for Archaeological Dataset Networking in Europe project. The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7-INFRASTRUCTURES-2012-1) under grant agreement no. 313193. The views and opinions expressed in this report are the sole responsibility of the authors and do not necessarily reflect the views of the European Commission.

Scientific Datasets in Archaeological Research

ABSTRACT

Recent trends in archaeological research dictate the incorporation of various analytical methods for dating, chemical and/or mineralogical characterization, morphological description, biological and environmental studies, etc. of various excavation findings, ancient objects and landscapes to support and corroborate the archaeologists' observations. Each method, when used individually, provides a different feedback within a relatively discrete scope, but once single studies are combined, the information produced by their amalgamation can significantly contribute to the solution of the archaeological puzzle, answering various explicit (e.g. age, place) and implicit (temporal evolution, connection between places) raised questions.

As a result, numerous scientific data (of various formats and types) and metadata are produced creating large and complex scientific datasets. The use of common terminology and definitions to describe these data along with the establishment of a formal standardized structure that would provide a common and extensible semantic framework that any cultural heritage information could be mapped to would assure the credibility and durability of scientific datasets in the archaeological research.

Though some attempts have been made towards this direction, most notably with the CIDOC-CRM extension CRMsci, a detailed general standard that covers the multitude of the scientific domains and specialties involved in Heritage Science is rather difficult to exist. Instead, general guidelines accompanied by a relatively simple and small set of metadata can be established to facilitate efficiently the use of scientific datasets in the archaeological research.

An attempt to illustrate the complexity of the problem and suggest a direction for its solution is attempted through the detailed presentation of the common practices used in pottery provenance studies which provides an overview of the various types and formats of the scientific data and metadata produced in the archaeological research.

The document has been prepared by Nikolaos Kazakis (ARC) and Nestor Tsirliganis (ARC) and revised to incorporate contributions and reflect suggestions by Attila Kreiter (MNM NÖK), Philip Buckland (Umeå University) and Guntram Geser (SRFG).

1 INTRODUCTION

In recent years, archaeology interacts increasingly with natural sciences and informatics to take advantage of the possibilities they offer and reduce the subjective element involved in the traditional approaches. The combined information from the various scientific disciplines allows the investigation of theories and interpretations that shed light on the human past and environment.

Numerous "components" of the archaeological research, such as Dating, Provenance and Dietary studies are leaning heavily on natural sciences and involve the use of advanced scientific methods applied on the archaeological finds. The information extracted from the materials using such methods produce a plethora of scientific data and metadata, which do not serve only as tools for answering individual archaeological questions, but also create large scientific reference datasets that allow the complete scientific documentation of the artifacts, environment and living beings of the past.

These datasets include data of various formats and types, while their metadata (e.g. experimental method, date of measurement etc) are essential for their evaluation and constructive use, especially when comparisons are attempted.

The specifications of the scientific data and metadata produced in the archaeological research and their importance from a scientific point of view are discussed below, while a minimal simplified structure for such datasets is also proposed as a guideline towards standardization based mainly on best practices and procedures followed in scientific laboratories. The above is accomplished through the detailed presentation of the common practices in pottery provenance studies, which can be considered as the most representative example of scientific data production in the archaeological research due to the large number of methods that can be involved and the variety of the obtained data.

2 APPROACHES IN THE ARCHAELOGICAL RESEARCH

Traditionally, the study of ancient artifacts was mainly focused on meticulous macroscopic observations by the archaeologists. Conclusions would be derived from their typology, decoration, iconography and chronology, based exclusively on stylistic considerations and aesthetic evaluation of the artifacts as well as their use, role and function. The above could also be complemented in some cases by available historic archive studies. In addition, evaluation of their characteristics would also shed light on the places (e.g. Attics or Corinthian pottery in ancient Greece) and the techniques of production (black-figure), the distribution, the trade routes, the organization and the contacts of the societies they created and exchange them.

In the late '50s, the term of Archaeometry was first introduced by Prof. C.F.C. Hawkes to refer to the use of methods and techniques coming from the natural sciences (physics, chemistry, mathematics, geology, geophysics, biology) and their application to archaeology as well as to art objects in order to provide solutions for specific questions and problems (Harding, 1994). The term today is expanded to include the computer technologies and their contribution to the study of the Cultural Heritage and is increasingly referred to as **Heritage Science**, which actually represents a modern and more sophisticated approach in the archaeological research.

Heritage Science can be regarded as the "forensics" in cultural investigations, since it seeks for information in retrospect trying to answer primary questions, like 'who', 'where', 'when' and 'how'. This is achieved through the amalgamation of several disciplines and the fusion of numerous scientific data, that include (though not exclusively):

- measurements of physicochemical parameters and properties of materials, environmental and biological samples
- analyses of the composition of materials, environmental and biological samples

- study of the ‘structure’ and form of materials, environmental and biological samples
- assessment of physicochemical processes in materials, environmental and biological samples
- spatial data
- statistical data

In the case of artifacts, in particular, it is evident and widely accepted that the physical properties and composition are significant descriptive characteristics of both raw materials and the finished products (e.g. Buko, 1984). In the same respect, the manufacture techniques can characterize the appearance and functional capabilities of the finished object. As a result, various methods, simple or complex, destructive or non-destructive, qualitative or quantitative, can be employed to study the above features and provide information about the past of the archaeological finds. Any advanced method employed for such purposes is eligible to describe an artifact by objective, precise and replicable standards, free from aesthetic or subjective judgments common to art-historical appraisals. In parallel, they allow the definition of units of measurement or comparison that are standardized and independent of the context of observation. Thus, for a pot that could be described as “soft and poorly fired” for one cultural region and “hard and well fired” for another, a measured hardness of 5.0 in the Mohs’ scale, and a firing temperature between 800 ^{o}C and 900 ^{o}C eliminates the ambiguity of the qualitative terms. It must be accentuated that the most important aspect of all these methods is that they yield a data base which can be used to relate material properties to human behavior and technology.

