Towards an Ontology for Toponyms in Nepalese Historical Documents

Sabine Tittel
Heidelberg Academy of Sciences and Humanities
Heidelberg
sabine.tittel@hadw-bw.de

Abstract

Nepalese historical legal documents contain a plethora of valuable information on the history of what is today Nepal. An empirical study based on such documents enables a deep understanding of religion and ritual, legal practice, rulership, and many other aspects of the society through time. The aim of the research project ‘Documents on the History of Religion and Law of Pre-modern Nepal’ is to make accessible a text corpus with 18th to 20th century documents both through cataloging and digital text editions, building a database called Documenta Nepalica. However, the lack of interoperability with other resources hampers its seamless integration into broader research contexts. To address this problem, we target the modeling of the Documenta Nepalica as Linked Data. This paper presents one module of this larger endeavour: It describes a proof of concept for an ontology for Nepalese toponyms that provides the means to classify toponyms attested in the documents and to model their entanglement with other toponyms, persons, events, and time. The ontology integrates and extends standard ontologies and increases interoperability through aligning the ontology individuals to the respective entries of geographic authority files such as GeoNames. Also, we establish a mapping of the individuals to DBpedia entities.

Keywords: Ontology, Nepal, Place Names, Toponymy, Linked Data, Text Edition

1. Introduction

Recent years have witnessed a growing estimation of modeling language data as resources of the Semantic Web. Linguistic resources in increasing numbers are converted into Linked Data (LD), a set of standard practices for representing and interlinking structured data on the web using Resource Description Framework (RDF). Ontologies of numerous domains provide the necessary structures to formalize the information in the resources and to embed the resources in a cross-discipline, cross-domain and cross-linguistic context. We here describe a proof of concept for the creation of an ontology suitable for the modeling of Nepalese place names and their entanglement with other place names, events, persons, and dates. The ontology, called NEPAL.PLACES, will be part of a bigger vision of LinkedOpenNepal: to build a set of LD data resources with a text corpus as its focal point. The text corpus is part of the Documenta Nepalica and comprises texts and documents on the history of religion and law of pre-modern Nepal. LinkedOpenNepal will further comprise an ontological model also for person names, and a lexicographic module. NEPAL.PLACES is the outcome of a collaboration of domain experts of South Asian studies and of ontology engineering. The Nepalese language is the official language of Nepal. In a country with 92 (to 124) different languages (and language varieties, resp.) [in 2001] (Dr. wasa et al., 2007 10), it is the mother tongue of approx. 45% of the population of >29MM people [in 2021] serving also as a lingua franca (Hutt (1988, 23); van Driem (2001) 1125–1128; 1130f.; 1142)). It is also spoken in north-eastern India, Myanmar, and Bhutan (Riccardi, 2003 539–541). Through different historical periods, Nepali has developed a large body of literary works (Hutt (1988 71–76); van Driem (2001 1136f.)), catalyzed also by the legal code from 1854 (the Muluki Ain, see Khatiwoda et al., 2021)) written entirely in Nepali (Riccardi, 2003 544). Nepali literature is clearly under-represented in the digital context, in particular with respect to historical language stages. The aim of NEPAL.PLACES (⊂ LinkedOpenNepal) is, thus, to not only increase the visibility and re-usability of the valuable historical Nepali documents but also to establish interoperability with other language resources. This is particularly relevant in the context of South Asian countries and societies that are historically connected to and share geographic, cultural, economic, philological and linguistic aspects with Nepalese society. The challenges of this task lie in extending existing ontological models, in ambiguities within the data to be modeled, linguistic hurdles, complex relations between toponyms and connected information as witnessed by the documents.

The paper is structured as follows: We describe the linguistic resource that provides the data in section 2 and introduce the paradigm of Linked Data, together with related work, in section 3. Section 4 shows the steps...
from the linguistic resource towards a Semantic Web resource, with a motivation in section 4.1 the digital status quo in section 4.2 the development of the ontology in section 4.3 and discussions guiding the ontology engineering process as well as ongoing work in section 4.4 We close with a conclusion in section 5.

