HUNGARIAN ARCHAEOLOGY
E-JOURNAL • 2016 AUTUMN
www.hungarianarchaeology.hu

NEW METHOD IN SARMATIAN RESEARCH: The Database of Carpathian Basin Grave Goods[1]

András Balogh – Dávid Hüse – Eszter Istvánovits – Roland Kruska – Valéria Kulcsár

A peculiar contradiction in 21st-century Hungarian archaeology is that it looks upon the natural sciences as fields that are able to provide answers to fundamental questions, and it tries to use their new methods as a 'wonder tool' in on-site research. In contrast to this, it still has yet to properly exploit the possibilities offered by information science. The reason for this is perhaps that while it receives the former as a final answer, the development of a database that can be properly utilized requires a serious investment of time from the archaeologist working in collaboration with an information scientist. The time has come for the research into Barbaricum of the imperial period in the Carpathian Basin to process its finds into a proper digital apparatus. This is why we created the Sarmatian grave goods database (this deals with the archaeological materials from a population group settled on the Great Hungarian Plain during the time of the Roman Empire, but outside its territory). The opportunity arose within the context of a National Scientific Research Fund Programs (OTKA) grant (2013–2017) to record and develop a chronology of find materials from Sarmatian graves. The project leader is Valéria Kulcsár, and the participants are Gyöngyi Gulyás, Eszter Istvánovits, Anita Korom, and Kornél Sóskúti, as well as Gábor Pintye, Zoltán Rózsa, Dávid Schwarcz, and Gabriella Vörös as consultants. The institutional framework for the grant project is provided by the University of Szeged.

BASIC PRINCIPLES OF THE DATABASE

A ready-made database based on the special features and find materials of a given period cannot ever be transplanted to a different area, period, or set of finds, but there is some possible 'borrowing' between periods. A developed and functioning database is easier to remodel taking into account the special features of another period than to make one without any precedent; the method can be drawn upon and the existing structure utilized.

In designing the database, the following considerations were taken into account: 1) it must be able to operate as part of a complex GIS framework system; 2) it must provide the opportunity for online publication; and 3) the serialization must be able to be 'built upon' (a mathematical/statistical method, which in this case is based upon the combination of finds and find contexts).

The GIS is an adaptation to computers of the traditional use of maps, and from our point of view a database is needed that can develop a two-way link between the cartographic information and other types of data entered into the system. Through the aid of this, the locations of particular features (e.g. kurgans, and piled stone graves), object types (e.g. *terra sigillata* wares and enameled fibulae) and their combinations (e.g. burials with enclosing ditches that contain knives), as well as the location of particular features/object types within the same cemetery can be quickly and effectively surveyed and analyzed. The link is two-way: on the one hand the screening can begin through the database with the charting of individual object types, the combination of types, ritual elements and the combination of rites, while on the other hand it can provide us the information related to a grave at a particular point within a cemetery map, if necessary.

The database will appear as part of an online publication. The advantage of this over traditional catalogues, text corpora, and essays is that it provides access to an amount of material that cannot be transmitted by traditional means. It can be used by 'guest researchers' according to their own points of view (they do

[1] The project outlined in this article is being implemented through support from grant number OTKA K 104980.

not need to print out the information and then organize the finds according to their own considerations). It can be constantly kept up-to-date. It must be kept in mind that a database always contains 'sensitive data' (e.g. geographical coordinates) that are not public, as well as unpublished ensembles that are still being excavated. Due to this, screening before publication must be dealt with. Among the ideas is the making parts of the database public with the sensitive data blocked out.

There is a dark side to keeping the database constantly up-to-date: the corrections cannot be traced as easily. It is worthwhile to pay attention to this and meticulously indicate modifications.