To this respect, very decisive is also the role of the digital technologies, which significantly contribute to the organization of the accumulated knowledge and set up new methodologies for analyzing the physicochemical properties of archaeological objects. Cultural databases, 3D digitization and reconstruction, as well as virtual reality, are used in order to make material remains more accessible and better understandable to both scholars and the general public. Last but not least, the Geographical Information Systems (GIS) introduced to archaeological studies since the late 20th century provide important information for matters such as trade routes, contacts, use and function etc, through the distribution of artifacts in place and time.

3 SCIENTIFIC DATASETS-CURRENT ISSUES OR CONCERNS

Large scientific datasets are frequently produced from several research infrastructures who serve the archaeological research analyzing the various archaeological finds using the latest instruments (Matsui et al., 2012). Yet, the straightforward comparison of data between different laboratories has remained problematic, despite the gradual improvement of the analytical techniques towards higher precision and accuracy. The above is attributed not only to technical reasons but also to the variable database formats used in the individual laboratories (Hein and Kilikoglou, 2012).

Scientists who support the archaeological research conducting measurements with advanced scientific methods usually adopt protocols and follow certain research paths, which however are not adequately and/or clearly documented, rendering the provenance of the extracted scientific data hard to determine. In addition, a different methodology can be followed by the various research installations making the interoperability of the scientific data almost impossible. In the same respect, the resulting scientific datasets, which may include data of various formats (e.g. numbers, spectra, charts) and types (e.g. raw data, calculation results), along with any metadata (e.g. experimental method, instrumentation, date), vary or contain inadequate information preventing their re-evaluation or use for comparison purposes.

The above is also reflected on the fact that in many case studies it has been preferable to undertake redundant new measurements of reference materials rather than accessing analytical data on the same material generated by another laboratory (Hein and Kilikoglou, 2012).

Availability and accessibility of scientific data could be a thorny issue, since in the archaeological community, scientific data (results) are not included in datasets in most cases, but they are presented in reports (for in-house or informal use) or similar documentation (Bisol et al., 2014). Thus, results may not be reproducible, not necessarily because of error, but because the data and/or metadata are absent or insufficient (e.g. Alsheikh-Ali et al*.,* 2011; Freese, 2007), or even lost (Vines et al., 2014). In most datasets created for the needs of the Archaeological Research, only the archaeological observations are included (e.g. dimensions of the object, estimated period, stylistic properties etc), while the scientific data, which would actually confirm or help the archaeologists to build theories about the past (e.g. age, provenance), are completely absent making the re-evaluation of each analysis in the future, in light of new or emerging knowledge, virtually unfeasible. In the same respect, important metadata, such as instruments and methods used, measuring conditions and standards employed etc. are rarely available. As a result, dependable comparison between data of the same scientific investigation method, but of different techniques (e.g. different equipment) and/or re-estimation of the precision of the measurement are hardly attainable.

However, accessibility to data of other research infrastructures should not be such an issue at present, since digitization of data across all domains in academic research and scholarship (e.g. Carlson, 2006; Borgman, 2009) has made them available more easily and distributed more quickly than ever before (Strasser et al., 2014). Summing up the above, the major current concerns of the scientific datasets produced in the archaeological research are the lack of efficiency and diffusion of the scientific data and the need for additional metadata, which would assure the reuse of the produced datasets, namely the re-evaluation of the samples and the inter-comparison of the results among various laboratories.

As Strasser et al. (2014) state, in order to fully take advantage of such heterogeneous data, a streamlined workflow for collection, organization and publication of data is needed. Data management, namely proper documentation, is required in order to assure the value of the scientific datasets over time thus allow the long-term storage, and easy discovery and access of the data.

4 STRUCTURE OF SCIENTIFIC DATASETS

4.1 Needs and challenges

The efficient usage of a scientific dataset has several requirements that a user should take into consideration. The most important is that the user must get acquainted with the available scientific methods which are used to support the archaeological investigation of a particular problem and / or question. Knowledge and understanding of the principles, requirements and type of data produced by the various scientific methods is a prerequisite in order to be able to understand the structure of the database and the nuances of the data (i.e.., 'read between the lines'). Secondly, in order to successfully tackle an archaeological problem, the user should also learn how to:

- interpret laboratory results/scientific datasets from an archaeological point of view
- manage and read beyond the data (metadata and paradata)
- incorporate laboratory results/scientific datasets into the investigation of an archaeological problem
- design archaeological projects that make use of scientific datasets

Several issues/challenges regarding the scientific datasets in the archaeological research also exist and should be confronted in order to optimize the availability, accessibility and interoperability of the scientific datasets. First of all, such scientific datasets should compose an inextricable part of each archaeological dataset rather than an independent set of additional information. Along with the archaeological observations (for instance condition of a finding, dimensions of an object, estimated period etc), all the available scientific data should also be recorded in order to create a complete ‘identity’ record for all archaeological findings or the environment, which should properly be updated after new analyses and/or findings.

Scientific datasets are not created only for in-house use, but also aim at corroborating the work of other researchers as well. However, the use of different terminology and definitions may create confusion to scientists of different scientific disciplines, which produce data for the archaeological research. Consequently, one of the keynote challenges is to agree on common terminology and definitions. Even the definition of the term *‘data’* may confuse the scientific community. For example, in the case of the determination of the chemical composition of a ceramic artifact by means of X-ray Fluorescence Spectroscopy, some would regard the concentration of the elements as data, while others would consider the unprocessed information, namely the acquired (utterly qualitative) spectrum, as data!

