

# LINKED DATA METHODOLOGIES IN GANDHĀRAN

## BUDDHIST ART AND TEXTS

### PELAGIOS WORKING GROUP FINAL REPORT

Frederik Elwert and Jessie Pons  
Center for Religious Studies  
Ruhr University Bochum



Suggestion for citation: Elwert, F. & Pons, J. 2020. 'Linked Data Methodologies in Gandhāran Buddhist Art and Texts. Pelagios Working Group Final Report', doi:10.13154/rub.148.125.

This work is licensed under a [Creative Commons "Attribution 4.0 International"](https://creativecommons.org/licenses/by/4.0/) license.  
DOI: [10.13154/rub.148.125](https://doi.org/10.13154/rub.148.125)

This publication is financed by a grant from the Pelagios Network ("Small grants", 2019).



Pelagios  
network

RUHR  
UNIVERSITÄT  
BOCHUM

RUB

# LINKED DATA METHODOLOGIES IN GANDHĀRAN BUDDHIST ART AND TEXTS

## PELAGIOS WORKING GROUP FINAL REPORT

Frederik Elwert and Jessie Pons  
Center for Religious Studies  
Ruhr University Bochum

### 1 INTRODUCTION

The aim of the Working Group “Linked Data Methodologies in Gandhāran Buddhist Art and Texts” is to explore potential uses of Linked Open Data principles in bridging different collections of Gandhāran Buddhist resources, also in different media (i.e. text and visual art).

The team includes art historians and archaeologists, philologists and historians of Buddhism as well as experts in Digital Humanities who have decided to combine their expertise to ponder means to foster interoperability between repositories – existing or prospective – and ultimately, to advance our knowledge of Buddhism in Gandhāra. The working group is coordinated by Frederik Elwert and Jessie Pons (Ruhr Universität Bochum) and includes: Antonio Amato (Ruhr Universität Bochum), Naomi Appleton (University of Edinburgh), Stephan Baums (LMU München), Chris Clark (University of Edinburgh), Anna Filigenzi (University of Naples, l’Orientale) Luca Maria Olivieri (Ca’ Foscari University of Venice), Gregory Parker (Classical Art Research Centre, University of Oxford), Élie Roux (Buddhist Digital Resource Center), and Peter Stewart (Classical Art Research Centre, University of Oxford).

The result of the discussion is a set of guidelines that should help in planning and implementing future collection management systems. These guidelines follow a few basic principles:

1. Our main purpose is bridging diverse collections, not building a comprehensive data model for any single collection.
2. We follow a progressive enhancement approach: Instead of envisioning a perfect system, we define a series of incremental changes that can be incorporated into existing databases with minimal effort and progressively enhance interoperability.
3. Wherever possible, we build on existing and established work, both digital and analog.

One of the challenges is to identify which parts of the required vocabularies are specific to Gandhāran studies and which should be more broadly relevant to the field of Buddhist

Studies. Where similar initiatives already exist, these should be used, extended or linked. In this respect, we are coordinating our efforts with other projects centered on equivalent or complementary collections and with a shared interest LOD. Their leaders or members have joined the working group. These projects are:

**Buddhist Digital Resource Center** The BDRC was established in 1999 as the Tibetan Buddhist Resource Center. This non-profit organisation which brings together partners in the field of preservation and research has since been dedicated to the preservation and the study of Buddhist literature as well as to the dissemination of knowledge thereof. Since 2015, it has expanded the ambit of its mission – originally limited to Tibetan texts – to include all Buddhist literary tradition. Its corpus now contains works in Pāli, Sanskrit and Chinese. The BDRC maintains the **BUDA** platform (launched in 2017) which provides **LOD vocabularies** and datasets (Roux 2018). The vocabularies primarily cover the categories of *topics* (i.e. works classified according to topics ranging from logic to deities), *places*, and *persons*. The BUDA platform strives to extend the range of vocabularies to meet the needs of its partners.

**Gandhari.org** Gandhari.org is a website for the presentation and study of the complete corpus of published Gāndhārī texts, established in 2002 and run by Stefan Baums and Andrew Glass. The corpus comprises Buddhist manuscripts, inscriptions, coins and secular documents. The first two of these sub-corpora are predominantly from Buddhist contexts and hence relevant to our purposes. They employ metadata important for the development of our categories (see below). The Gandhari.org server is in progress of being updated to a new software system, using the open-source Research Environment for Ancient Documents (or **READ**) software as a new backend for managing its source texts. This upgrade will enable many new features such as the linked presentation of images and multiple readings for each inscribed object.