THE INFORMATION SCIENCE BACKGROUND TO AN ARCHAEOLOGICAL DATABASE: POSSIBILITIES AND LIMITATIONS

The *security* of the data is essential, including safe storage and access linked to levels of authorization. The human factor can be essentially fully excluded from our multiphase data archiving system, the data archiving automatically runs at given intervals and then an information specialist checks the data and the processes for saving the information. The saving of the information occurs in a multiphase system: first on data storage devices that mirror one another, and then at intervals on a network storage device as a back-up copy (a copy is made of both the descriptive data and the entered stock, which is often a great deal of data). In the third phase, saving information following changes occurs on a storage device at a different location and/or town (a large bandwidth is necessary for this). The processes run during non-working hours, so that they do not take up all the bandwidth and do not burden the system. (We do not record on paper, but it is theoretically possible.)

Several user groups and the authorization levels linked to these have been developed. Registration and identification of the user is necessary for use; and the functions and content are available in accordance with one's authorization.

In the last two decades, museums have had to perform significant professional tasks in relation to large projects. Excavations on large areas have provided enormous – nearly unmanageable – amounts of information. An important consideration for our database was that in addition to new excavations *the earlier research, documented according to a different system, should be able to be integrated*. Research cannot be deprived of these observations, finds, and results just because large-scale excavations have taken place more recently. Today, a proper information system is indispensible for the storage and processing of data. This has been recognized by quite a few institutions, but due to the restraints of time and finances they tried to wrestle with these tasks each on their own, although there is a great need for a national, or even an international database covering several countries and a complex GIS system. The initial momentum bogged down as a result of continuous and repeated institutional reorganizations and numerous unfavorable processes. The provincial museums did not receive any kind of central assistance or support. Despite this, they are developing their own databases on their own and under difficult circumstances. There are institutions – at least among the provincial museums, which are the most affected by the large projects and institutional reorganizations – that are just struggling with the maintenance of the previously existing databases.

At a few collections, they have continued to actively utilize and develop their databases, sometimes entering quite a large amount of data. It is hardly biased to state that among the museum systems one of the most capable is the EMIR (Egységes Múzeumi Információs Rendszer – Unified Museum Information System) developed by Pazirik Kft., which served as a precedent for our database. Taking into account the history of their creation, this is not by chance. Following the wave of large projects in the first half of the 2000s, its development was aided by consultation with archaeologists and information specialists of institutions where previously they had tried using various databases in isolation. The accumulated experience and ideas have contributed to the creation of the EMIR system.

Currently it is being used by eleven institutions, who store several terabytes of data. The flagship of the development is the Jósa András Museum, where currently data on 13,500 features from nearly 2,000 sites have been entered with 33,000 documents (drawings, photographs, journals, maps, archival materials, etc.). In the inventory record that was recently begun more than 8,000 items can be found.

EMIR provides the opportunity to serve special, individual demands. One of its great advantages is its flexibility. The alteration of the structure that has been developed is of course a delicate operation (since every alteration has an impact on the entire system, and many things are interrelated in a complex system that becomes even more intricate through its development), but the individual elements can be quickly and easily expanded and modulated with proper discretion. This can be related to new types of objects, measurements, unique characteristics, etc.

The database is made up of *three modules that build upon one another*. The first is the site (*Fig. 1*). This was taken essentially unchanged from the version used by the museums. The individual features found at the sites can be linked to this, which is the second module, where special information appears. Since burials are the focus, it is possible to search according to more variables than in the database of the museums (e.g. marking of the graves, gender of the deceased, unique characteristics of the skeleton, individual elements of the burial rite, and size and character of the coffin). Linked to the features are the objects found within them (third module). The major innovations were introduced in connection with this, concentrating on objects characteristic of the Sarmatians, their dating, etc.