To avoid any further confusion in the present document the term *‘Data’* or *‘Scientific Data’* refers to raw data, i.e., results of every measurement directly produced by the various scientific instruments without any intermediate calculations. Also the term ‘*metadata’* as is used here, unless explicitly mentioned otherwise, includes also the *‘paradata’* related to a measurement.

4.2 Metadata and paradata

There is seldom a perfect, complete and devoid of any ambiguities and/or outliers set of data. Consequently, to ensure that data is not misused, the assumptions and limitations affecting the creation of data must be fully documented, meaning that besides the acquired data, appropriate *metadata* and/or *paradata* should also be considered in a complete scientific dataset. Metadata allows a producer to describe a dataset fully so that users:

- can understand the assumptions and limitations
- evaluate the dataset's applicability for their intended use
- efficiently use scientific datasets in their research (especially when comparisons are attempted)

Few examples of metadata that could or should be documented in a scientific dataset are given below:

- laboratory, personnel who produced/created the data
- date(s) when the measurements/analyses were done
- experimental method/technique used
- instrumentation used for the production of data
- calibration method-data and standards used

In the same respect, paradata (a.k.a. operational parameters/conditions and procedure) can also be very important for the certification, assessment of limitations, and ‘reproducibility’ of the scientific datasets.

Documentation of all the above and their incorporation in a scientific dataset may be more significant for research purposes, such as:

- to compare the results and/or re-estimate the precision of the measurement (e.g. instrument and/or method employed)
- to repeat the experiment (e.g. protocols/conditions of the measurement)
- to recalibrate the measurement (e.g. reference data and/or standards used)
- to use exactly the same method/conditions for new samples in light of new/additional archaeological finds

- to determine the provenance of the extracted data (spatiotemporal information, such as who, where and when the measurement was conducted)
- to evaluate and certify the interoperability of the data

4.3 Metadata standards

According to the above, the adoption of a metadata standard seems imperative in order to assure the credibility and durability of a scientific database, since it would encompass only the necessary data and metadata devoid of any excessive information. Furthermore, the database should be structured in a way that could be expanded towards different kinds of information in order to permit inter-comparison among different techniques, towards the implementation of multidisciplinary studies (Hein and Kilikoglou, 2012). A metadata standard should meet several requirements and define (ISO, 2003):

- mandatory and conditional metadata sections and elements
- a minimum set of metadata required to serve the full range of metadata applications (data discovery, determining data fitness for use, data access, data transfer, and use of digital data)
- optional metadata elements – to allow for a more extensive standard description of the data, if required
- a method for extending metadata to fit specialized needs, since numerous scientific methods are used as diagnostic tools in the various fields of the archaeological research (e.g. dating, provenance, dietary studies, gene studies etc.)

Implementation of a metadata standard in the scientific datasets produced in the archaeological research would have many benefits, since it would (ISO, 2003):

- provide data producers with appropriate information to characterize their data properly
- facilitate the organization and management of metadata
- enable users to apply data in the most efficient way by knowing its basic characteristics
- facilitate data discovery, retrieval and reuse. Users should be better able to locate, access, evaluate, purchase and utilize the data
- enable users to determine whether data in a holding will be of use to them.

Agreement on common metadata standards is vital for the scientific community which serves the broader field of the archaeological research in order to bring together and integrate existing archaeological research data infrastructures and facilitate their accessibility and exploitation by researchers. A formal standardized structure that provides a common and extensible semantic framework that any cultural heritage information can be mapped to, should be established. Such a modeling of information will assure the credibility and durability of a scientific database, since it will encompass only the necessary data and metadata, which along with their interdisciplinary character should contribute to the creation of an integral archaeological research infrastructure.

In this respect, several ontologies/metadata standards are available, which are adopted for various natural sciences. Few examples are the DIF (Directory Interchange Format) (Olsen & Chiddo, 2008), used for exchanging information about scientific data sets primarily for earth sciences and the ISO 19100 series (especially ISO 19115) for geographic information metadata (ISO, 2003). In addition, the CIDOC-CRM is intended to promote a shared understanding of cultural heritage information by providing a common and extensible semantic framework that any cultural heritage information can be mapped to. However, CIDOC-CRM defines how cultural heritage information can be exchanged and how semantic relations can be arranged, but does not define what should be documented (Schröttner et al., 2012). The development of an extension to this (CIDOC-CRMsci) with modifications and improvements has being iniated in the framework of the ARIADNE project (FP7-INFRASTRUCTURES-2012-1-313193) (e.g. Doerr and Hiebel, 2014), with the aim to:

- assure credibility and durability of datasets
- maximize the info with minimum requirements
- establish the use of common terminology and definitions
- provide the common denominator for all sciences and studies involved (i.e., facilitate the interdisciplinary character)
- facilitate the building of an integral archaeological research infrastructure

4.4 Metadata standard in scientific datasets of archaeological research

As already discussed, contemporary archaeological research composes several scientific disciplines and incorporates different measurements and analyses conducted each one with its own identity and scientific background (e.g. isotope analysis, TL/OSL dating, DNA studies, composition analysis of organic samples or materials etc). To establish a 'universal' standard that covers all areas and their individual nuances and special details seems extremely complicated and difficult if not impossible. As an alternative, a common denominator should be found, namely the minimum required metadata fields and dataset structure in order to adequately record and describe all (most) possible (scientific) measurements that may be conducted to serve as tools for answering archaeological questions.

Table 1 presents the common metadata fields required (according to the authors' opinion) in a dataset, independently of the scientific field they originate from, in order to allow the verifiable accurate and efficient use, interoperability, comparison, update and reusability of these datasets. The Table is based on standard 'best practice' procedures (GLP, ISO) followed by scientific laboratories, as well as existent and accessible scientific datasets (web and publications). In addition, the Directory Interchange Format (DIF) (Olsen & Chiddo, 2008), which is an approved standard recommended for use in NASA Earth Science Data Systems and provides a metadata format used to create directory entries that describe scientific data sets, has also been taken into account with some necessary modifications.