2. The Linguistic Resource ‘Documenta Nepalica’

The research project ‘Documents on the History of Religion and Law of Pre-modern Nepal’ (Heidelberg Academy of Sciences and Humanities, with research units in Heidelberg, Germany, and Patan, Nepal) makes accessible for the first time a corpus of historical texts from the early Sāha (1769-1846) and Rāṇā (1846-1951) periods. This rich textual material is held by the National Archives of Nepal and other archives and collections. It consists of temple documents and administrative and legal documents, and it essentially lays the ground of our knowledge about topics still to a large extent unexplored: the history of religious institutions in Nepal, of legal practice in South Asia, the developments entailed by the formation of the Himalayan state, such as the restructuring of social institutions, elite cultures, the legitimization and affirmation of rulership, and the expansion of Hindu rule.

The project’s research results accessible through the Documenta Nepalica consist of a comprehensive catalogue of descriptive metadata of the documents (>65,000 entries) and of scholarly digital text editions (>450) including English translations, comments, notes, and facsimiles (published in open access and partly already with DOIs, in cooperation with Heidelberg University Library). These resources are provided through a freely accessible virtual register, with a motivation in section 4.1. They are currently not open to the public, the bibliography and a glossary growing into a future dictionary. For an example, see Fig. 1.

The creation of an accompanying, comprehensive bibliography and a glossary growing into a future dictionary is also underway. A newly added task is the identification of named entities within the corpus (person and place names with significant spelling variation) and their organization within a register. The digital research results are groundbreaking in their contribution to text-based empirical studies on the history of Nepal. However, they constitute a typical data silo (Berners-Lee, 2009), i.e., a valuable resource in the World Wide Web (WWW) but with limited access to the catalogue and text editions, not allowing queries beyond the implemented possibilities; the glossary and named entity register are currently not open to the public, the bibliography only partly. Thus, interoperability with related resources is not given. To address this problem, we adopt the paradigm of Linked Data.

3. Linked Data

Over the last years, the paradigm of LD (LOD respectively, with the ‘O’ symbolising open access) has developed into a widespread, powerful means to turn the heterogeneous, unstructured information of the WWW into machine-readable, semantically accessible data (Berners-Lee, 2009; Bizer et al., 2009). As the most solid grounding of the Semantic Web it provides a set of best practices for the interlinking of datasets, with Resource Description Framework (RDF, Cyganiak et al., 2014) as a standard data model (Wood et al., 2014) 4-12. RDF represents data in the form of graphs, i.e., triples with a source node (a subject), and a target node (an object), connected through a directed edge (predicate) pointing from the former to the latter; subject, object and predicate are each identified through URIs that must be accessible via HTTP (the object can also be a literal described as a string).

There are many advantages to representing linguistic resources as LD, such as structural interoperability (through same format and same query language), conceptual interoperability (through shared vocabularies and ontologies), uniform access (through the use of standard Web protocols), and resource integration (linking resources) and federation (cross-resource access); cf. Chiarcos et al. (2013).

3.1. Related Work

Given the potential of LOD to create interoperability crossing borders of disciplines, content, languages, original formats and places of publication of resources, the last decade has witnessed a significant growth in contributions to the LOD ecosystem. This is also reflected by the strong increase of the LOD cloud (Gandon et al., 2017) with resources from geography, life sciences, (political) administration, social media, etc., showing semantically interconnected information across all fields. Also in linguistics, many proposals have been made to integrate language resources and their lexical, semantic, and morpho-syntactic aspects into the Semantic Web (Declerck et al., 2015; Aguado de Cea et al., 2016; Jarrar et al., 2019; Bellandi et al., 2018), for the recent state of the art see Bosque-Guil et al. (2018) and Cimiano et al. (2020).

