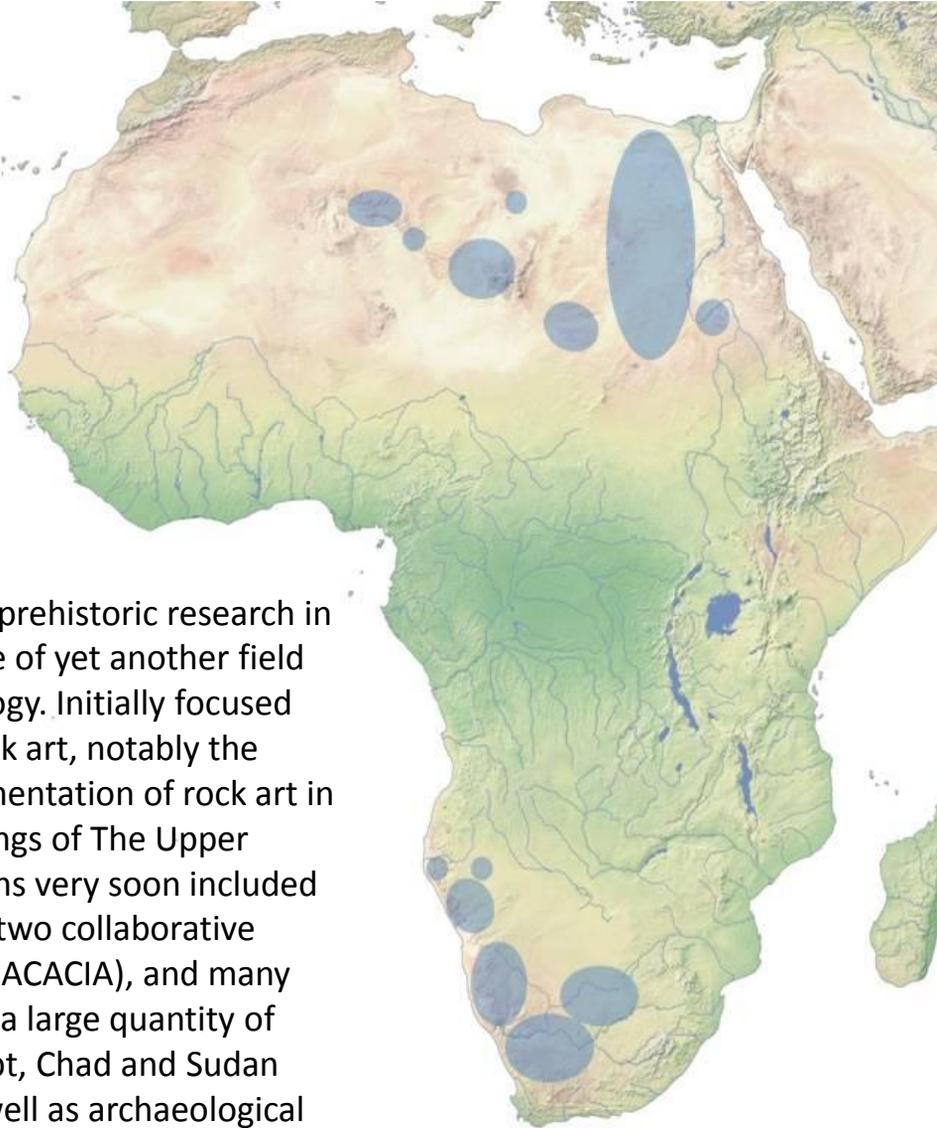


The open Archive Cologne for Archaeology and environmental history in Africa

Eymard Fäder M.A.

Ladies and gentlemen, dear colleagues, please allow me to express my gratitude for being given the opportunity to present our digital archive project within today's workshop. AAARC, another acronym, will hopefully facilitate memorizing; even more so if one doesn't find it appealing.

Over 50 years ago, in 1963, prehistoric research in Cologne saw the emergence of yet another field of study – African Archaeology. Initially focused on Libyan and Namibian rock art, notably the most comprehensive documentation of rock art in the world – The Rock Paintings of The Upper Brandberg – the field seasons very soon included excavations as well. Within two collaborative research projects (BOS, SFB ACACIA), and many other single theme studies, a large quantity of research material from Egypt, Chad and Sudan was to be accumulated as well as archaeological records of many kinds.