The required metadata could be divided into three levels, Project level, Object level and Measurement level metadata. Each one provides different information. The project level metadata mainly refers to information related to the archaeological site under investigation (where the samples were collected). The object level metadata provide information (mainly archaeological) and description of every 'object' (or 'sample' in general) to be measured/analyzed. Finally, the measurement level metadata is related to all available information about the experimental/measuring procedure(s) including any post-measurement data treatment. Based on the above a project may include multiple measurements of several objects (samples) found (collected) in the same excavation site. Although the more metadata the better, in most cases a certain number of metadata is enough to establish their usability for future purposes.

It should be noted that the compilation *Table 1* was heavily based on the information needed for provenance studies since these are usually more involved, can be conducted using various methodologies and measurement techniques, and the respective data, by the very nature of these studies, are most often used for comparisons while the augmentation of these datasets is a constant pursuit of the researchers involved. The metadata included in Table 1 should be sufficient to cover most other cases. However, due to the multitude of methodologies and measurements used in archaeological studies and the continuous enrichment of those, thanks to the advancements in physical sciences, the possibility of some necessary additions cannot be excluded. As a result, at this stage, no attempt has been made to suggest/provide a schema that conforms to or is aligned with other established schemata or initiatives (DCMI, CIDOC-CRM). It is the authors' opinion that this Table should be first updated and streamlined to cover the needs of as many as possible scientific disciplines (or at least the most used) and then propose a final schema in accordance to internationally accepted standards.

Table 1. Metadata fields that should be filled when archiving scientific data

Field	Description	Field importance*
	Project level	
Name entry	Title of the data set described by metadata	M
Summary	Description of the data set along with the purpose of the data	M
Keywords	Keywords that are representative of the data set	M
Language	Language used in the preparation, storage and description of the data	D
Project name	Name of the project	M
Project date	Date the project started	M
Location	Geographical area, where the object(s)/sample(s) was (were) found/collected	M
Site geographical longitude	Longitude of the Location	D
Site geographical latitude	Latitude of the Location	D
	Object level	
Object/sample ID code	The inventory ID of the object/sample	M
Object/sample lab ID (if applicable)	The ID of the object/sample as registered by the laboratory performing the measurement (if different than above)	M
Object/sample lab date	Date the object/sample was delivered to the laboratory	M
Batch protocol number	The protocol number of the batch in which the object/sample belongs to	D
Excavation section	The section of the excavation site at which the object/sample was found/collected	O
Area description	Description of the area/environment where the object/sample was found/collected	M
Object/sample type	The type of the object/sample (e.g. soil, clay, sherd, pollen, seed, other)	M
Object/sample classification	The 'category' where the object/sample belongs to (taxon, vessel type etc. depending on the object/sample)	M/D (depending on the object/sample)
Object/sample dimensions (if applicable)	The dimensions of the object/sample	O
Object/sample description	Free text describing the object (form, condition, colors, texture, etc)	O
Object/sample picture	Picture of the object in tiff or jpg format	D
	Measurement level	
Type of analysis	Dating, chemical analysis etc.	M
Laboratory_ID	The laboratory at which the measurement was conducted and the data were collected	M
Personnel	The person (researcher) who conducted the	O

	analysis and can be contacted for more information about the data and metadata	
Preparation method	Description of the preparation/pretreatment of the sample (cleaning, sieving, acid digestion etc)	M
Analytical method	Which method was employed for the analysis (for instance, XRF or AAS or… for chemical analysis, C14 or thermoluminescence or dendrochronology or… for dating etc)	M
Technique	The specific technique used (for instance flame or graphite furnace in AAS)	M
Instrument(s)	Instrument(s) used for the measurement (brand name, model and serial number)	M
Instrument(s) details	Other instrument details (especially those that are unique for the specific instrument (s) used)	D
Start measurement date	The date the measurement of the sample started	M
End measurement date	The date the measurement of the sample ended	M
Measurement protocol	Steps and operational conditions of the measurement	M
Calibration method and standard materials/reference specimens	Calibration procedure and/or the standard materials/reference specimens used	M
Other conditions	Description of specific experimental conditions not stated explicitly elsewhere but can affect the measurement results (e.g. optical filters used in TL/OSL measurements)	M
Software (acquisition)	The specialized or commercial software used to acquire the data	O
Raw data	Original measured values without any mathematical treatment	M
Results	Results of the measurement following any necessary mathematical treatment of raw data	M
Error	The estimated error of the measurement	M
Methodology (alternatively software used)	Methodology or commercial software used to produce the results from the raw data	M
Additional files	Images, Spectra or charts corroborating the interpretation of the results	O

**M: Mandatory, D: Desirable, O: Optional*

5 CASE STUDY-POTTERY PROVENANCE STUDIES

To elaborate on the above and better realize the identity and complexity of scientific data in archaeological research and the need of a concise and complete set of metadata, the procedure followed during a pottery provenance study is described in detail below. A provenance study is one of the best examples of the synergy of various sciences and methodologies in archaeological studies (chemistry, physics, geology, petrography, mathematics, informatics, etc).

Provenance of ceramics, namely the answer to the question of where the artifacts were produced, is often a critical issue in archaeological studies, since it can illuminate various aspects of the civilization that produced them, such as trade routes and patterns, economic interactions between communities in a broader area and even of the technological level of an era. Using analytical methods several measurements are conducted to determine the chemical composition, physicochemical properties and structure of the materials along with the morphology and the physicochemical parameters of the excavation environment. The above, in conjunction with the statistical processing of the results, provide numerous important scientific data which represent a unique "fingerprint" for each artifact that can shed light on its origin, since featural variations between sources should be greater than within sources (Rice, 1987; Sterba et al., 2009; Tite, 2008).