**Jātaka Stories** The Jātaka Stories project was launched in 2018 by Dr Naomi Appleton and Dr Chris Clark at the University of Edinburgh. The project set itself to establish a searchable database of previous births of the Buddha which, since late 2019, is accessible online and free. It is composed of two units – stories in texts and stories in art – and integrates several fields which are pertinent to our endeavour (see below). At the moment, the corresponding corpora focus on literary and visual traditions from India but will incorporate material from other parts of Asia at a later stage. One planned extension is the inclusion of previous births preserved in the literary and visual records from Gandhāra.

## 2 VOCABULARIES

At this point, the group does not aim at defining an all encompassing linked open data ontology as data model. Its core aim is defining a set of vocabularies and creating stable identifiers

for each entry. Where the complexity of the data exceeds the features of vocabulary systems like SKOS, slightly more complex data models might be incorporated.

At this point, we focus on four core vocabularies: Places, motifs, narratives, and persons. For each, we look for best practice models that teach us *how* to model the information. Then, we define a very limited set of primary sources that give us information about *what* we want to model. In most cases, our base vocabulary will be a LOD representation of a single source or a very limited selection of sources. We also identify additional sources that we might use in order to broaden the scope of the vocabularies and increase coverage.

## 2.1 PLACES

### GENERAL SCOPE

**Nature of sites covered** The primary focus of the Working Group are archaeological sites. Under this, we understand:

- sites where Gandhāran artefacts have been excavated under official excavation campaigns, and
- sites which have been identified during the course of official archaeological surveys.

These prerequisites are crucial for issues of reliability, particularly relevant for the provenance of artefacts and the sensitive problem of their authenticity.

Although we are primarily interested in tangible traces (i.e. archaeological sites), linkage to other vocabularies which focus on historical entities is intended. These are for instance monasteries in the sense of an (im-)material institution that could, say, be referred to in a Buddhist text. This historical entity might overlap with the archaeological site (Baums 2019, 167–74). Our aim is therefore to ensure the compatibility of databases and the access to mutually complementary information. In this respect, we will work in close collaboration with researchers developing the [Gandhari.org](http://Gandhari.org) and the [BUDA](http://BUDA) platform, which document place-names in Gāndhārī inscriptions and manuscripts and more generally in Buddhist literature.

**Geographical zone covered** Scholars are sometimes at odds over the limits of the geographical zone referred to as Gandhāra or Greater Gandhāra (Pons 2019; Salomon 2018). While all agree that Gandhāra designates a geographical zone with a shared political and cultural history, philologists, art historians, archaeologists might diverge on what defines this sameness. We opt for a pragmatic and bottom-up approach by which archaeology and artefacts define contours. According to the present state of archaeological research, the western boundary can be placed at the confluences of the Panjshir, Kabul and Logar rivers southeast of the Hindu Kush, the northern boundary along the Jambil tributary of the Swat River, the eastern boundary along the Siran river and the south-eastern limit, by the mount Kohat.

**Period covered** The chronological scope of the Gazetteer is difficult to delimit categorically in light of issues connected to the uncertain date of Gandhāran Buddhist sculptures. The heyday of the production is traditionally attributed to the rule of the Kushans (c. from the 1st to the 4th centuries CE) but evidence shows that the chronological span within which Buddhist artefacts were made can be extended to the Indo-Scythian (1st century BCE-1st century CE) and the early or Turk Sahi periods (7th to 8th century CE) (on the issue of chronology and a reassessment of the conventional taxonomy of Gandhāran art see Filigenzi 2010, 389–406, 2015; Pons 2017).

## ROLE MODELS

There are a couple of initiatives that model places (sometimes among other things) and that can serve as examples for our encoding of the gazetteer.

**Linked Places** Linked Places is both a data interchange format specification and an abstract data model for historical gazetteers. It was developed as a shared format for use in both Pelagios and the World Historical Gazetteer projects (see also [this post](#) for elaboration). It particularly stands out in its ability to date not just place records themselves, but many of their properties, like names, geometries or relations to other places.

**Linked Art** Linked Art is an upcoming specification to model information about pieces of art in a LOD way. It heavily builds on CIDOC-CRM. It includes a quite elaborate **Place** type. However, in its framework, Places are primarily conceptualized as “extents in space,” not as historical entities. Also, the Linked Art recommends to use external gazetteers as reference systems.

**BDRC/BUDA** BDRC/BUDA integrate an extensive **Place** class which invites to broaden the scope within which Places are normally conceived in databases. The classification system is based on a wider range of place types (Natural features, Regions, Sites, Contemporary Place, Printeries) and a distinction between Buddhist traditions. Particularly inspiring are properties provided for each site, which exceed what other initiatives envision, such as associated works, associated persons and literary production.

