Future Generation Computer Systems 56 (2016) 558–570

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

A generic framework for synchronized distributed data management in archaeological related disciplines

Johannes-Y. Lohrer [a,*], Daniel Kaltenthaler [a], Peer Kröger [a], Christiaan van der Meijden [b], Henriette Obermaier [c]

[a] Institute for Informatics, Ludwig–Maximilians-Universität, Oettingenstraße 67, 80538 Munich, Germany

[b] IT Group, Veterinarian Faculty, Ludwig–Maximilians-Universität, Veterinärstraße 13, 80539 Munich, Germany

[c] State Collection for Anthropology and Palaeoanatomy Munich, Kaulbachstraße 37, 80529 Munich, Germany

HIGHLIGHTS

- We list a set of requirements for a synchronized distributed data management.
- A discussion about existing solutions for synchronized distributed data management.
- The xBook framework has the aforementioned features and meets the requirements.
- We describe the dynamic synchronization process of xBook in more detail.
- The synchronization is realized within the archaeological application OssoBook.

ARTICLE INFO

Article history:
Received 12 March 2015
Received in revised form
25 June 2015
Accepted 1 July 2015
Available online 14 July 2015

Keywords:
Synchronization
Databases
OssoBook
xBook
Framework
GUI

ABSTRACT

In this paper, we describe xBook, a generic, open-source e-Science infrastructure for distributed, relational data management that is particularly designed for the needs of archaeological related disciplines. The key feature of xBook is that it can be used as an offline resource at remote sites during excavations and can be synchronized with a central server at any time. While some scientists can record data in xBook in the field where no internet connection is available, colleagues can already work with and analyse the previously synchronized data via the central server at any location in the world. Due to the modular implementation, any relational data schema can be implemented in xBook. This way, xBook provides a flexible data management service that is not tailored to a specific discipline or view but can be customized to any specific need to support e-Science applications. It could potentially be used in any application and offers the synchronization feature to any domain. Incarnations of the xBook framework are used in archaeology, and archaeobiology (anthropology and archaezoology). We will highlight one of them, OssoBook, an e-Science service that implements a data model for animal remains from archaeological sites (mainly bones) and has emerged as one of the European standards for archaeozoology.

1. Introduction

As in many other applications in archaeology a main part of the work comprises in collecting, sharing and analysing data. Often many researchers from different institutions and even varying countries are involved in excavation projects. Therefore entering data directly into databases is required to easily access data from different places and work simultaneously on recording as well as analysing the data. Archaeological data is often gathered in field work, i.e., at remote sites that do not offer a convenient environment for IT services, it is typically not possible to enter the data into databases that must be accessed via an internet connection. As a consequence, IT services are hardly used in these projects. Rather, data is typically recorded on paper and is (if at all) later processed electronically using proprietary and/or file-based data management tools like Excel, etc. for doing simple descriptive statistics. This is significantly inconsistent with the need to sustainably store data on the cultural heritage claimed by the UNESCO.[1]

* Corresponding author.
E-mail address: lohrer@dbs.ifi.lmu.de (J.-Y. Lohrer).

http://dx.doi.org/10.1016/j.future.2015.07.001
0167-739X/

[1] http://www.unesco.org.

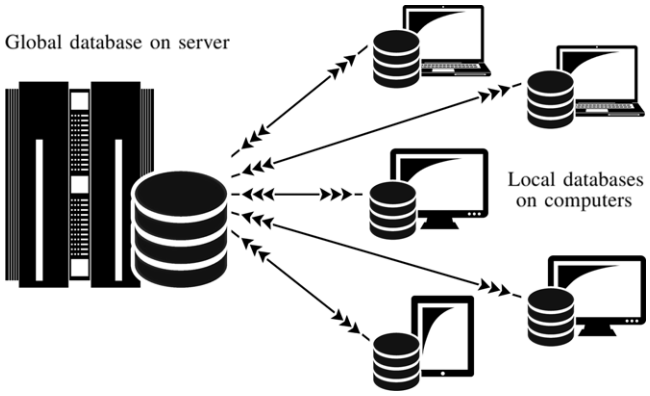

Fig. 1. The local clients are connected to the global server. The synchronization allows data exchange, so data can be recorded on the local machines, but can be backupped and shared via the server.

Obviously, researchers from these archaeological domains would significantly benefit from a profound e-Science infrastructure that supports digital recording, implements sustainable data management and storage as well as offering powerful analysis tools. The key limitation of such an IT service is the problem of multiple users that need access to data recording and data analysis even if a permanent internet connection cannot be established. A synchronization process is required, implementing a client–server architecture as visualized in Fig. 1, to ensure working offline at remote places, but also storing data globally, where it can be shared with other users is the solution.