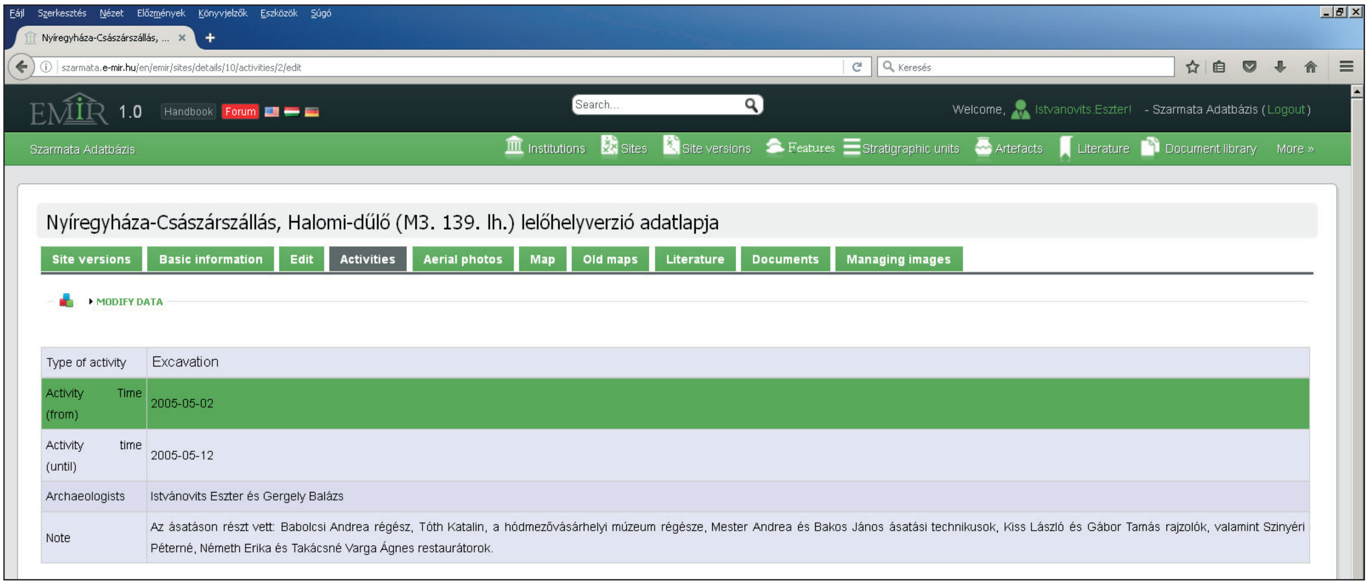

Fig. 1: The information supplementing the base data can be uploaded on the so-called tabs, for example data related to the excavation and its leader

Within the sites, the site version(s) is (are) always entered (*Fig. 2*). Research performed according to the various archaeological methods (field walks, excavation, etc.) provide differing types of information. The individual tasks performed at the sites can sometimes be linked to different experts. It is possible to indicate the later identification or verification of old sites. There are cases where the numbering of features was reworked. This problem cannot be dealt with by GIS either. It is significant that in EMIR these are uploaded as site versions, so the results of research performed on single sites but in different periods and often with differing methods are processed and analyzed separately. The versions are brought together at the level of the site.

The documents related to a site (photographs, aerial photographs, maps, old maps, excavation journals, etc.) can be uploaded. Further important information related to a site/site version (e.g. when it was excavated and by whom, who participated in the excavation and the literature) can be accessed by the tab appearing at the top of the page.

For a portion of the field work from the last decade the application of a legal protocol has been prescribed. A significant point in this is the use in parallel of stratigraphic numbers (phenomena, for example a stratum, are observed and isolated at the excavation and are given identification numbers), as well as the feature numbers (phenomena comprised of several stratigraphic units and therefore recorded with stratigraphic numbers, such as dwellings). In the case of earlier excavations, it was not done in this way. The