## PRIMARY SOURCE

**Pons & Olivieri, Census of Archaeological Sites** This inventory (Pons 2019, 27–36) includes archaeological sites which have yielded Buddhist material. It results from a methodical review of published excavation reports and archaeological surveys undertaken by official institutions in Pakistan and Afghanistan since the beginning of research in the region in the 19th century until today. It records a total of 260 sites across the region presented in a tabular form. The list of sites follows a two-fold classification which integrates information on:

1. Physical geography. This includes: river valleys (e.g. Swat, Kabul, etc.), their stretches and banks as well as their tributaries.
2. Political and administrative divisions. This includes: the country (Pakistan and Afghanistan); the modern district (e.g. Buner, Mardan, etc.) and well as the modern city in the vicinity of the sites (i.e. zone).

This information can feed the attribute element Relation(s) for each site (see model used by [Linked Places](#)) for which we follow a two-tier hierarchical system that takes into account both physical geography and administrative geography.

However extensive, this inventory is not exhaustive and lacks relevant information. To develop into an archaeological Gazetteer and comply with standard requirements, the census must be complemented on seven points:

1. The number of sites: The census needs to be updated with the latest results from archaeological excavations and surveys undertaken and/or published during the last year.
2. Name variants: The census uses a specific format for the transliteration of toponyms in Pashto (i.e. hyphenated form and a non-capitalised second element). Other variants are also possible and need to be integrated.
3. Geolocation: The census only gives general indications about the position of the sites in relation to physical features (i.e. river system). One task will be to implement the geolocation for each site.
4. Description: The census provides no information about the sites it lists other than their name and their localisation. A description of the nature of the remains uncovered, albeit kept to a minimum, needs to be added.
5. History of archaeological research: This includes the date of the mission, the auspices as well as the person in charge of the archaeological mission.
6. Bibliography: This includes mentions of the site in surveys, reports of archaeological campaigns, studies of the site.
7. Collection: When applicable, the record should also indicate the collection in which the excavated artefacts are housed.

## **ADDITIONAL SOURCES**

In order to cover these insufficiencies, we can rely on the following additional sources:

**Pons (2011)** This PhD dedicated to the stylistic study of Buddhist sculptures from Gandhāra compiles a corpus 157 archaeological sites preserved in a database conceived under *FileMakerPro*. The information pertaining to the major features of the site, the chronology of the archaeological research, as well as the corresponding bibliography has been collected for one-third of the entries.

**Ball (2019)** This updated edition of the *Archaeological Gazetteer of Afghanistan* published by W. Ball in 1982 – since become a reference work for archaeology in Afghanistan – is a compendium of large scope with 2292 entries. This includes sites either reported in notes of travellers or excavated, and covers a time span ranging from the Lower Palaeolithic to the Timurid period. Not all entries are relevant to our purpose. It however incorporates the newest discoveries of Buddhist remains located east of the Pamghan mounts and the Hindu Kush in the Logar Province (e.g. Kharwar, Mes Aynak), traditionally understood as part of the Gandhāran cultural zone. For each site, the Gazetteer gives information on the name, the location, the date, field-work, as well as a short description. It also indicates the collection in which the finds are kept and provides an exhaustive bibliography. Each site is attributed with a catalogue Gazetteer number which will be retained to create stable identifiers for the Place items.

**Archaeological Map of the Swat Valley Project (AMSV)** AMSV project (2000-2006) was conducted within the framework of the IsIAO archaeological mission in the Swat Valley and led by Luca Maria Olivieri and Massimo Vidale. Most of the archaeological sites and rock carvings listed in the preliminary report (Olivieri and Vidale 2006, 73–150) have already been integrated in Pons and Olivieri 2019. The bibliography that the article provides for each entry however needs to be included and the census needs to be updated with the newest results of the project.

**Archaeological Gazetteer of Gandhāra** The Archaeological Gazetteer of Gandhāra is an initiative of Antonio Amato and Fabrizio Sammarco to create an online database of archaeological sites ranging from the Achaemenid period until the Arab conquest. Currently, it records over a hundred Buddhist sites for which the following information has already been entered: site name, a brief description of site, georeferentiation (location, city, region/province, state, GPS coordinates, satellite map), chronology and period, bibliography and photographic documentation.

## TASKS

Since the information relevant to an archaeological gazetteer is extensive and difficult to capture using a spreadsheet or a similar application, and since no off-the-shelf tools are available for the creation of gazetteers, we are currently looking into building a simple web application that can be used for collecting the information.

A [very simple version](#) of the gazetteer is available in the linked places TSV format and could be imported into a future application.

## 2.2 MOTIFS

### GENERAL SCOPE

“Motifs” is a rather broad category that relates to typical and recurring elements in visual art. For art historical research, identifying and characterizing motifs allows retrieving certain depictions (e.g. searching for the motif of a thunderbolt to retrieve representations of Indra and/or Vajrapāṇi and the narratives in which they play a role) and to analyse usage or occurrence patterns (e.g. some motifs may be endemic to a single site or found across several sites). It is important to note that although the term motif can be applied to both iconographic and formal features (i.e. treatment of proportions, measurements, volumes), the set of vocabulary that we define only pertains to the former aspect.

