NFDI4Objects
Research Data Infrastructure
for the Material Remains of
Human History

TRAIL 3.1:

ArboDat+: Developement of a concept for a centralised standardised ArboDat service

Partner **Lead**: Matthias Renz (Kiel University)

Members: Felix Bittmann (Lower Saxony Institute for Historical Coastal Research), Moritz Mennenga (Lower Saxony Institute for Historical Coastal Research), Wiebke Kirleis (Kiel University), Toralf Kirsten (Mittweida University of Applied Sciences), Nicole Boenke (Ruhr-Universität Bochum, Institute of Archaeological Studies) and Christoph Rinne (Kiel University)

Contact Matthias Renz (Kiel University) / mr@informatik.uni-kiel.de

Summary

The program package ArboDat is currently used for capturing archaeobotanical data and widely established in the (European) community with isolated installations. The goal of this TRAIL is to redesign ArboDat as a user friendly, extensible web-based software system in order to overcome challenges to collaboration and data sharing. ArboDat+ shifts individual, incompatible, decentralised and isolated offline ArboDat instances that are based on a commercial product (MS Access) to a (new) common web-based system. This will retain relevant functionalities and make data FAIRer to register, access, handle, analyse, retrieve, integrate and store.

ArboDat+ will be integrated into the Research Data Commons (RDC) of the Common Infrastructures (infra) section of the NFDI e.V. This multi-cloud-based infrastructure, which is being developed as part of a number of NFDI consortia (especially NFDI4BioDiversity and the NFDI/GAIA-X project FAIR Data Spaces), is intended to provide uniform access to data, software and computer resources as well as enable confident data exchange and collaborative work. This integration and its connection to PANGEA will assure long-term sustainable access to ArboDat+ services.

The main objectives of this TRAIL are: 1) Comprehensive review of objects, fields, attributes and used functionalities of the current ArboDat instances with the involvement of the ArboDat community (current and potential users) and other experts in archaeobotany. 2) Conceptual design of a common scheme, scheme mapping, metadata handling and requirement analysis for retaining functionalities. 3) Study of the targeted data model. 4) Implementation and integration of the new ArboDat+ system into the RDC. 5) Adoption of the existing connection to PANGAEA, developed at the NIhK.

This TRAIL will make a significant contribution to the harmonisation, long-term archiving and publication of archaeobotanical data. Through its integration into the RDC, it will offer an internationally available platform for flexible and efficient reuse of archaeobotanical data. The result of this TRAIL is an online software application service to register and analyse archaeobotanical data and related metadata in a standard database in order to make them freely accessible to the scientific community, especially national and international archaeobotanists and scientists in related disciplines (cultural and natural sciences).

Description

The TRAIL aims at developing an Interoperability Service (IntS), discovery service (DiS) and software application service (SAS). Existing ArboDat data sets from the applicants are used as test data to develop the tool. Once the tool is implemented, the ultimate aim is to include the data of all ArboDat users.

Currently, at the institution where ArboDat is used, it is dependent on a commercial product: MS Access. There are several versions of the software in use and compatibility issues arise with each update of MS Office. Organisation-specific adaptations complicate the exchange of data between institutions enormously – this is only possible manually.

A generic data scheme which is extensible but requires a rich metadata specification of each data element will enable FAIR use of data. The TRAIL produce this scheme based on the following standards: the ArboDat standard concerning acquisition of archaeobotanical data; an access and dissemination standard supported through the RDC with PANGAEA and T5; standards of spatial (GIS) and temporal attributes. Originally developed as part of a DFG priority programme, the current 2018 version of ArboDat (compatible with Microsoft Office 2016) is available in German, English and French and is used in around 20 European countries. This allows data to be exchanged, but does not guarantee free access to it. Since the data acquisition and storage depends on the skills and knowledge of individual users, backup is not guaranteed and the risk of losing quality data is very high. Therefore, an export interface was developed to transfer analysis datasets from local ArboDat instances to PANGAEA, where the data is open access and has a citable DOI. To develop and sustain ArboDat in the long term, it needs to be converted into a more modern database format, retaining the export, publication and other functions via PANGAEA.

Within PANGAEA, plant taxa are linked to the Euro+Med Plant Base Project so that taxonomic changes can be synchronised. The data can be used to analyse (past) biodiversities and their changes over time. ArboDat+ will not only improve networking within the archaeobotanical community, but also make data interoperable via key fields and common references to a wide range of space- and time-related data (geographical information, excavations, archaeozoology, methods, experiments, metallurgy, isotopes, dating, sedimentology/geoarchaeology and much more).

Relevance

This TRAIL integrates ArboDat into a sustainable framework to enable standardised data access, organisation, retrieval and dissemination. ArboDat already has over 150 users and working groups distributed around Europe. Arbodat+ will serve both archaeobotanists and other communities. The TRAIL relates closely to TA1: Documentation, because the metadata is recorded in ArboDat by coordinates relating to the geographical or excavation context. The relation to TA2: Collecting is that plant remains are part of collections and must be stored and described accordingly. ArboDat can help to set standards for this type of material. ArboDat+ will lay the foundations for a standardised long-term archiving and storage of archaeobotanical research data and with strong implications for TA5. The link with PANGAEA will ensure open access to and dissemination of publications of site- or period-related datasets in the future. Using macros, legacy data that was available in different formats can be integrated into ArboDat.

This TRAIL is connected to different NFDI initiatives. As archaeobotanical data contributes to (past) biodiversity and its ecological and palaeoenvironmental proxies, there is a close thematic link to NFDI4BioDiversity, underlined by the cooperation with PANGEA. The same applies to NFDI4Earth, where plant remains and related measurements (e.g. aDNA, isotopes, radiocarbon dates) are proxies for different research parameters, e.g. earth system research, climate change, water scarcity (hydrology), land use change or development, pollution, natural disasters. With regard to NFDI4Culture, plant remains are important indicators of human activities and are part of cultural heritage. Through the standardised data scheme, ArboDat+ will significantly improve the accessibility and interoperability of archaeobotanical datasets. With the interface to PANGAEA, it will also contribute to shared knowledge commons based on the FAIR principles.

Deliverables

1. RDC-compatible scheme: Central ArboDat+ DBMS scheme that consolidates existing ArboDat instances and is compatible with the RDC platform.
Interface to PANGAEA: the software tool enables the export and open access publication via PANGAEA.
2. The ArboDat data scheme is formalised and published in the Commons as a reference for archaeobotanical research. ArboDat+ could easily be integrated into existing courses and trainee programmes, contributing to a broad competence framework.

Work plan

Topics	Months											
	1	2	3	4	5	6	7	8	9	10	11	12
Kick-off meeting: introduction of deliverables and responsibilities												
Clarification of existing ArboDat instances, data and data origins and required ArboDat functionalities												
After 3 months: 1st milestone meeting Discussion and evaluation of a common ArboDat scheme and scheme mappings												
Prototype implementation (proof of concept), development of data access and online service strategies												
After 9 months: 2nd milestone meeting Members of the ArboDat community evaluate service and functionality quality												
Plan for providing services: StoS; DaS; DiS; PaS; first online version of ArboDat+												