Research Data in South Asian Studies

WORKSHOP

THURSDAY, 7TH NOVEMBER 2024

FRIDAY, 8TH NOVEMBER 2024

Thursday, 7 November 2024

12:00-12:40 **Get Together / Lunch**

12:40-12:45 **Welcome Address**
             (Rike Balzuweit)

Introduction

12:45-13:00 **Research Data in South Asian Studies**
             (Ute Hüsken)

13:00-13:15 **Research Data and the FID4SA**
             (Nicole Merkel-Hilf)

Use Cases – NFDI Text+ – Legal Aspects

13:15-13:45 **Reading the Ṛgveda – and Producing Textual Research Data**
             (Anne Keßler-Persaud)

13:45-14:15 **The Transmigration of Indic Epigraphy to the Digital Plane of Existence**
             (Dániel Balogh)

14:15-14:45 **The Vedic Prose Corpus**
             (Oliver Hellwig, Sven Sellmer)

14:45-15:15 *Coffee Break*

15:15-15:45 **Some Considerations on Research Data Management in Long-term Projects: Creating a RDM Plan for the Project „Hindu Temple Legends in South India" (HAdW)**
             (Liudmila Olalde, Frank Grieshaber)

15:45-16:15 **Legal Framework for Publishing Research Data**
             (Fabian Rack)

16:15-16:45 **Text+: An Infrastructure for Language-orientated Data within the NFDI**
             (Andreas Witt)

16:45-17:00 **Wrap-up and Discussion**

Dinner from 6:30pm (self-paying) at "Da Baggio"

Friday, 8 November 2024

09:00-09:15 **Get Together**

Use Cases – Ethics – NFDI4Culture – RDM

09:15-09:45 **The OMnibus of Mantras: Creating a Multimedia Digital Archive for Mantra Research**
(Borayin Larios, Andrea Acri)

09:45-10:15 **Digitization of Gandharan Artefacts: Research Data Perspectives**
(Jessie Pons, Frederik Elwert)

10:15-10:45 **Archiving and Sharing Ethnographic Research Data**
(Michaela Rizzolli)

10:45-11:15 *Coffee Break*

11:15-11:45 **NFDI4Culture**
(Matthias Arnold)

11:45-12:45 **Research Data Management Practice**
(Georg Schwesinger)

12:45-13:15 **Final Discussion**

13:15-14:00 *Lunch and Closing of the Workshop*

Abstracts

Reading the Ṛgveda – and Producing Textual Research Data
(*Anne Keßler-Persaud*, Göttingen)

In my project on the Ṛgveda, I will use text-linguistic and hermeneutic methods to develop new interpretations of a series of hymns from this ancient collection of texts. My interpretations will be notable for their consistent focus on the Ṛgveda itself as the context of the individual hymns. In the workshop paper, I will discuss how the digital platforms and search engines that exist for the Ṛgveda can be effectively combined with classical philological analysis in finding – ideally all – meaningful comparative text passages. I will further consider which of the data that will be generated in my research process are re-usable; two types of textual data come into question. Finally, I will address the matter of suitable formatting and publication of this research data. In addition to permanent availability, visibility is an important factor in making actual re-use a realistic possibility.

The Transmigration of Indic Epigraphy to the Digital Plane of Existence
(*Dániel Balogh*, ERC project DHARMA, HU Berlin)

The advantages provided by electronic texts incorporating descriptive markup are increasingly evident. Epigraphic sources are foundational to research in many subfields of South Asian studies, and digital editions of such sources constitute a backbone of the ERC project DHARMA. The EpiDoc markup system, a customisation of TEI adapted to epigraphy, is now a universal standard for such editions, but its application to our discipline still has its pitfalls. On the one hand, EpiDoc is a broad standard in need of further specification and customisation to suit the characteristics of the scripts, languages, physical formats and genres of Indic inscriptions. On the other hand, EpiDoc makes a point of compatibility and reverse translatability with the Leiden conventions for print-based markup in (primarily) Graeco-Roman epigraphic editions, which to some extent clips the wings of born-digital editing. The DHARMA project aims to establish standards for the encoding of Indic inscriptions that, while remaining compatible with TEI in general and EpiDoc in particular, offer at least partial solutions for these problems. In addition to discussing the whys and hows of this endeavour, this paper will highlight the challenges involved in linking editions to metadata, in structuring metadata, and in setting a desirable level of granularity in text encoding.