Also for the modeling of text editions as LD, we can build on previous work. Approaches to transform text editions from XML/TEI (TEI Consortium, 2017) into RDF have been proposed, e.g., a mapping of TEI markup to CIDOC CRM (Crofts et al., 2003) in Eide (2014 2015, 23-38), a formalization of the TEI data model into an ontology (Ciotti and Tomasi, 2016 2017), and the integration of RDFa (Herman et al., 2015) into XML/TEI (Titel et al., 2018), adapted by Cimiano et al. (2020, 253-262). All approaches depend on an interpretation of each individual TEI markup; however, efforts to define a binding mapping of TEI and OntoLex

---

Cf. http://narchives.gov.np/
Cf. the project’s website http://www.hadw-bw.de/Nepal
Cf. http://narchives.gov.np/
Cf. http://lod-cloud.net
 Effective 03-18-2022, cf. https://nepalica.hadw-bw.de/nepal/
Figure 1: A text edition as part of the Documenta Nepalica (annotated named entities are highlighted in green).
1 The
2 <seg about="occ:219" type="n.">
3 <w property="rdfs:seeAlso" resource="gloss:1931">cakrapuñja</w>
4 <gloss property="skos:definition">endowed lands or
5 and charitable functions</gloss>
6 <seg about="occ:213" type="n.">ropani</seg> 
7 [...]
8 <gloss property="skos:definition">unit of land
9 measurement in the hill region, including the
10 Kathmandu Valley, comprising four muris</gloss>
11 <seg about="occ:235" type="n.">gut.hi</seg>
12 <gloss property="skos:definition">endowed lands or
13 and charitable functions</gloss>
14 <seg about="occ:2051" type="n.">ropani</seg>
15 <gloss property="skos:definition">unit of land
16 measurement in the hill region, including the
17 Kathmandu Valley, comprising four muris</gloss>
18 <seg about="occ:2051" type="n.">gut.hi</seg>
19 <gloss property="skos:definition">unit of land
20 measurement in the hill region, including the
21 Kathmandu Valley, comprising four muris</gloss>
22 <seg about="occ:176" type="n.">Deupathap</seg>
23 <gloss property="skos:definition">unit of land
24 measurement in the hill region, including the
25 Kathmandu Valley, comprising four muris</gloss>
26 <seg about="occ:176" type="n.">Deupathap</seg>
27 <gloss property="skos:definition">unit of land
28 measurement in the hill region, including the
29 Kathmandu Valley, comprising four muris</gloss>
30 <seg about="occ:235" type="n.">gut.hi</seg>
31 <gloss property="skos:definition">unit of land
32 measurement in the hill region, including the
33 Kathmandu Valley, comprising four muris</gloss>
34 <seg about="occ:219" type="n.">cakrapuñja</seg>
35 <gloss property="skos:definition">endowed lands or
36 and charitable functions</gloss>
37 <seg about="occ:213" type="n.">ropani</seg> 
38 [...]
39 <gloss property="skos:definition">unit of land
40 measurement in the hill region, including the
41 Kathmandu Valley, comprising four muris</gloss>
42 <seg about="occ:2051" type="n.">gut.hi</seg>
43 <gloss property="skos:definition">unit of land
44 measurement in the hill region, including the
45 Kathmandu Valley, comprising four muris</gloss>
46 <seg about="occ:176" type="n.">Deupathap</seg>
47 <gloss property="skos:definition">unit of land
48 measurement in the hill region, including the
49 Kathmandu Valley, comprising four muris</gloss>
48 <seg about="occ:176" type="n.">Deupathap</seg>
49 <gloss property="skos:definition">unit of land
50 measurement in the hill region, including the
51 Kathmandu Valley, comprising four muris</gloss>
52 <seg about="occ:219" type="n.">cakrapuñja</seg>
53 <gloss property="skos:definition">endowed lands or
54 and charitable functions</gloss>
55 <seg about="occ:213" type="n.">ropani</seg> 
56 [...]
57 <gloss property="skos:definition">unit of land
58 measurement in the hill region, including the
59 Kathmandu Valley, comprising four muris</gloss>
59 <seg about="occ:235" type="n.">gut.hi</seg>
60 <gloss property="skos:definition">unit of land
61 measurement in the hill region, including the
62 Kathmandu Valley, comprising four muris</gloss>
63 Figure 2: Text edition of K.0440.0007.