### ROLE MODELS

**Iconclass** Iconclass is a standard tool to index, describe and search the subjects of visual documents used by museum and art institutions worldwide. Its hierarchical classification system comprises 450 basic categories which deal with general topics (e.g. 1 Religion and Magic), more specific ones (e.g. 7. The Bible) as well as abstract art (0 Abstract, Non-representational Art). Additional subdivisions within these three main divisions allows attaining a greater level of specificity and precision in the description of subjects (e.g. 1. Religion and Magic includes a subdivision 12. Non-Christian religions which itself includes a subdivision 12H. Hinduism, Buddhism and Jainism). Each annotation is accompanied by a textual definition in several languages and ascribed a URI to ensure LOD availability.

**Getty Vocabularies** The *Art & Architecture Thesaurus* and *The Getty Iconography Authority* are two of the six authoritative vocabularies developed by the Getty Research Institute for cataloguing and researching objects of art, architecture, material culture and archival material. While the AAT is a thesaurus of generic terms (e.g. “stupa”), the IA includes proper names of historical or fictional characters and events (e.g. “reclining Buddha”). Therefore, the AAT broadly aligns with our “Motifs” category and the IA with both our “Narratives” and “Persons” categories. Both Getty Vocabularies follow a conceptual framework of facets (e.g. the records stupa and reclining Buddha belong to the facets “Objects” and “Allegory/Symbolism/Theme” respectively) and are polyhierarchical. This means that a single record can have several relationships which can be hierarchical, based on equivalence, or associative. Each record is identified with a URI to secure long-term consistency and linkage (under Open Data Commons Attribution License (ODC-By 1.0) to other resources such as Iconclass.

However useful and established these tools are, they are principally oriented towards Western art. Despite increasing efforts to encompass non-Western topics, their treatment of motifs typical of Buddhist, let alone Gandhāran art is limited. The IA for instance, includes less than

ten events in the Buddha's life. While these resources cannot be used as primary reference vocabularies, certain categories will be mapped with our discipline-specific vocabulary when relevant. Fortunately, such a vocabulary does exist for the field of Gandhāran studies: The *Repertorio Terminologico per la Schedatura delle Sculture dell'Arte Gandharica* (Faccenna and Filigenzi, Anna 2007, hereafter *Repertorio*).

## PRIMARY SOURCE

**Repertorio** The structure of the *Repertorio* is straightforward. It covers glossaries related to material and technique, architecture and subject. It includes a wide range of subjects: decorative motifs, people, fauna, flora, weapons, musical instruments, ceremonial objects, everyday objects, furniture, and means of transport. Furthermore, it uses a numerical identifier system and it is already available in English, Italian, and Chinese. A substantial part has also already been translated in French. However, this vocabulary is currently only available in print form and as a PDF document.

## TASKS

A first task would be to transform the *Repertorio* into a LOD-compatible vocabulary, preferably in SKOS. Secondly, its elements should be mapped to existing art historical vocabularies to increase compatibility with existing (Western) collections. A [preliminary draft](#) of such a mapping has been created as part of the DiGA project.

The question of translation (and transcription) of terms employed to describe Buddhist architecture and subject matters poses some challenges. Currently, the *Repertorio* adopts a hybrid system (e.g. the building is designated by the Indic term but its features by terms borrowed from Western architecture). SKOS allows assigning multiple labels in different languages to a concept, which mitigates the issue of choosing between Indic terms and translations. For the SKOS version, a subsequent task will be to enrich the *Repertorio* with Indic terminology to avoid any semantic dichotomy between the terms used and the objects they describe. With respect to architectural motifs, we will notably draw on the work by O. von Hinüber on the architectural terminology used in Buddhist sources to describe stupa architecture (von Hinüber 2016). For the mappings to (Western) vocabularies, careful consideration is required when deciding if a concept traditionally applied to Western art and architecture is actually the same as (i.e., an `exactMatch` in SKOS terminology) a concept describing Gandhāran art and architecture.

## 2.3 NARRATIVES

### GENERAL SCOPE

In a Buddhist context, narratives can designate stories of the Buddha's lifetime, stories of his previous lives (the *jātakas*) as well as stories related to important Buddhist figures either in the past (*avadānas*, on issues of terminology see Lenz 2010, 92–93) or in the present. These narratives are preserved in a variety of media: written, visual, oral, theatrical, choreographic, musical. Documents allow tracing several versions of the same story. These generally share the same storyline – the narrative nucleus – but can differ in detail – the peripheral narrative units – thereby constituting multiple (re)tellings. Comparing versions of a narrative within and across media allows identifying overlaps as well as discrepancies which bespeak convergent or divergent threads in the dissemination of narratives and can point to media specific conventions and logics in their transmission.