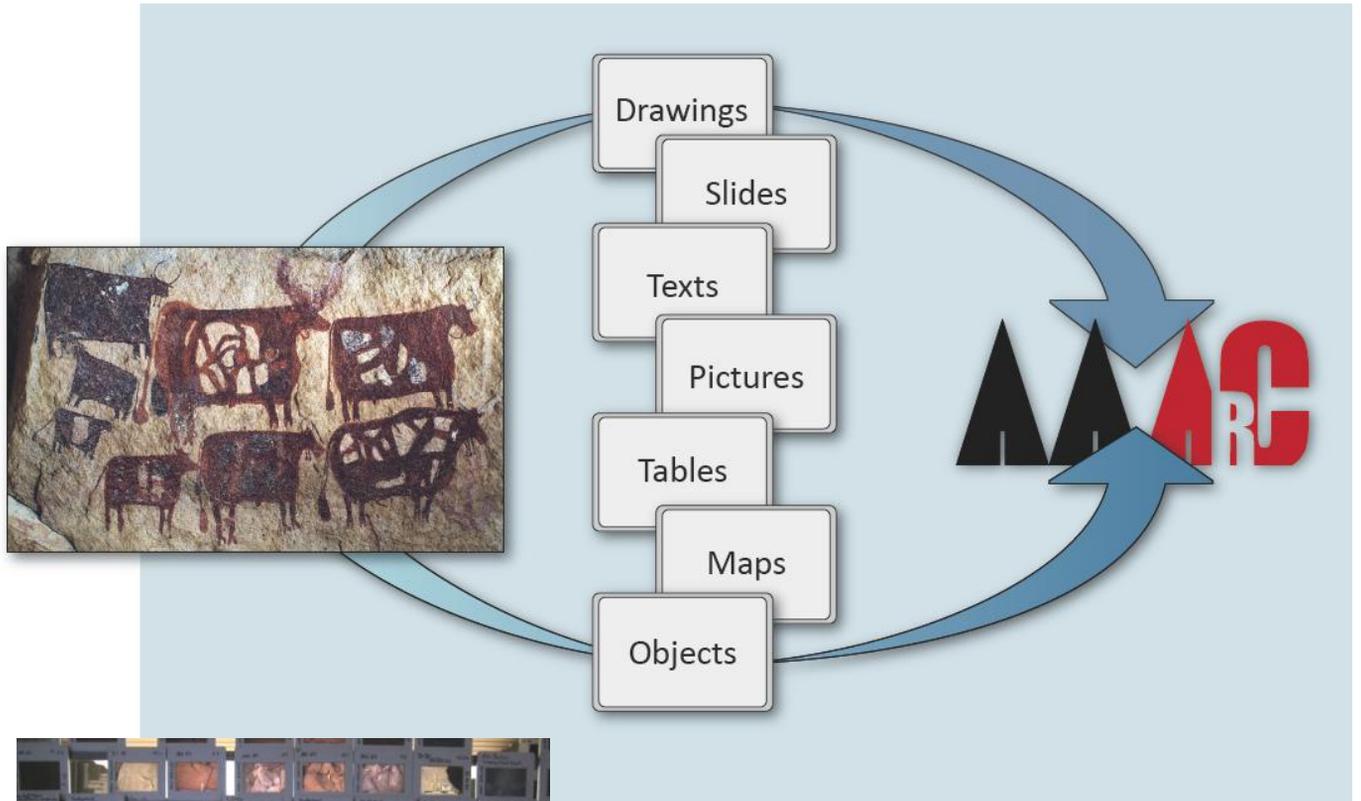
Later on, the initial focus on NE and SW Africa was revived and extended to multidisciplinary research on the cultural dynamics of arid regions. As you all are aware, the present CRC 806 studies the emergence of the modern human out of the African continent.