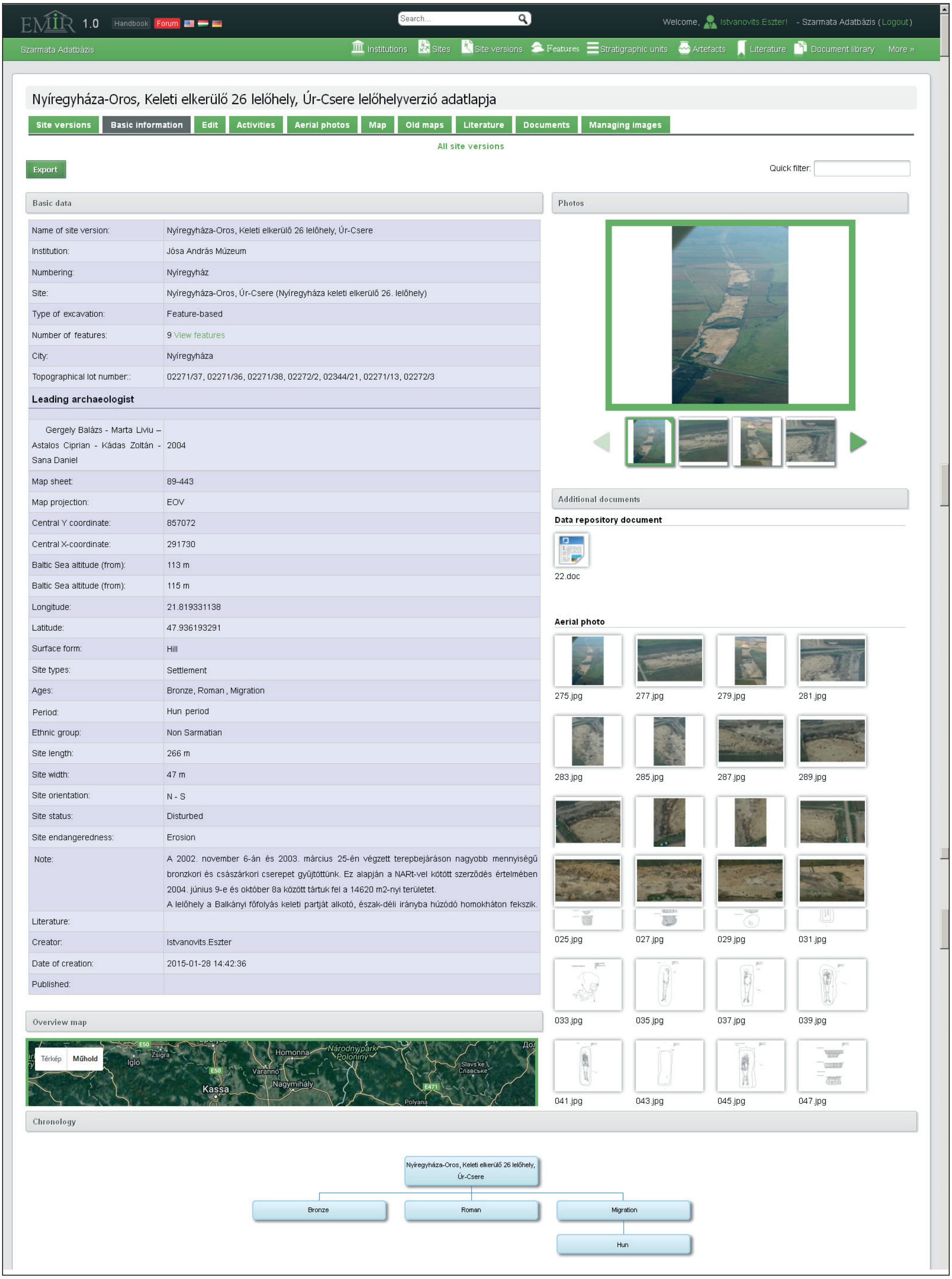

Fig. 2: The presentation of a site version in the database

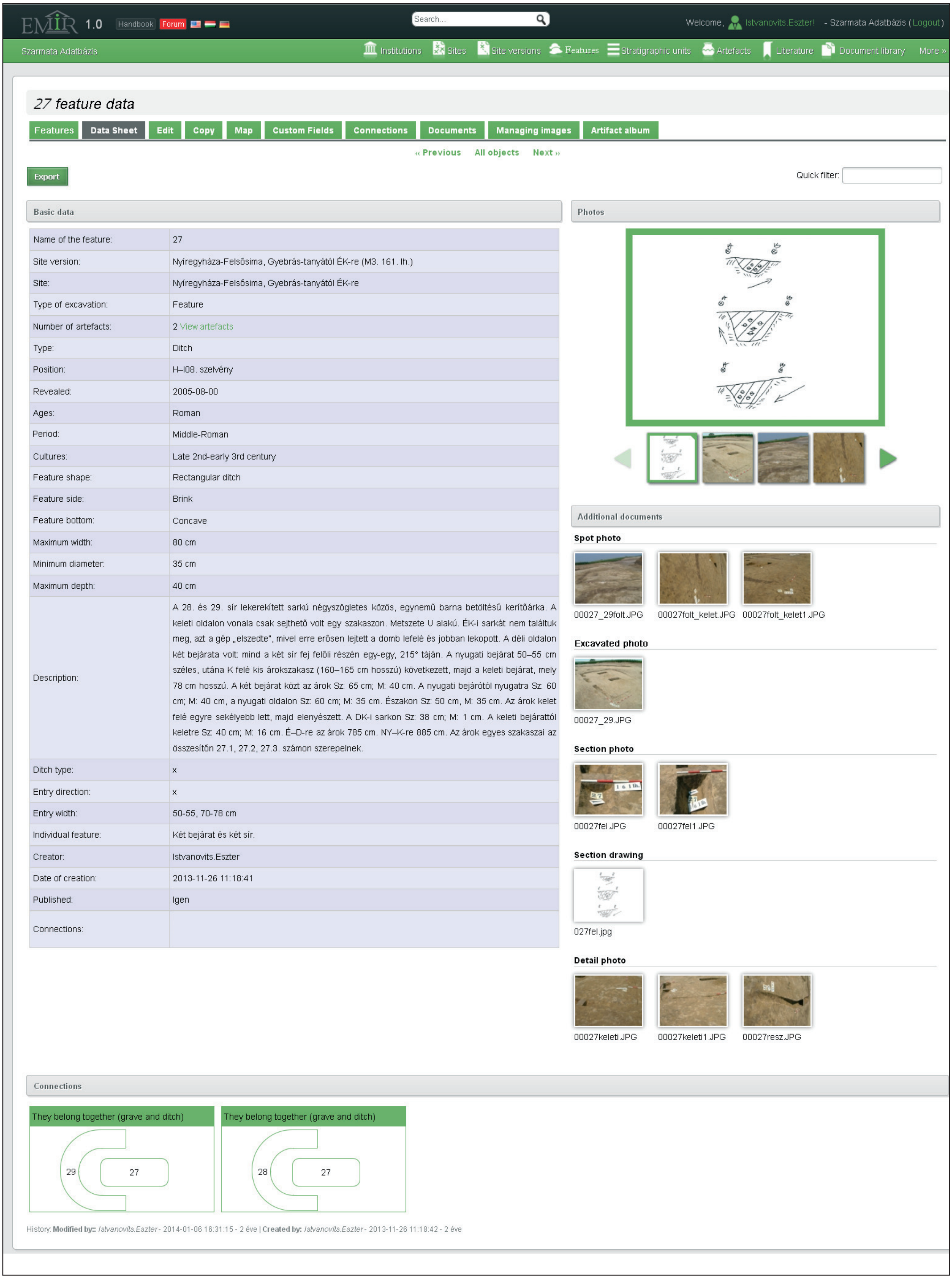

Fig. 3: General data related to a feature

use of the two systems together arose as a problem in the database. In EMIR, it was possible to resolve this issue in a manageable form. Which method was used was indicated on the site version. If stratigraphic numbers were used, several strata numbers are connected by a mutual feature number. If feature numbers were given, the relationship between two phenomena can be provided as a 'link', where its nature can also be defined (for example, superposition, that is strata located above or under one another, as well as their former connection), and this can also be indicated graphically (see the bottom of *Fig. 3*).