Two types of Buddhist narratives are at the center of our current efforts. These are:

1. The life story of Siddhārtha Gautama. This includes events from the four surveys in the Tuṣita Heaven to the Mahāparinirvāṇa and the worship of the Buddha's relics.
2. Stories about previous births of the Buddha, the *jātakas*.

A third type of scene will be added to these, namely scenes showing hunters, revelry scenes, figures worshipping the Buddha's bowl and the like. Some of these depiction are generic in nature and although they do not count as narratives per se, they should nevertheless be considered in our effort. Other scenes however, might refer to stories which have until now defied any identification attempt.

In the study of visual art, identifying a narrative depicted on a relief or a painting can be a challenge in its own right. Usually, individual elements like figures or certain objects (i.e. motifs and persons) can be recognized. While these can often feed hypotheses about the story represented, they are sometimes not sufficient to settle for a one or the other story. Thus, an annotation schema for narratives would ideally allow tagging individual elements and the relationship between these as well as to capture hypotheses about the narratives in a structured way.

A model that is very promising in this regard is currently being developed in the [ONAMA](#) project. We will follow the project's output and evaluate its use for the study of Gandhāran Buddhism.

### ROLE MODELS

The vocabularies for motifs and narratives show considerable overlap. The Iconclass and Getty AAT/Iconographic Authority vocabularies discussed [above](#) are also relevant here.

**ONAMA** ONAMA develops a model for a cross-media description of actions, actors, settings and temporal structures. In addition to the constituting basic elements of transmedial “narrative nuclei”, in these descriptions the respective textual and pictorial realizations are recorded in detail ([project’s description](#)). ONAMA’s preliminary model has been [presented](#) at the DHd2020 conference. It approaches narratives as (potentially sequenced) scenes which include key actors, objects, and actions. This allows labeling not only a narrative as a whole (e.g. a certain *jātaka*), but also the elements that it comprises.

## PRIMARY SOURCES

**Jātaka Stories** In its current form, the Jātaka Stories database records a total of 749 stories across eight textual collections in Indic languages as well as 62 depictions from sites in South Asia. A particularly valuable contribution of this project is the list of story clusters it developed. A story cluster is a label designating a group of stories “that share a very similar plot architecture, with key storyline events in essentially the same sequence involving essentially the same character types.” To give a concrete example, the Sivi-jātaka from the *Jātakatthavaṇṇanā* is part of two story clusters, namely “King Śibi/Sivi gives away parts of his body” and “A man makes a gift of his eye”. Other versions (e.g. the Śibi-jātaka from Āryaśūra’s *Jātakamālā* or the Śibi from the *Avadānaśataka*) are also both grouped under this label although the king gives different parts of his body, namely his leg and his eyes respectively. Being generic in nature, the label constructs types regardless – but not oblivious – of the distinctive features of each version. These story clusters respond to two concerns: Firstly, to avoid misrepresentations about the multi-directionality in the transmission of narratives which could result from applying a specific title (e.g. the Pāli Sivi-jātaka) to all variants of the same story; secondly, to facilitate the mapping of narratives within and across text and image corpora. What is more, by facilitating the comparison of images with potential literary parallels, the Story Cluster constitutes a tool to ascertain unidentified scenes.

**Faccenna (1962)** Faccenna’s opus *Sculptures from the sacred area of Butkara I, Swat, Pakistan* records over 700 sculptures coming from the same area, namely the Swat Valley. Most of these have been excavated at Butkara I during scientific excavations while others come from Saidu Sharif I as well as various collections (e.g. Wali Sahab collection). This is the most extensive archaeological repertory of sculptures, with entries written by Maurizio Taddei. It is equipped with a detailed index and covers over thirty Buddhist narratives (mostly events from the life story of Siddhārtha Gautama) well as a very wide array of generic scenes. The provenance of the majority of artefacts is known which makes for a reliable corpus from which to extract a list of Buddhist narratives.

**Zwalf (1996)** Zwalf’s reference book *A Catalogue of Gandhara Sculptures in the British Museum* documents 670 sculptures. Over two hundred of these illustrate the life story of Siddhārtha

Gautama and *jātakas* altogether covering about forty events. Most of these sculptures were discovered during the course of excavations conducted by Officers of the British Government of India and the Archaeological Survey of India in the 19th and early 20th century.

## ADDITIONAL SOURCES

**Pons (2011)** The substantial number of stories covered by Faccenna and Zwalf can be augmented with the narratives identified by Jessie Pons within the framework of her PhD. The descriptors which she uses for her database are adopted from A. Foucher's seminal work *L'art gréco-bouddhique du Gandhāra: étude sur les origines de l'influence classique dans l'art bouddhique de l'Inde et de l'Extrême-Orient* (Foucher 1905–1951) and have already been translated in English. This additional list of descriptors needs to be assessed by members of the working group and agreed upon.

