



Linked European Archives for Ethnomusicological Research

## ethnoArc – technical info sheet for archives four steps towards ethnoArc

The ethnoArc project wants to improve the accessibility and valorisation of European ethnomusicological archives. It attempts to do this by creating one common interface for distributed field collections from different sources, enabling access to resources for various application and research purposes.

The ethnoMARS search tool is designed to conduct multi-archive searches and to analyse and compare retrieved data and focuses on the specialized needs of researchers. The challenge was to define, devise and implement a system that allows the digital representation of rich and complex information originating from heterogeneously structured archives without transferring data to a common metadata format.

In ethnoArc the participating archives are *not required* to change their internal database or catalogue system. They will have to generate a second database (“ethnoArc database”) that typically mirrors an existing database. In order to make the structure intelligible for archive users, the archives must also provide English descriptions for their metadata categories.

When running, ethnoArc will help the participating ethnomusicological archives to connect among themselves (“interoperability”) and create better working/studying opportunities for researchers and other external users.

### Assumptions

This info sheet assumes that

- you work at an archive which has metadata available in some sort of digital database,
- you want to experiment with or evaluate the ethnoArc tools and
- you want to make your metadata searchable with the ethnoMARS search tool.

This paper addresses the most common situation in which an archive already possesses catalogue information in a digital format (in almost any kind of database or other electronic document). ethnoArc can also be used to build a new archive catalogue from scratch, but this procedure is not the focus of the present info sheet.

More information on the ethnoArc project and the experiences that the archives have made with implementing ethnoArc can be found on the project website ([www.ethnoArc.org](http://www.ethnoArc.org)). Please feel free to contact the members of the ethnoArc consortium if you have more questions (see contacts below).

## **1 Defining your database structure**

The first step of generating an ethnoArc database from your metadata is to specify the structure of your database. The new structure will usually mimic the structure of already existing data and therefore typically is a relatively easy task. The format is specified as a set of objects and their relations and written as an XML file: the ethnoArc Metadata Specification Scheme. Documentation of the specification format, description guidelines and example descriptions are available (see contacts below).

## **2 Setting up the database**

The ethnoArc database is based on mySQL. The mySQL database is freely available from <http://dev.mysql.com/downloads/mysql/5.0.html>. In the ethnoArc project, we are using the 5.0 version of mySQL, but we are using no 5.x specific features, so the tools are likely to work on legacy 4.x installations of mySQL.

After installing mySQL, an ethnoArc tool, the ethnoArc parser, will create an archive specific database structure from the provided XML database description.

The technical requirements for running a mySQL database are moderate. Our current server is a four year old 2.6 GHz PC with 1 GB Memory and a 130 GB hard disk. Storage requirement for the metadata of an archive with 30000 entries is about 200 MB.

## **3 Filling the database**

Importing metadata from the existing format to the ethnoArc database will most likely be the highest initial investment. Since the effort depends largely on the accessibility of the existing format and on how much the new database structure differs from that, it is difficult to give a specific estimate.

For the archives involved in ethnoArc from the start, the export formats ranged from Excel sheets to ethnoArc specific XML files. The average time for implementing an importer to the ethnoArc structure was about three working days for a programmer. For a programmer who has no prior knowledge of the ethnoArc structures, this would probably require a week.

This assumes that, in the beginning, only a simple import is required, i.e. the data from the existing database is exported and then imported to the ethnoArc database. This is usually sufficient at the beginning and, since archive data rarely requires ‘up to the minute’ updates, also often sufficient in the long run, with periodic imports every month or some other convenient frequency. Dynamic synchronization between the existing database and the ethnoArc database would probably require twice the programming effort.

## **4 Evaluating ethnoMARS and making the data public**

After importing data into the database, it can already be searched in with the ethnoMARS search engine. While this is sufficient for internal evaluation, it is not recommended for general access, since it requires direct access to the mySQL database, as well as a user account and a password for it, which should not be made available to external users. To ensure a secure access system, ethnoArc provides a query server, which acts as a proxy between mySQL and external users.

This allows archives to block the access port for the mySQL database (usually port 3306) from outside access. Users need only to be able to access the proxy (ethnoArcQueryServer) on two freely selectable ports. The proxy will then handle all SQL traffic and user data behind the firewall, inaccessible by external users.

The proxy (ethnoArcQueryServer) is a Java application, which only requires a Java runtime environment.

**Contact:**

For more information about the ethnoArc project, please visit our website at [www.ethnoArc.org](http://www.ethnoArc.org) and contact the project coordinator at the mail address found at that site.