The individual features – whether they are graves, pits dug by robbers, ditches encircling the grave, kurgans, etc. – naturally have individual identification numbers and are closely linked with the site version. On the first page of the feature module (*Fig. 3*) general information is recorded that can be provided for any type of phenomenon (e.g. infill, size, form, etc.). There are characteristics that are only significant for certain types of features (e.g. graves) that are unnecessary for other types. In this case, special input fields are linked to the type, which only appear when this type is selected during the uploading of the general data. In the case of a grave, for example the orientation, the position of the body in the grave, gender, age, and individual elements of the burial rite are significant. If the (above) general information is entered, and a grave is indicated, then after these data are saved another window automatically pops up where the special information can be entered. Naturally, these appear together on the data sheet (see *Fig. 3*).

The finds are connected to the feature. For this module, it was necessary to establish a precise and consistently utilized typology. It is a fundamental expectation of databases that, on the one hand, it *should be simple and secure to upload the data*, and, on the other hand, that *the system be transparent, with queries being simple and multilayered*. A database fulfills these expectations if it is not roundabout, but is sufficiently articulated. A happy medium must be reached. Researchers are inclined to emphasize the unique qualities of an object. The database provides the opportunity for the indication of unique characteristics, but for the system to be unambiguous, these must be put in the background at first, and the common qualities emphasized.

The typology was constructed upon the model of a *plant identification system*. The 'classes', 'types', and 'variations' were made understandable through *drawings* so as to facilitate designation. Of course, the characteristics are far more multifaceted, and in no way could every significant factor be built in to the format (*Fig. 4*). The objects were placed within a simplified scheme, and then in the second step the unique qualities were taken into account. This was resolved through the ability to attach *special input fields* or *tabs* appearing at the top of the page to particular elements in the database. An example of this is shown in *Figure 5*.

The size, decoration, materials analysis, restoration, and literature, as well as the location of the object in the grave can be provided in a similar manner. The latter is able to link related objects (e.g. various types of beads on necklaces or the elements of a belt: buckle, strap end, "shepherd's purse", etc.). It is also possible to search for these data. Illustrative materials can be linked to the object. In the case of each feature, there is an opportunity to review the find assemblage.

From the first moment, attention was paid to make the database able to be serialized, or in other words that the finds and find groups be open to mathematical-statistical analysis, and the relationships and combinations that can be shown between them be open to presentation. The chronological determinability of the objects (ensembles) and phenomena (ensembles) was also designed in this way. The more precise definition of the chronology of the Sarmatian period was set as a final objective, and serialization as a method through the use of information science – for example on the basis of the experience from the Avar period – can provide a great step forward towards this. So-called contextual serialization is what we are talking about, when objects are categorized according to how commonly and in what composition they are found in each group of finds.

A 'cruder' and a 'more refined" chronological approach were separated from one another. The first was for the determination within a broad range, such as 'early imperial period', 'after the Marcomannic wars', etc. ('traditional dating'). This could be designated for both the features and the objects.