**Jongeward et al. (forthcoming)** The forthcoming monograph *Buddhist Rebirth Narratives in Literary and Visual Cultures of Gandhāra* gathers the results of the interdisciplinary project on the religious and cultural significance of literary and visual narratives of the past lives of the Buddha during the early centuries CE in Gandhāra. It is led by Jason Neelis and financed through a grant from the Robert H.N. Ho Family Foundation Program in Buddhist Studies. The monograph contains a comprehensive catalogue of previous birth stories documented in Gāndhārī manuscripts and Gandhāran art (reliefs and paintings). This corpus can be fed into the list of narratives and story clusters from the Jātaka Stories project.

## TASKS

As a first step, an agreeable list of *jātakas* and life-time stories should be compiled and each entry should get a permanent ID. At a later stage, a more complex modelling of narratives using the ONAMA ontology can be carried out.

## 2.4 PERSONS

### GENERAL SCOPE

The category “Persons” was not included in the original programme of the working group but emerged as a natural extension of its efforts during our discussions. Three types of persons have been distinguished and will be taken into consideration:

- Historical persons that appear as donors, owners or comparable functions, e.g. in inscriptions
- Literary persons mentioned in Buddhist texts or represented in Buddhist art. This includes human (e.g. Ānanda, Jaya) as well as supra-human beings (e.g. the Buddha, the Bodhisattva Maitreya)

- Archetypes or classes of persons. This category is probably the broadest in scope as it includes:
  - professions or roles (musician, brahmin, prince)
  - generic (i.e. unnamed) mythological figures (nāga, yakṣa)
  - animals when they are actors in a narrative

The boundary between the former two types can sometimes be blurred, as donors might also appear in scenes alongside the literary personae. This is only hypothetical in figurative art since there is no way to identify the donor portrayed but occurs in literary tradition. Several *avadānas* in Gāndhārī cast Aśpavarman, a member of the Apraca lineage (BL1.175-184, *avadāna* 7 and BL1.185-204, *avadāna* 8). The boundary between the last two types can also be blurred with the “Motifs” category, as typical depictions of persons and animals can also be regarded as motifs (e.g. depictions of rows of buddhas and donors).

## ROLE MODELS

There are multiple attempts to bring prosopographical data into the linked data world. Some initiatives that are potential role models for our data include:

**LAWD** LAWD is an ontology for Linked Ancient World Data. Despite its generic name, its main focus lies on texts and persons. It has basic classes for Persons and Names, including Deities.

**SNAP:DRGN** Standards for Networking Ancient Prosopographies, builds on LAWD and extends the ontology with classes for personal relations. SNAP is interesting because it specifically positions its ontology as a means to import data from multiple partner datasets to merge them into a single authoritative file. Thus, SNAP comprises an ontology, an interchange format, and a dataset.

**GND** GND is the German Integrated Authority File. Since it supersedes the older Name Authority File, it includes many records on persons. It also defines an RDF ontology and exports its datasets in RDF under a CC0 license.

## PRIMARY SOURCES

Several sources can be used in the compilation of the vocabulary associated with the person category, without any of them supplying the bulk of terms. In general, each feeds one of the three types distinguished. All of these sources have been mentioned already, and we will content ourselves with highlighting the way in which they respectively contribute to the person vocabulary.

**Jātaka Stories** The Jātaka Stories database includes a field “Characters” which lists over 1500 terms. These are characters who occur in either the “story body” (the story of the past itself), the frame story (the story of the present) or the rebirth identification. This field is two-tiered and distinguishes “Character Name” and “Character Descriptors”. These correspond to our “Literary person” and “Archetypes” respectively. The Jātaka Stories project covers a substantial number of terms but since its focus – at least with regard to the “Persons” category – is on literary tradition, it lacks some terminology more specific to the visual tradition, in particular Gandhāran. This inadequacy can be compensated by the next source.

**Pons (2011)** The list of subjects represented in the art of Gandhāra established by Pons includes:

1. Deities more typical of the Gandhāran regional context such as figures depicted under Western guises (e.g. the Dioscuri, Tyche, Dionysus) or Indic gods and semi-gods (e.g. Pāñcika, Hāriti, Vajrapāṇi, Sūrya).
2. Figures who populate the visual repertoire of Gandhāra (e.g. Atlantes, donors, devotees, wreath-bearers).

These terms remain to be translated in English and approved by the working group. The “Persons” category must be completed with the individual names of protagonists in episodes of the Buddha’s life. Some of these names are mentioned in the frame story or rebirth identifications of jātakas and have been documented by the Jātaka Stories project but a more systematic compilation still awaits.