The Vedic Corpus Project
(*Oliver Hellwig*, Berlin / *Sven Sellmer*, Adam Mickiewicz University Poznań, Polen)

In preparation for a larger project on information retrieval in Vedic, a pre-modern South Asian language (ca. 1300–300 BCE) whose texts deal primarily with the Vedic sacrificial ritual, a group of researchers from Germany (O. Hellwig), Poland (S. Sellmer) and Japan (K. Amano) are currently compiling a corpus of linguistically annotated Vedic prose texts that will bring together in a standardized format as many surviving works as possible, their post-Vedic Sanskrit commentaries and line-by-line aligned linguistic analyses as well as translations into modern languages. In addition, we are building a database of parallels for selected Vedic rituals. On the one hand, the project aims to create a comprehensive database for the training of machine translation algorithms. On the other hand, the database of parallel passages and the translations will hopefully help researchers to find relevant information for linguistic, philological and text-historical purposes more quickly in this extensive and complex collection of texts. The first part of the presentation provides an overview of the aims of the project and its current (technical) status. In the second part, we offer detailed insights into selected parts of the data collected so far.

Some Considerations on Research Data Management in Long-term Projects: Creating a RDM Plan for the Project „Hindu Temple Legends in South India" (HAdW)
(*Liudmila Olalde*, Project Hindu Temple Legends in South India, HAdW / *Frank Grieshaber*, Project Hindu Temple Legends in South India, HAdW)

Submitting an RDM plan has become a requirement for funding applications, yet, it is equally important for ongoing projects to create one and keep it up to date. Taking the preliminary RDM plan of the long-term project HTL as a basis, we argue that the FAIR principles (Findability, Accessibility, Interoperability, and Reusability) are not only relevant for research data published as project output or archived once the research project has come to an end, but should also guide the production, exchange and storage of data throughout the duration of a project. Adhering to these principles when designing workflows can, for instance, facilitate the collaboration among project members and reduce the amount of time necessary for coordinating tasks.

Legal Framework for Publishing Research Data
(*Fabian Rack*, Research Associate at FIZ Karlsruhe - Leibniz Institute for Information Infrastructure and iRights.Law in Berlin)

The presentation will provide an overview of the legal framework for publishing and making research data available (data protection, copyright, confidentiality, etc.). The implementation of Open Science and research data licensing practices will also be discussed.

Text+: A Digital Infrastructure for Language-Oriented Data within the NFDI (not limited to Southeast Asia)
(*Andreas Witt*, NFDI Consortium Text+, University of Mannheim)

Text+ is a consortium within the German National Research Data Infrastructure (NFDI) that focuses on creating a digital infrastructure for language- and text-related data. Its services are aimed at many disciplines, including philologies, particularly with a focus on Southeast Asia. The NFDI is an initiative for the long-term preservation and indexing of research data. The primary goal of Text+ is to establish an infrastructure for digital collections, lexical resources, and editions. This promotes the use of digital resources in research and teaching to ensure the interoperability and reusability of data in accordance with the FAIR principles (Findable, Accessible, Interoperable, Reusable). Text+ offers various services and infrastructures, including data centres for the provision and archiving of research data, a registry system for describing and networking resources, the Federated Content Search (FCS) for full-text search within Text+ resources, and the GND (Integrated Authority Files) agency, which provides advice and services for standard data. The community is closely involved in the project, with regular open consultations and collaboration with professional associations and networks. Text+ supports the creation and use of research data and promotes networking and cooperation with national and international partners and standardization committees. This ensures integration into the European and international research landscape.

This presentation will showcase various services offered by Text+, particularly in the context of use in humanities research and networking with Specialised Information Services (FIDs).

The OMnibus of Mantras: Creating a Multimedia Digital Archive for Mantra Research
(*Borayin Larios*, ERC project MANTRAM, University of Vienna / *Andrea Acri,* ERC project MANTRAM, École Pratique des Hautes Études, Paris)

The MANTRAMS project explores the multifaceted dimensions of mantras as they transcend various media and cultural contexts across South Asia. Focusing on the interplay of sound, materiality, and embodiment, the project investigates how mantras function as powerful mediators between the human, divine, and cosmological realms. The presentation will delve into the project's key themes: the synaesthetic and multisensory experiences of mantras, their role in negotiating power dynamics and social hierarchies, and the integration of mantras into digital media. A central component of the project is the ambition to create an OMnibus of Mantras (OM), an interactive digital platform designed to map and document mantra traditions through the collection of new sonic, visual, and textual data. This presentation will highlight the potential synergies between MANTRAMS and Andrea Acri's ERC project "MANTRATANTRAM," particularly in terms of creating a comprehensive digital database that integrates both textual and sonic data to enhance accessibility and cross-disciplinary collaboration. We will reflect on the potential of digital sonic archives to take us on cross-cultural sound journeys and to display sounds in a way that is ethically aligned to the concepts of knowledge accessibility of the communities we work with.