Listing 1: Data sample from a text edition of the Documenta Nepalica with XML+RDFa.

For RDF to be Linked Data it must adhere to principles defined by Berners-Lee (2009), one of which reads: “When someone looks up a URI, provide useful information, using the standards” (called a dereferenceable URI). The reference to the Deopatan entry (normalized spelling) of the named entities register in l. 28–31 does not meet this condition. To enhance the information provided and to reach compliance with other resources, a link to the respective entry in a gazetteer or an authority file in the domain of geographic references can be added. These are, e.g., the Getty Thesaurus of Geographical Names (TGN), GeoNames, the Virtual International Authority File (VIAF), and the Gemeinsame Normdatei (GND). The city of Kathmandu, for example, can be connected to GeoNames for <.../nepal/ontologies/viewitem/> (Nota bene: Both resources will be made open access in the course of 2022).

Cf. https://www.getty.edu/research/tools/vocabularies/tgn/ and https://www.geonames.org/ and http://viaf.org/. 15

4.1. Why an Ontology for Toponyms

The texts of the Documenta Nepalica attest complex connections of norms, ideas, and rules to places, practices, persons, castes, and the material world. This makes them “such exciting material” (Cubelic et al., 2018) and a key component for understanding the history of Nepal. However, they have not yet been sufficiently studied, neither as a self-sustained textual category, nor as source material for a historiography of South Asia, nor in relation to other texts, such as inscriptions, straíric texts, chronicles, belles lettres, etc. Their content is varied: festivals and rituals of religious communities, emerging scribal and administrative elites, court proceedings and litigation, the development of bureaucratic policies, offices, social roles, military positions, and state duties, cp. ib. 11–13. They also give insight into the social and cultural circumstances and forms of slavery in Nepal (and South Asia) differing significantly from better-studied African and North American forms, being more familiar and strongly related to land and landownership.

http://www.dbn.de/gnd. 16

Attested in numerous documents, e.g., in doc. RRC_0062_0180. https://nepalica.hadw-bw.de/nepal/editions/show/47509.
(Bajracharya and Michaels, 2022, 1f.). All aspects are entangled with places, and all require a careful regional contextualisation to enable a deeper understanding of Nepalese socio-historical development, especially considering trans-regional migration patterns (e.g., of scribal groups) and fluctuating borders with emerging or perishing kingdoms, cp. Khatiwoda et al. (2021 47–56).

Therefore, as a major step towards a LOD transformation of data from the Documenta Nepalic, we develop an ontology that will provide the necessary classes and properties for a modeling of this entanglement. We will populate the ontology with individuals taken from the Documenta Nepalic, turning it into a means to create a bird’s eye view on the interlacing of Nepalese places with events and personalia through time.

### 4.2. Digital Status Quo

While the format of the text editions is in XML conform with the standard TEI P5 (published online as HTML) and can thus be utilized for a future automated transformation into RDF statements, the entries of the named entities register are formatted using HTML:

```
<doc id="19806">
  <body>
    <d:p>
      capital of Nepal, declared capital by...
    </d:p>
    <d:p>
      a district in Central Nepal, part of...
    </d:p>
    <d:p>
      a district in Central Nepal, part of...
    </d:p>
    <d:p>
      Bagmati Province</a>
  </body>
</doc>
```

Listing 2: HTML entry for Kathmandu.