## ADDITIONAL SOURCES

**Buddhist Studies Person Authority Databases** DILA created an authority database of persons containing a large part of the persons/characters referenced in the Chinese Canon. It also provides an LOD export using the BUDA ontology.

**Gandhari.org** Gandhari.org manages a list of names recorded by gandhari.org. The database already contains 112 names of historical figures mentioned in Gāndhārī inscriptions. They are presented as the donors of an object (e.g. an image of the Buddha or a reliquary), its owner, the beneficiaries of the donation, or the artisan. In the future, this list should be augmented with the names of protagonists of stories preserved in Gāndhārī manuscripts, particularly *pūrvayogas* and *avadānas*. These stories only occasionally find a correspondence in other Indic literature and would provide a list of hitherto unknown names.

**BDRC** BDRC contains historical persons mentioned in colophons, roles, as well as deities (though with a focus on Tibet).

**British Museum** The collection of the British Museum uses controlled vocabularies to catalogue its collection. Its vocabulary of person names includes Buddhist figures and deities.

## TASKS

A first task is to complete the lists corresponding to the three main categories outlined above. As a second step, the entries will have to be mapped with Iconclass, Getty IA and BUDA, although one can only expect limited matches with these established vocabularies.

## 3 INTEROPERABILITY AND LINKING

Interoperability and linking are primary concerns of the working group, even the main reason for this initiative. This involves two levels:

1. Establishing shared identifiers for places, persons, motifs and narratives will enable the cross-linking of resources from different partners *within* the field of Gandhāran and Buddhist studies.
2. Through the cross-linking of our vocabularies with much larger, established vocabularies, our work can be integrated with data sources *beyond* this particular field, especially with Western art history.

For this to succeed, two steps must be taken:

1. When creating the vocabularies themselves, we will include means of linking to established vocabularies and populate them with references to the most prominent ones.
2. We encourage our partners to include the permanent identifiers into their own databases.

We know that many projects are on low budgets and have limited technical capacities. So re-implementing their databases in an LOD-friendly form from the ground up is usually not an option. Thus, we recommend a series of progressive enhancements that projects might take:

1. We suggest the use of permanent identifiers for the shared vocabularies as a first, relatively lightweight step towards more interoperable infrastructures. For instance, projects that already have place as data type in their system might add a column for the place ID in their database table.
2. We encourage projects to also create permanent identifiers for items in their databases, e.g. texts or objects.
3. Where possible, a machine-readable export of individual entries should be implemented in any format that suits the project (e.g. XML, JSON).
4. A dedicated URI should list all entries that are available in machine-readable form to allow for harvesting of the data.

For the permanent identifiers, we chose not to use a base domain that we own, because we are not in a position to make institutional guarantees for the longevity of any domain of ours. Thus, we will choose a general-purpose provider like [PURL](#) or [w3id](#).

## 4 VISION OF A FUTURE ARCHITECTURE/PLATFORM

To achieve cross-linking, we would need a system that allows for searches and analyses that span collections. This would require three additional layers that are currently outside the scope of this working group:

1. An actual LOD ontology that defines not only vocabularies, but different classes and their relations.
2. Export facilities within each partner's systems to map their collections to the shared ontology (potentially through intermediate formats like LIDO).
3. A system that harvests and ingests the individual data files and provides a common search and analysis infrastructure.

We acknowledge that this is a huge task. But we also believe that this is an important goal and vision for a future of Gandhāran studies where the availability of digital collections will actually lead to new insights. In pursuing this goal, we will follow a model of progressive enhancement, making small steps that each provide benefits of their own, but that are guided by a broader vision for a future research infrastructure.

### ROLE MODELS

There are a couple of projects that follow a similar approach of integrating data from diverse sources through shared interchange of formats and ontologies. \* [BUDA](#), mentioned above, develops into a platform that ingests data from different Buddhist studies projects and makes them available under a shared data model. This makes BUDA not only a role model, but also a potential candidate for a shared platform. The main challenge is catering to the specific needs of art history within the framework. \* [Europeana](#) is a large portal that aims at making visible the European cultural heritage by providing a platform for browsing and searching numerous collections from museums, libraries, archives and galleries. Due to the breadth of data types, its schema is rather basic. \* [Nomisma.org](#) defines an [ontology](#) for numismatic research. Based on this, it incorporates numerous [datasets](#). Besides a simple search, it also provides research tools that allow more complex analyses of the collected data. \* [SNAP:DRGN](#) follows a similar approach for prosopographical data (though the platform seems to be dormant). It builds on the [LAWD ontology](#), extends this with classes for relations between persons, and incorporates three partner datasets into a consolidated dataset, providing stable URIs for each record. \* [World Historical Gazetteer](#) is a project that aims at building a similar platform for integrating historical gazetteers. It uses the linked places and linked traces ontologies developed in cooperation with the Pelagios network.