Digitization of Gandharan Artefacts: Research Data Perspectives
(*Jessie Pons*, CERES, Ruhr-Universität Bochum, *Frederik Elwert,* CERES, Ruhr-Universität Bochum)

The DiGA project, short for "Digitization of Gandharan Artefacts: A Project for the Preservation and Study of the Buddhist Art of Pakistan" is a cultural heritage project that ran from 2021 to 2024, funded by the German Federal Ministry of Education and Research. It digitized and catalogued a corpus of around 1500 Buddhist sculptures from the ancient region of Gandhara. The typical funding mode for digitization projects presents inherent challenges, as the funding is for a specific time frame, but the expectation is to produce long-lived, reusable data for future research. The project addressed this challenge in two ways. First, it sought coordination and consultation with various projects and stakeholders in the field of Gandharan studies to ensure that early design decisions were in line with the needs of the community and compatible with current developments elsewhere. This resulted in the creation of a thesaurus for describing Gandharan art, based on existing reference works. Secondly, the project chose infrastructure providers that would guarantee a long-term perspective for the data created in the project, rather than building its own platform. In addition, it was important to incorporate the perspectives and needs of stakeholders in Pakistan to ensure that the process and output of the project met local expectations and ethical standards.

Archiving and Sharing Ethnographic Research Data
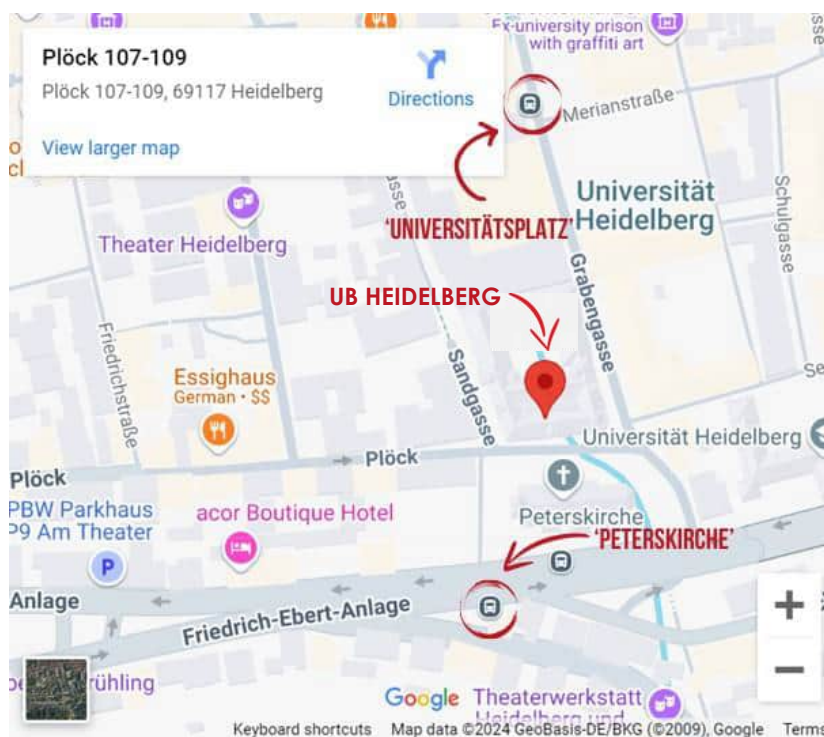(*Michaela Rizzolli*, FID SKA, Qualiservice, Universität Bremen)

How can sensitive ethnographic data be archived and made available for re-use? The lecture discusses which ethical and legal aspects need to be taken into account, presents how ethnographic data can be archived at the research data center Qualiservice and what role concepts such as the CARE Principles might play in the future.

Research Data Management Practice
(*Georg Schwesinger*, UB Heidelberg)

The presentation introduces some of the most important best practices for research data management. This includes information on data storage and archiving, selection of data formats, data migration, organization of spreadsheets and researching suitable repositories.

Getting There

Venue

Universitätsbibliothek
Handschriftenlesesaal

Plöck 107-109
69117 Heidelberg

Contact

FID Südasien

fid-suedasien@ub.uni-heidelberg.de

Tel.: (06221) 54-15047

UB Heidelberg / Handschriftenlesesaal

Tel.: (06221) 54-2392

Hauptbahnhof ⟶ Universitätsplatz

Bus 32 (Platform C) to Universitätsplatz, every 10 minutes.

Hauptbahnhof ⟶ Peterskirche

Bus 20 (Platform C) to S-Bahnhof Altstadt, every 20 minutes.