5.1 Pottery in scientific studies

Selection of pottery in provenance studies (and in the majority of the scientific studies assisting the archaeological research) seems more advantageous compared to any other material due to the features they exhibit, which allow the evaluation of their initial state and the extraction of valuable information from their composition (e.g. Padeletti and Fermo, 2010). More specifically:

- they have a long history of existence and can be found in great abundance virtually all over the world
- they exhibit remarkable resistance to weathering and erosion along with tolerance through time
- they are non-perishable, although they may break, fragments (sherds) are virtually indestructible
- pottery have virtually endless shapes and decorations and multiple functions, were commonly used in everyday life and easily transported.

As a result the scientific study of pottery can provide valuable information regarding the manufacturing method, which can be described as "an additive process in which the successive steps are recorded in the final product", gaining more knowledge about the past since pottery manufacture, like any other productive technology, represents a point where a cultural system interacts directly with its surrounding socio-cultural and natural environment.

According to the above, pottery represents not only a category of useful containers, but also a simple and convenient means of:

- dating sites
- tracing trade patterns (local and long distance trade arrangements)
- studying ancient technology (resource selection, forming techniques, firing strategies)
- investigating settlement patterns and demographic factors
- studying a variety of social aspects (dietary habits, ceremonial or ritual activities etc).

5.2 Objectives of pottery provenance studies

Pottery provenance studies can have various scopes and the used methodology may involve additional steps or follow specific constrains. More specifically, provenance studies can aim at the (Kazakis & Tsirliganis, 2015):

- *Identification of the provenance of ceramic artifacts of unknown origin*

In such a case, the archaeological observation is not adequate to serve as a representative guide for the provenance study and consequently, the use of reference samples, namely samples of known provenance (ideally ceramics found at a workshop) is required.

- *Classification of ceramics of the same region*

In this case, the provenance study is conducted towards the identification of know-how and craftsmanship of certain production workshops in the same region or different geographical location. Such a successful study requires an adequate number of samples of each potential group in order to assure the credibility of the results.

- *Confirmation of the provenance of ceramic samples with “suspected” origin*

A pottery provenance study may also be employed to determine the origination of pottery which appears to have many stylistic similarities with major and “popular” ceramic categories with historically known production workshops (e.g. Attics in ancient Greece). Such a study can basically be regarded as an authenticity test, while it requires a reference group of the ‘genuine’ articles.

- *Building archaeological databanks*

All scientific data produced during a pottery provenance study in conjunction with the archaeological data allow the full documentation of the ceramics studied and the creation of large archaeological reference datasets. As a result, new reference groups are produced, while existing ones can be expanded, improved and/or modified.

5.3 General methodology in pottery provenance studies

The methodology adopted during pottery provenance studies can be briefly described with the following steps:

- a small fragment of a ceramic sherd is subjected to one or more laboratory analytical methods towards its qualitative and quantitative characterization and the identification of its unique fingerprint that allows its discrimination among others of different origin or its grouping with those exhibiting similar features
- next, multivariate statistics are used to determine similarities and differences between the specimens, which ultimately lead to their classification into distinctive groups according to their provenance
- final conclusions are usually reached with the juxtaposition of the laboratory/statistical and the archaeological data (primarily based on stylistic features)

5.4 Procedures and scientific data sources in pottery provenance studies

Besides the macroscopic visual observation of any stylistic properties on the under-study pottery, which is in principle carried out by archaeologists, several other laboratory scientific procedures are involved in pottery provenance studies which produce numerous data. Easily employed measurements of various physical properties of the clay, such as color (e.g. Munsell color system), density and hardness take place first, while more advanced approaches follow for the in-depth material characterization. The latter composes mainly the chemical analysis of the elemental bulk concentrations and the mineralogical investigation.

The *chemical analysis* is oriented towards the identification of the chemical elements constituting the ceramic fabric, present in major, minor or trace amounts, which could provide a unique chemical profile allowing ceramics made from the same raw materials to group together during the statistical analysis of the data. For this purpose, various methods can be employed (***Table 2***) (Kazakis & Tsirliganis, 2015) individually or in combination (two or more) in the same study in order to increase the validity of the results. Furthermore, in some cases one method may complement another, since accuracy and sensitivity of the various methods varies and in some cases they are element dependent. Consequently, one method may be used for the determination of the major elements, while a second one may be employed for the trace elements.

On the other hand, the *mineralogical investigation* aims at identifying the speciation of the geological components in the ceramic fabric and assessing the potential origin of the raw materials. Mineralogical study is focused on the analysis of the temper, which was added by potters in order to modify as desired the properties of the clay and the quality of the final product. The geological data and information acquired by such a study can increase knowledge regarding the selection and use of local and non-local resources by potters. As in the case of the chemical analysis, various methods can be employed for the mineralogical investigation of pottery, the most important of which are presented in ***Table 2***.

Table 2. Instrumental methods of chemical and mineralogical analysis of ceramics.