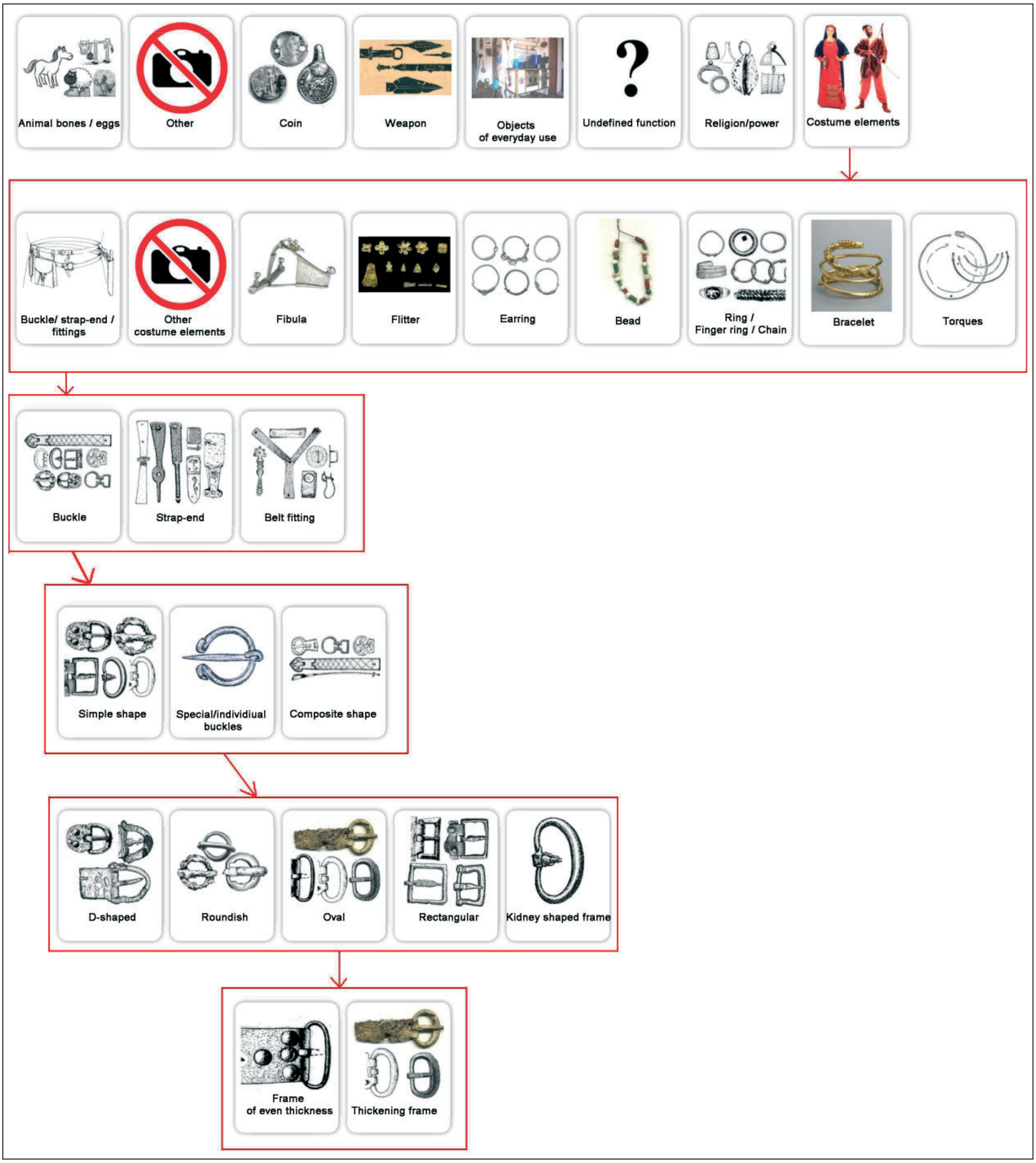

Fig. 4: Detail of the illustrated 'type species' for Sarmatian objects from the Carpathian Basin

The opportunity for a more precise chronological determination was also provided for the objects. These were separated into two categories. There are objects in Sarmatian finds whose age can be determined more precisely (for example, pieces of Roman glass, *terra sigillata* wares, coins, certain types of fibulae, etc.).[2] These are managed with a date range in a separate field ('exact dating'), indicating the related literature and the basis of the dating. It is not possible to disregard dating worked out from earlier research. There are finds

[2] We are aware of the problems from dual or even triple dating, but this is not the topic of the present essay.