The documentation of the Cologne African prehistory studies is stored mainly in the institute's repositories while the results are being published mainly in HBI's *Africa Praehistorica* monograph series. Please apologize this enumeration that may appear sophomoric to guests and redundant to my colleagues, its only purpose is to illustrate both the plethora as well as the variety of archival material accumulated. An accumulation rooted in, as every student of field archaeology learns from the very first day of academic education, invasive fieldwork that cannot ever be repeated because it destroys its subject of interest. While the paradigm of open access and open data may appear dogmatic in different fields, in archaeology it must necessarily be redundant. The accumulation of archaeological research data is destructive upon collection in most cases – it cannot be regarded other than obsolete if not made accessible. The weight of responsibility even increases in regard to the countries whose cultural heritage is the object of academic research and that may not be able to afford the luxury of first world scholarly research ; they are therefore in dire need of a “restitution” in knowledge from the beneficiaries of the research activity.

In 2012, African Archaeological Archive Cologne (AAARC), in collaboration with the CoDArchLab and funded by the DFG, has stepped up to digitally record and make available this research “heritage”. I will not leave unmentioned that AAARC is part of the Arachne digital repository of the DAI and is hosted by the RRZK. This project would have not been possible without this invaluable cooperation. In accordance to the funding agent's set of standards, the principal aim of this archive is to comprehensively digitize the records and enable their accessibility by use of the interfaces provided by the housing repository, the DAI. With the project having processed a mere 60,000 digital copies, the digital presentation instruments still are in an unsatisfactory state. After just having received funding continuation for another 18 months after a mere two years of work, I can only attempt a progress report as well as sketch the technical and conceptual complexity of the task.



Archival Material





Neither do I want to overwhelm the uninitiated nor bring coal to Newcastle to the disciples of digital data management with too technical an exposition. Those of you about to get involved in a similar undertaking might find the naïve approach both refreshing and reassuring that made us start with a workshop, wanting for advice from colleagues already experienced in the then rather novel field of open digital data management in archaeology. With the funding proposal in view, the already well-structured documentation repository had to be re-assessed as to the necessary hardware and software infrastructure. Moreover, retrospectively adapting an already efficient and well-functional data management model into a digitally functional one turned out to be the complicated key aspect of the whole enterprise. It is our firm conviction that in retro-digitization, the archival structure mirrors the research goals and background and is thus an indispensable component to be preserved.

Standards

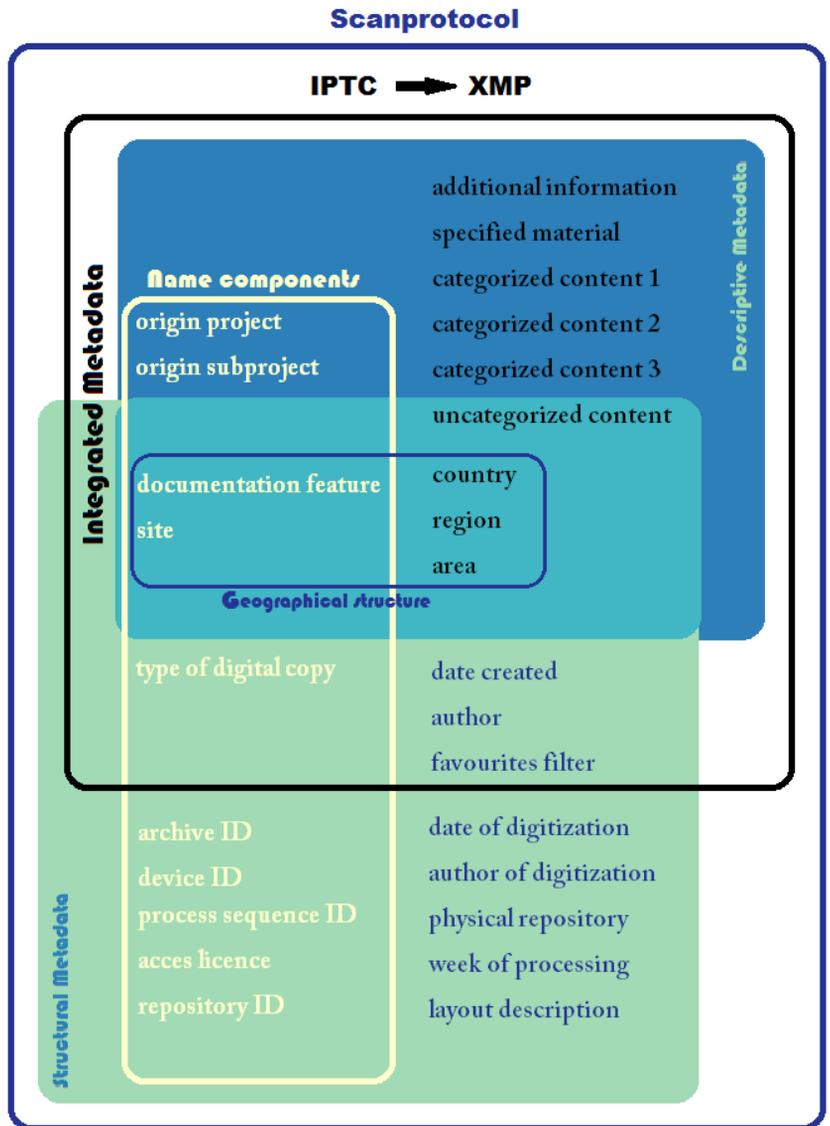
- Digitizing Standards set by the German Science Foundation (DFG):
 - Open Access computing (software and share-/freeware formats)
 - Metadata interoperability
 - Open Data accessibility