Chemical analysis	*Mineralogical investigation*
Neutron activation analysis	Thin-section petrographic analysis
X-ray fluorescence spectroscopy	Scanning electron microscope
Atomic absorption spectrometry	X-ray diffraction
Electron microprobe analysis	
Proton-induced X-ray emission	
Scanning electron microscope + EDS/WDS	
Inductively coupled plasma-atomic emission spectroscopy-mass spectrometry	
Mossbauer spectroscopy	

The above two approaches (chemical and mineralogical analysis) of ceramics characterization provide different kinds of scientific data which complement each other (Rice 1987; Wilson 1978). In most cases, the use of solely one of the above analyses is not adequate to determine with high certainty the provenance of the pottery studied (Mommsen, 2001; 2004). Chemical data alone do not identify fully the geological resources (raw materials) used and do not provide information on ceramic technology and culture specific traditions or traditions of households or potters within a settlement (e.g. quality of raw material preparation; orientation of non-plastics; mixing of raw materials – different raw materials "meet" when potters did not homogenize them properly; type/size/amount of temper – which may be culture specific in a given period, such as chaff or grog tempering), while petrographical analysis alone cannot account for possible phase transitions of the minerals due to the firing of the ceramics and very small inclusions cannot be identified properly even though they may be vital for appropriate provenancing of imported wares or distinguishing between local clay sources thus between different coexisting traditions. As a result, the *state-of-the-art* in the provenance studies of ceramics dictates the adoption of an "*integrated*" approach, in which the above analyses are combined, while additional measurements of the microstructure and other physicochemical and mechanical properties of the pottery are more than desirable (e.g. Mirti et al., 1996; Moropoulou et al., 1995; Rice 1987; Wilson 1978). The amalgamation of all data obtained from the above scientific activities can offer the optimal characterization of ceramics which, aided by advanced specially tailored statistical methods (Principal Component Analysis or Hierarchical Cluster Analysis), will provide an accurate and distinct classification of ceramics leading to satisfactory, secure conclusions on their origin.

5.5 Scientific data in pottery provenance studies

According to the above, it is evident that a pottery provenance study combines many natural sciences which serve as detection tools to extract all possible underlying information that will promote the archaeological research. Regardless the approach adopted in a pottery provenance study, even if a single analytical method is used, various scientific data are produced demanding different handling and processing. To elaborate on the above, ***Figure 1*** illustrates the various scientific data produced during a pottery provenance study.

Figure 1 reflects the variety and complexity of the scientific data produced in general in the archaeological research, which can have:

- Various ***formats***, such as numbers, spectra, diagrams, pictures, models and maps.
- Various ***types***, such as raw data, calculation results, tables, plotted data and statistical data.

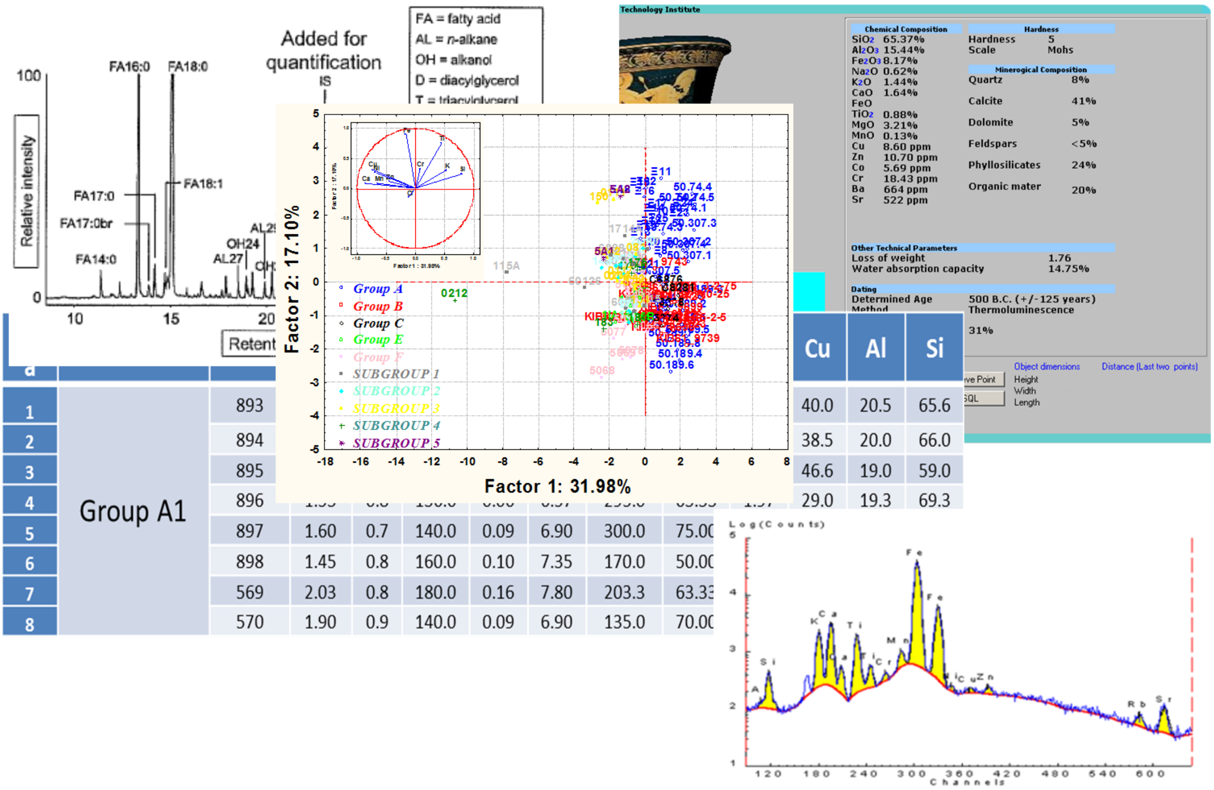

Figure 1. Example of scientific data produced during pottery provenance studies.

All the extracted data can be used either for a general approach to an archaeological problem or a 'customized' solution to a specific case. According to the above, the scientific datasets demand/involve a three-level interpretation (***Figure 2***):

- *1st level → scientific interpretation (results-conclusions)*

In the 1st level, data/results are interpreted and evaluated strictly from a scientific point of view. Inspection of the various data (e.g. elemental concentration) and comparison between the values is made, allowing the extraction of conclusions, which do not have tangible meaning to the archaeologist, but provide the basis on which the 2nd level of interpretation will take place.