Existing commercial solutions for this problem are typically integrated in a dedicated database management system and/or cloud service. For licence or financial reasons as well as due to privacy concerns however, not all institutions can or want to resort to these systems. Budgets for archaeological excavation projects are typically optimized in terms of logistics and man-power. Reserving a considerable part for IT infrastructure is completely unrealistic. In addition, as long as the data is not yet analysed and the results are not yet published, the participating researchers are very wary about giving their data into the hands of commercial cloud services. In this paper, we propose the framework xBook, a solution for the sketched problem that follows the architecture depicted in Fig. 1 is directly included into the application and, thus, can be used independently of the underlying database software. In particular, it can also be used with non-commercial and open-source database management systems; in fact, xBook uses a MySQL database. In addition, xBook implements a sophisticated privacy management. The solution can both be run as a server and can be installed on remote clients (e.g. laptops, computers). Thus, each institution or consortium running an excavation can implement their own e-Science service without the need to give the data to a third, potentially not trusted party. Finally, the xBook framework is independent with respect to the data model. Thus, each research institution or consortium is free to implement its specialized data model reflecting different working paradigms, different ways of recording data, etc.

The xBook framework is developed by the ArchaeoBioCenter[2] that is dedicated to the study of biodiversity, resource use, climate change, subsistence strategies, cultural change and landscape development in the emergence and development of anthropogenic ecosystems from the Stone Age to today. The objects of research are archaeological archives that are analysed macroscopically, microscopically and with the use of methods from molecular biology. These also use methods from molecular biology. The goal is the exploration of the relationship between humans and their natural environment and anthropogenic impact on the environment. Members of the ArchaeoBioCenter are associated with five different faculties at the Ludwig–Maximilian-University of Munich (Veterinary Medicine; Cultural Sciences and Archaeology; Mathematics, Computer Sciences, and Statistics; Biology; and Earth Sciences), and active cooperation between the university and the Bavarian Natural History Collections (Bavarian State Archaeological Collection; Bavarian State Office for the Preservation of Monuments; Bavarian State Collection for Palaeontology and Geology; Bavarian State Collection for Mineralogy; Bavarian State Collection for Anthropology and Palaeoanatomy; and Bavarian Academy of Sciences and Humanities).

In summary, the main contributions are as follows: We list a set of requirements that should be addressed by e-Science infrastructures for a synchronized distributed data management and data analysis in the archaeological sciences that have been extracted from comprehensive discussions with domain experts, reflecting their typical working procedures (cf. Section 2). We discuss existing solutions for synchronized distributed data management and data analysis in Section 3. We propose the xBook framework that has the aforementioned features and meets the identified requirements (cf. Section 4). In particular, we describe the synchronization process of xBook in more detail (cf. Section 5) that is independent of the data model and could potentially be used in any application domain if needed. We describe the realization of the synchronization within the application (cf. Section 6) and also sketch an incarnation of the xBook framework for archaeozoological data called OssoBook [1] which has emerged as a standard for a large European community and is used in many archaeozoological projects (cf. Section 7). Finally, with this paper, we hope to draw the attention of the e-Science community to archaeological applications that provide a rich spectrum of open issues and challenges for eScience methods and services.

2. Problem formulation

In this section, we discuss the requirements of an e-Science infrastructure for the archaeological sciences in more detail. By developing a database for archaeozoologists we have identified the following requirements of a synchronization, that have to be fulfilled to make working with the data possible.

Let us note that depending on the use case of the application, some requirements listed below are obligatory and additional requirements might exist. However, the following list has been extracted after extensive communication with domain scientists analysing their particular working procedures.

A synchronized distributed e-Science infrastructure for data management and data analysis in the archaeological sciences should address the following issues (the order of appearance is arbitrary and does not reflect priorities):

- **Distinctability of entries**: Two different entries must be distinct from each other, no matter on which local database they were created.
- **Conflict Handling**: Different users are able to work on the same data simultaneously on different local databases. The synchronization must recognize that a conflict occurred and provide options to solve it.
- **Time-delayed execution**: It cannot be guaranteed that a user of the database can always execute the synchronization process. This could be technical reasons like temporary internet disconnects, but also for logistic reasons that there is no internet connection available. It must be possible to continue working offline and synchronize the data later as soon as the computer is reconnected to an internet connection again.

[2] http://www.en.archaeobiocenter.uni-muenchen.de.