Interfaces

- ARACHNE (German Archaeological Institute (DAI) Digital Archive)
- DAI-Gazetteer and DAI Zenon

Metadata frameworks of the currently favored ontological structure (e.g. CIDOC-CRM) are available, but not only lacked in a digital interface, it would have been too costly in the given time resources to adapt that kind of digital data model into the historically developed “search instrument” of the past research repository.

The archive team instead opted for an easy transfer to the IPTC-XML picture metadata technical solution that would facilitate the mapping into further developments in the not yet standardized field of archaeological data management. Unsurprisingly for archaeological data, the main structural axis inherited with the original 1980s “site catalogue” is the geographical hierarchy, where each subset is contained in the higher level, top-down from the (actual) country of provenance to the sub-site entity like e.g. single features. The simplicity of it allows for being mirrored in the digital folder hierarchy. The second axis of reference – encoded in the name of the digital copy - projects the genesis of the documentation as well as the physical place of reposition of actual archival material.



Metadata Structure

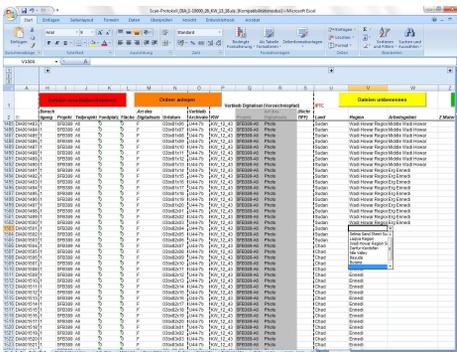


The kernel of descriptive curation, the third metadata axis consists however in content encoding of the material, in a controlled vocabulary as well as free text. Most of this metadata is carried in an integrated XML-tag that can be automatically harvested. Occurring redundancies are fully intentional.

One technical aspect is for the metadata to be mapped to a search-engine and linked to GIS- and bibliographical interfaces to enhance interlinked research opportunities.

The other technical challenge addresses the necessity to batch-process vast amounts of digital copies using an array of combined software tools, both freeware and proprietary. Developing an appropriate workflow and technical solutions appeared paramount in the light of the vast volume of documentation: Digital genuines and copies enter the curation process by record into a staff-developed datasheet (MS-Excel) that by means of macro changes the names and generates a batch-ready list for the integrated metadata. These are then simply copied and pasted, then edited by means of the appliances EXIFutils and EXIFcell into the IPTC-tags.