## 5 ACKNOWLEDGMENTS

We would like to thank the members of the Working Group for their enthusiastic participation in the discussion around Linked Data Methodologies applied to Gandhāran Buddhist Art and Texts and for their invaluable comments on this report. Any errors are the sole responsibility of the authors.

## BIBLIOGRAPHY

- Ball, Warwick. 2019. *Archaeological Gazetteer of Afghanistan: Revised Edition*. Second. Oxford: Oxford University Press.
- Baums, Stefan. 2019. “A Survey of Place-Names in Gāndhārī Inscriptions and a New Oil Lamp from Malakand.” In *The Geography of Gandhāran Art Proceedings of the Second International Workshop of the Gandhāra Connections Project, University of Oxford, 22nd-23rd March, 2018*, edited by Wannaporn Rienjang and Peter Stewart, 167–74. Oxford: Archaeopress.
- Faccenna, Domenico. 1962. *Sculptures from the Sacred Area of Butkara I, Swat, Pakistan*. Vols. II, 2-3. IsMEO Reports and Memoirs. Rome: Istituto poligrafico dello Stato, Libreria dello Stato.
- Faccenna, Domenico, and Filigenzi, Anna. 2007. *Repertorio terminologico per la schedatura delle sculture dell’arte gandharica – Sulla base dei materiali provenienti dagli scavi della Missione Archeologica Italiana dell’IsIAO nello Swat, Pakistan*. Rome: IsIAO.
- Filigenzi, Anna. 2010. “Post-Gandharan/Non-Gandhara: An Archaeological Inquiry into a Still Nameless Period.” In *Coins, Art and Chronology II. The First Millennium C.E. In the Indo-Iranian Borderlands*, edited by Michael Alam, Deborah Klimburg-Salter, and Soshin Kuwayama, 389–406. Vienna: Verlag der Österreichischen Akademie der Wissenschaften.
- . 2015. *Art and Landscape. Buddhist Rock Sculptures of Late Antique Swat/Uḍḍiyāna. With Contributions by Luca Maria Olivieri and a Note by Peter Rockwell*. Vienna: Verlag der Österreichischen Akademie der Wissenschaften.
- Foucher, Alfred. 1905–1951. *L’art Gréco-Bouddhique Du Gandhāra: Étude Sur Les Origines de L’influence Classique Dans L’art Bouddhique de L’Inde et de L’Extrême-Orient*. Ernest Leroux. Paris: Imprimerie Nationale.
- Hinüber, Oskar von. 2016. “Some Remarks on Technical Terms of Stūpa Architecture.” *Annual Report of the International Research Institute for Advanced Buddhology at Soka University* 19: 29–46.
- Jongeward, David, Timothy Lenz, Jason Neelis, and Jessie Pons. Forthcoming. *Buddhist Rebirth Narratives in Literary and Visual Cultures of Gandhāra*. Seattle: University of Washington Press.
- Lenz, Timothy. 2010. *Gandhāran Avadānas, British Library Kharoṣṭhī Fragments 1-3 and 21 and Supplementary Fragments A-C, (Gandhāran Buddhist Texts, 6)*. Seattle: University of

Washington Press.

- Olivieri, Luca Maria, and Massimo Vidale. 2006. "Archaeology and Settlement History in a Test Area of the Swat Valley Preliminary Report on the AMSV Project (1st Phase)." *East and West* 56 (1-3): 73–150.
- Pons, Jessie. 2011. *Inventaire et Étude Systématiques Des Sites et Des Sculptures Bouddhiques Du Gandhāra: Ateliers, Centres de Productions*. Unpublished PhD thesis.
- . 2017. "Archaeology in Gandhāra: A Review of Research at the Crossroads of Disciplines." In *The Diversity of Classical Archaeology*, edited by Achim Lichtenberger and Rubina Raja, 199–219. *Studies in Classical Archaeology* 1. Turnhout, Belgium: Brepols Publishers.
- . 2019. "Gandhāran Art(s): Methodologies and Preliminary Results of a Stylistic Analysis." In *The Geography of Gandhāran Art Proceedings of the Second International Workshop of the Gandhāra Connections Project, University of Oxford, 22nd-23rd March, 2018*, edited by Wannaporn Rienjang and Peter Stewart, 3–40.
- Roux, Elie. 2018. "The BUDA Platform. LOD Platform and Model for Buddhist Studies: Bibliography, Prosopography, Geography." In Mainz. <https://doi.org/10.17613/kqzr-hp97>.
- Salomon, Richard. 2018. *The Buddhist Literature of Ancient Gandhāra. An Introduction with Selected Translations*. Summerville: Wisdom.
- Zwalf, Wladimir. 1996. *A Catalogue of Gandhara Sculptures in the British Museum*. 2 volumes vols. London: British Museum.