- *2nd level → 'real life' meaning*

The 2nd level of interpretation involves the decoding of the scientific data and the conclusions of the 1st level into 'real life' meaning results. In this level, scientific data actually answer the archaeological problem and become of practical use.

- *3rd level → implications for the specific archaeological problem*

In the final level of interpretation, the new knowledge gained by the acquired scientific data may induce implications for the specific archaeological problem and create new doubts and/or questions to the archaeologist (e.g. the "known" provenance of one or more ceramic samples may be put into question).

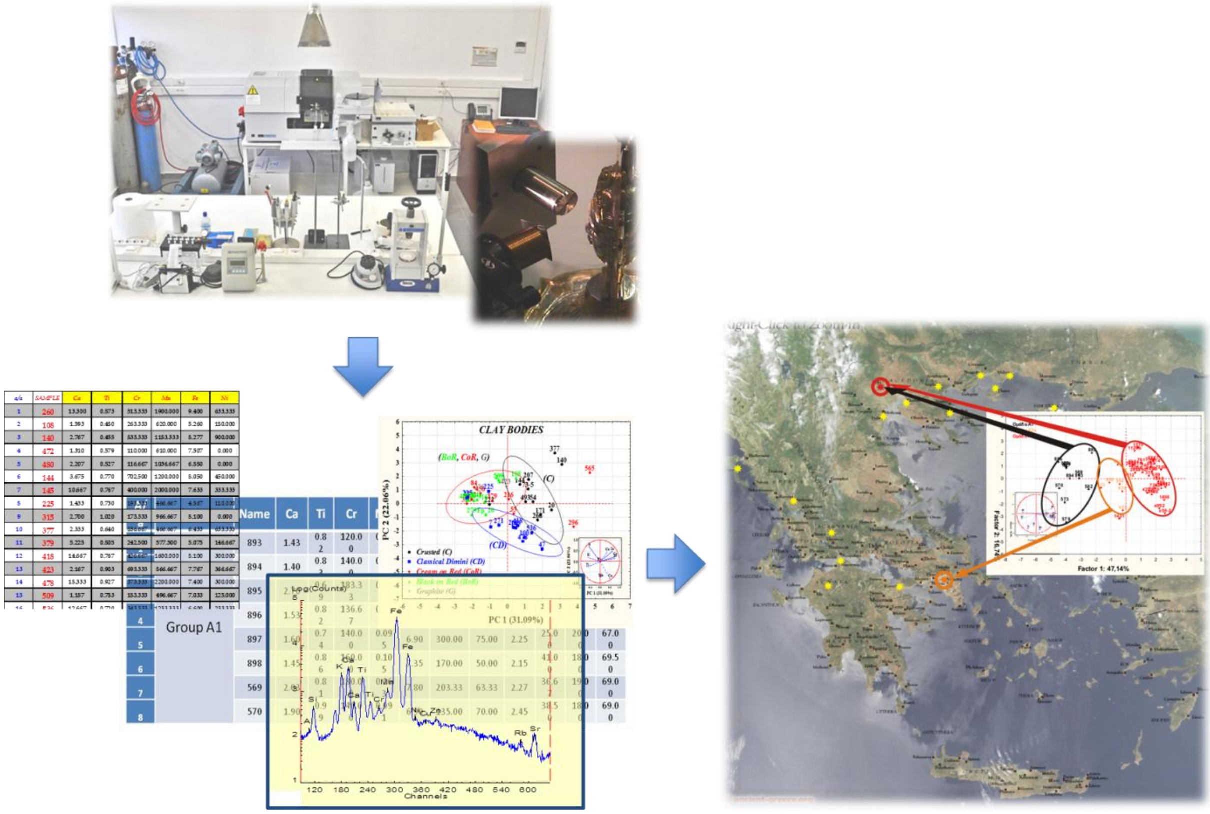

Figure 2. From the production of scientific data to their 2nd level interpretation.

6 CONCLUSIONS

Archaeology uses a variety of methods and tools to reconstruct the cultural past. Based on excavation finds, concerted efforts are made to shed light on the structure of the ancient societies, their environment and their socio-economical interactions. The above is mainly accomplished through the synergy of archaeology and various fields of natural sciences and informatics, which allow the extraction of information by means of advanced laboratory methods.

As a result a plethora of scientific data (of various formats and types) and metadata are produced creating large and complex scientific datasets. The role of these datasets is twofold. First, they can be used either for a general approach to an archaeological problem or a ‘customized’ solution to a specific case. Secondly, they are very important in order to repeat an experiment, to validate the results in light of new measurements and allow comparison between different measurements.

However, such scientific reference datasets used in the archaeological research exhibit large variability among the various research infrastructures. Agreement on common definitions, purposes and structure as well as the establishment of standards regarding these scientific datasets is a requisite in order to bring together and integrate existing archaeological research data infrastructures and to facilitate their accessibility and exploitation by researchers.

In this respect, several ontologies/metadata standards are available, which are adopted for various natural sciences, however, the establishment of a ‘universal’ standard that would cover all disciplines is extremely complicated and difficult. As an alternative, a simplified dataset structure including the minimum required metadata fields is suggested in order to in order to allow the verifiable accurate and efficient use, interoperability, comparison, update and reusability of these datasets independently of the scientific field they originate. This structure was based on standard ‘best practice’ procedures followed by scientific laboratories, as well as existent and accessible scientific datasets (web and publications), while no attempt has been made at this point to provide a schema that conforms to or is aligned with other established schemata or initiatives.

Due to the multitude of methodologies and measurements used in archaeological studies the possibility of some necessary additions to the suggested structure cannot be excluded. These additions and the finalization of this structure with its compliance to international standards is an on-going work that requires the collaboration of specialized interested parties.