“Scanprotocol“



ID	Name	Date	...
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025

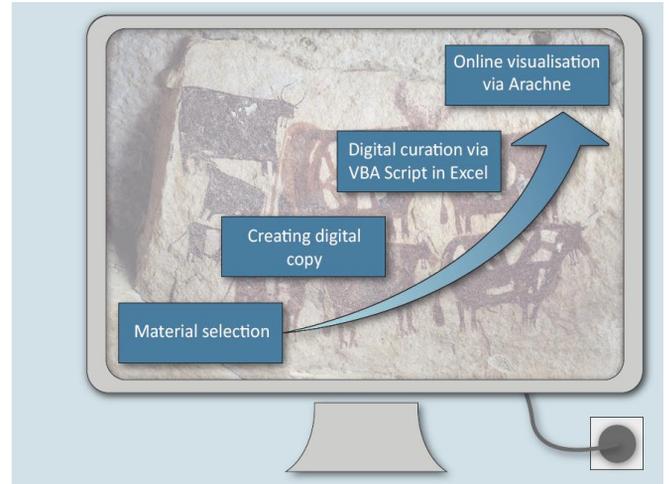
employs:

- MS Excel
- MS EXIFutils
- EXIFcell
- Adobe Photoshop
- MS Windows
- JOE - Rename
- BulkRenameUtility

- input interface
- computing:
 - digital curation
 - structural realisation
- processing reference



Adobe Photoshop then encodes these as [XMP tags](#) into the code header. Back to the MS-Excel application (named “scan-protocol”), a VB-application generates and distributes the mapping-ready copies into the meticulously maintained geographical folder hierarchy. This is the genuine archival structure and digital repository, which in turn is processed by the CoDArchLab into web-visualization and search-engine operability.



This is accomplished through a custom-developed C++ tool that employs the Adobe XMP-toolkit to readout the metadata and makes it operable by Qt-framework for different software systems. The tool generates SQL update instructions to be mapped directly to the database. By following this comparatively simple technical, but in practice complex workflow we have achieved a considerable output.

basic ARACHNE interface

basic DAI Metadata-set: AAArC example

 interface

 metadata-set



Digital copies so far

➤ Pictures

ca. **67,000** copies (ca. 3.8 TB)

of which ca. **60.000** are fully
curated and imported in
ARACHNE



➤ Documentations

ca. **9,000** copies (ca. 370 GB)



So far, the main focus laid on the management of visual data (evaluated by human interface). Making data computable by IT-interface available and not only storing it in a digital archive is the goal of future phases within the project. The site catalogue with its 6000 entries will provide the interlinked framework to GIS-visualization and bibliographical reference, all within the interoperable DAI interfaces, like iDAI-ZENON and iDAI-gazetteer. We are well aware that no reasonable run-time will allow for a complete digitization of the available physical archive. The intended accomplishment therefore aims at a deep-structured digital archive that will allow for both colleagues and interested public to access the available research data, and especially the heritage management of the countries where the data was collected to build upon for own digital repositories and education tools. The English operation language was chosen with the latter aspect in mind. Sustainability beyond project runtime includes not only the comprehensive documentation of workflow and technology, but also the development and implementation of concepts and procedures that will enable colleagues to digitize and digitally curate future research data and thus add on to the stock and continuation of AAaRC.

Finally, please allow me to draw your attention and express my gratitude to the cooperators and team members who indispensably enabled the success of this endeavour:

AAArC project team and contributors:

Tilman Lenssen-Erz, Rudolph Kuper
 Eymard Fäder, Friederike Jesse, Maya von Czerniewicz , Jan Kuper, Oliver Vogels
 Johanna Dreier, Norman Klahre, Andreea Darida, Tim Piccolini

CoDArchLab: Reinhard Förtsch, Simon Hohl, Sebastian Cuy

Institutional Infrastructure and Cooperation

- Cologne Digital Archaeology Laboratory (CoDArchLab)
- University of Cologne, Regional Computing Centre (RRZK)
- Center for eHumanities at the University of Cologne (CCeH)
- NESPOS Neanderthal-Museum Online Digital Archive)
- SARADA (South African Rock Art Digital Archive)

The Future of AArC

Not only an online repository

but

a digital research archive

Thank you for allowing me to briefly sketch the challenges of a digital research archive.