(certain types of torcs, beads, ceramics, etc.) whose approximate time of use has been determined. These are placed in another – so-called 'traditional dating' – field; they aid in dating and with sufficient data their chronology can be made more precise. The professional literature is also indicated here. The third chronology field is the serialized date, which the object 'brings with it' from the composition of the find assemblage. For example, if a piece of *terra sigillata* and a Roman fibula that can be clearly dated are found in a grave, then it is possible to determine a *terminus post quem*, and the date of the burial can be determined with relative accuracy. All of the finds from the assemblage obviously were in use during the period determined by the two clearly dated finds. For these objects, the program enters the determined values in the serialized date field. All of these are data that can exclusively be filled in for individual finds, and their combination together provides the age of the grave, which the system fills in automatically. According to our concept, the program's serialized dating must be followed by an individual subjective decision, in which it is possible to approve or reject the dating, since, for example, the number of objects that provided the basis of this is of consequence. For example, if the type is known from fifty graves and each one is from the end of the 2nd century, then it must be accepted, and a type has been found that can further date finds coming from identical ensembles.[3] It may happen that alongside fifty earlier occurrences, one time a type may come from a later environment; this may have been inherited, and the chronology is acceptable. In contrast, if the number of identical types of objects is two, for example, their identical dating cannot be generalized.

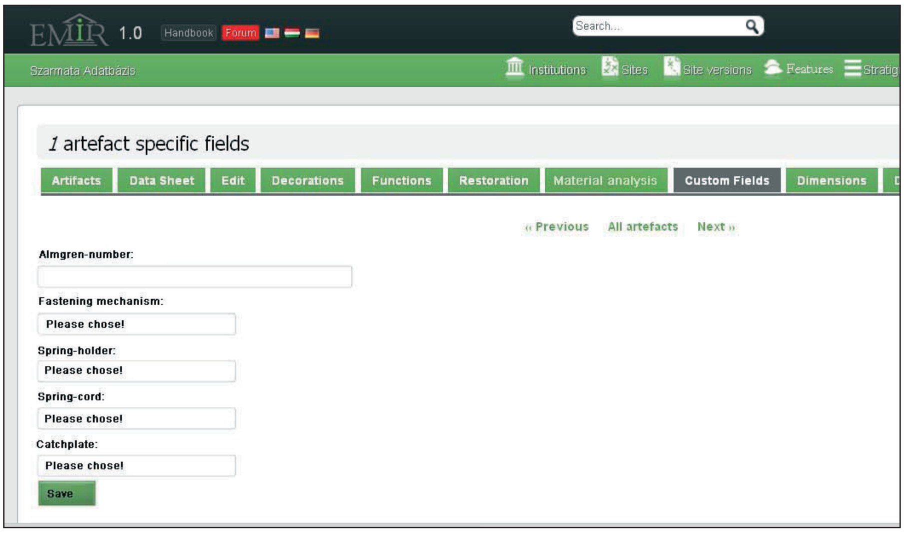

Fig. 5: Special fields for fibulae

FORUM AND HANDBOOK

Finally it is necessary to talk about the two tried and tested supplements for the use of EMIR, the *Forum* and the *Handbook*.

At the beginning of development, the need to continuously document changes arose. A joint Forum, independent of institutions and authorization, was created. It is here that alterations to be made, ideas and improvements are discussed. Users may open new topics and comment on previous ones. Every topic and comment can be accessed (indicating whether it is closed or not), and so the development of the database can be traced. Users receive an e-mail about new entries on the Forum. Images and text can be uploaded. The Handbook has also been completed with brief descriptions, images, and video materials.

It can be stated that a well-developed, thoroughly tested online database that works well as a part of a GIS system and has a great deal of uploaded information[4] is at our disposal for the processing of archaeological material, and this can serve as a basis for managing nationwide recording, a new inventory, or the accumulated data in thematic text corpora.[5]

[3] It follows from this that the more data available, the more likely it is that there will be a clue about the precise dating of each grave assemblage. The domino theory begins to work.

[4] As of June of 2016, 62 sites, 2,314 features (graves and archaeological phenomena related to them), 10,902 finds, and 21,223 documents have been uploaded.

[5] Our thanks go out to Dóra Mérai and Katalin Pintér Nagy for reading the manuscript of this essay and for making it more accessible to the reader through their comments.