7 BIBLIOGRAPHY

ALSHEIKH-ALI, A., QURESHI W., AL-MALLAH M.H., IOANNIDIS J.P.A. (2011) Public availability of published research data in high-impact journals. *Plos One*, 6: e24357.

BISOL, G.D., ANAGNOSTOU, P., CAPOCASA, M., BENCIVELLI, S., CERRONI, A., CONTRERAS, J., ENKE, N., FANTINI, B., GRECO, P., HEENEY, C., LUZI, D., MANGHI, P., MASCALZONI, D., MOLLOY, J.C., PARENTI, F., WICHERTS, J. M., BOULTON, G. (2014) Perspectives on Open Science and scientific data sharing: an interdisciplinary workshop. *Journal of Anthropological Sciences*. 92, 179-200.

BORGMAN, C. (2009) The digital future is now: A call to action for the humanities. *Digital Humanities Quarterly*. 3(4).

BUKO, A. (1984) Problems and research prospects in the determination of the provenance of pottery. *World Archaeology*. 15(3), 348-365.

CARLSON, S. (2006) Lost in a sea of science data. *The Chronicle of Higher Education*. 52(42). A35.

DOERR, M., and HIEBEL, G. (2014) *Developing common European archaeological concepts through extending the CIDOC CRM within ARIADNE*. [Online] http://www.ariadne-infrastructure.eu/content/download/1844/10313/version/1/file/CHNT-2013_04_GHiebel-MDoerr_cidoc-crm_archaeo-extensions_20131113.pdf [Accessed: 5th June 2015].

FREESE, J. (2007). Replication standards for quantitative social science. *Sociological Methods & Research*. 36(2), 153-172.

HARDING, D.W. (1994) Charles Francis Christopher Hawkes 1905-1992. *Proceedings of British Academy*. 84, 323-344.

HEIN, A., KILIKOGLOU, V. (2012) ceraDAT-Prototype of a web-based relational database for archaeological ceramics. *Archaeometry.* 54(2), 30-243.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. (2003) *ISO 19115: Geographic information - Metadata*. First edition, Switzerland. [Online] ftp://podaac.jpl.nasa.gov/misc/outgoing/ed/pre_2013/GHRSST_metadata/ISO%2019115%20.pdf. [Accessed: 5th June 2015].

Kazakis, N.A., Tsirliganis, N.C., 2015 Provenance of ceramics: Methods and Practices. In: Sarris A. ed. *Best Practices of GeoInformatic Technologies for the Mapping of Archaeolandscapes*. Oxford: Archaeopress (pp. 241-252).

MATSUI, Y., GIDDING, A., LEVY, TH.E., KUESTER, F., DEFANTI, T.A. (2012) ArchaeoSTOR Map: Publishing Archaeological Geodata on the Web. *COM.Geo '12 Proceedings of the 3rd International Conference on Computing for Geospatial Research and Applications*, Washington, D.C., USA, July 1-3.

MIRTI, P., CASOLI, A., and CALZETTI, L. (1996) Technology of production of fine pottery excavated on a western Greek site investigated by scanning electron microscopy coupled with energy-dispersive X-ray detection. *X-ray Spectrometry*. 25(3), 103-109.

MOMMSEN, H. (2001) Provenance determination of pottery by trace element analysis: Problems, solutions and applications. *Journal of Radioanalytical and Nuclear Chemistry*. 247(3), 657-662.

MOMMSEN, H. (2004) Short note: Provenancing of pottery-The need for an integrated approach?. *Archaeometry*. 46(2), 267-271.

MOROPOULOU, A., BAKOLAS, A., and BISBIKOU, K. (1995) Thermal analysis as a method of characterizing ancient ceramic technologies. *Thermochimica Acta*. 269-270, 743-753.

OLSEN, L., and CHIDDO, C. (2008) *Directory Interchange Format (DIF) Standard*. [Online] https://earthdata.nasa.gov/files/ESDS-RFC-012v1.pdf [Accessed: 5th June 2015].

PADELETTI, G., and FERMO, P. (2010) A scientific approach to the attribution problem of renaissance ceramic productions based on chemical and mineralogical markers. *Applied Physics A: Materials Science and Processing.* 100(3), 771-784.

RICE, P.M. (1987) *Pottery Analysis: A Sourcebook*, The University of Chicago Press, Chicago and London.

SCHRÖTTNER, M., HAVEMANN, S., THEODORIDOU, M., DOERR, M., FELLNER, D.W. (2012) A generic approach for generating cultural heritage metadata. *EuroMed'12 Proceedings of the 4th international conference on Progress in Cultural Heritage Preservation*. Limassol, Cyprus, October 29-November 03, 231-240.

STERBA, J.H., MOMMSEN, H., STEINHAUSER, G., and BICHLER, M. (2009) The influence of different tempers on the composition of pottery. *Journal of Archaeological Science*. 36(7), 1582-1589.

STRASSER, C., KUNZE, J., ABRAMS, S., CRUSE, P. (2014) DataUp: A tool to help researchers describe and share tabular data [version 2; referees: 2 approved]. *F1000Research*, **3**:6.

TITE, M.S. (2008) Ceramic Production, Provenance and Use-A Review. *Archaeometry*. 50(2), 216-231.

VINES, T.H., ALBERT, A.Y., ANDREW, R.L., DEBARRE, F., BOCK, D.G., FRANKLIN, M.T., GILBERT, K.J., MOORE, J.S., RENAUT, S., RENNISON, D.J. (2014) The availability of research data declines rapidly with article age. *Current Biology*. 24, 94-97.

WILSON, A.L. (1978) Elemental Analysis of Pottery in the Study of its Provenance: A Review. *Journal of Archaeological Science*. 5(3), 219-236.