The code example shows three aspects that are perfectly suited for the current data use but that are shortcomings with respect to automated data processing: (i) all entries of the register are annotated as undifferentiated hyperlinks: in the example, the person name Pṛthvīnārāyaṇa Śāhā, a king, and the toponym Bagmati Province; (ii) the information relevant for the classification of the entry is an unstructured string, in this case, the fact that Kathmandu is both the capital of Nepal and a district; (iii) the founding date of Nepal’s capital is not annotated. For the time being, we resolve this by extracting the information manually.

### 4.3. Ontology Development

We construct a conceptual model through the Protégé v5.2.0 desktop version 17, which we formalize in OWL (Bechhofer et al., 2004), integrating existing standards for cross-resource compatibility. As for metadata, this is straightforward: We use DublinCore properties (creator, title, issued, description, rights) 18 and VANN (vann:preferredNamespacePrefix, vann:preferredNamespaceUri) 19, the license is defined as Creative Commons Zero. We also use RDF, RDFS (Brickley and Guha, 2014), and the datatype xsd:dateTime (Fallside and Walmsley, 2004). However, to represent the information in focus, this does not suffice. Two aspects are crucial: the establishment of the classes and properties that will enable (i) a classification of the individuals of the ontology (e.g.: city of Kathmandu) and (ii) a modeling of their entanglement with other individuals, i.e., toponyms, events, dates, agents, etc. (e.g.: Deopatan is a historic town in the Kathmandu valley, now part of the city of Kathmandu). For the first point we are fortunate to be able to adopt the ontology established by DBpedia, starting with Place as the hierarchically highest class and including only classes relevant for our data 20.

The second aspect is more time-consuming. It is work in progress: The incremental population with entries from Documenta Nepalic as ontology individuals leads to repeated revisions of the created properties (and, to a lesser extent, of classes). At this point, the properties are as generic as possible, with specifications focusing on rivers (with object properties hasTributary, isTributaryOf, flowsIntoRiverAt). The property connectsTo with Place as its domain and Person as its range facilitates the connection to the person name ontology NEPALPEOPLE, see infra. For alternative designations of a toponym we use the GeoNames property alternateName 21.

We give an example for the river related properties with the modeling of the Trisuli river, serialized with Turtle Syntax (Prud’hommeaux and Carothers, 2014):

```
### https://example.org/nepalplaces#

rdfs:label "Trisuli river" ;
:hasTributary :Marsyangdi_River ;
:isLocatedIn :Nepal ;
:rdfs:comment "One of the major tributaries of the Narayani River basin in central Nepal; originates in Tibet as a stream and enters Nepal at Gyirong Town; joins the Narayani River at Devghat."
;rdfs:seeAlso <https://nepalica.hadw-bw.de/nepal/ontologies/viewitem/213> ;
owl:sameAs dbr:Trishuli_River .
```

Listing 3: Individual Trisuli river.

Currently, 57 individuals, 76 classes (used 39 times), seven object properties (used 47 times), and 15 annotation properties are registered in NEPALPLACES.

### 4.4. Ongoing Work

The ontology development is a joint activity with domain experts from the research project’s team resulting...
in and profiting from numerous fruitful discussions, exemplified in the following.

### 4.4.1. Identification of Shortcomings of the Class Structure

NEPALPLACES must provide the means to map Nepalese toponyms to a class. As mentioned above, we can draw on the DBpedia ontology. However, the DBpedia ontology proved to be (i) not fine-grained enough and (ii) not accurate for the modeling of Nepalese reality and—for that matter—of the reality of the Indian subcontinent. Thus, we extend it, integrating new sub-classes into the hierarchy. This is a process driven by several aspects:

**Population with Individuals.** The population of NEPALPLACES with data from the named entities register from Documenta Nepalica uncovers shortcomings of the class structure in a straightforward way. E.g., the aforementioned Pashupati Aryaghat is a ghat and needs to be classified as such in the ontology. Since this is not possible with the DBpedia ontology we introduce Ghat as a sub-class of ArchitecturalStructure (on the level of Building), annotated with the comment “A set of steps leading down to a body of water with a platform for bathing typically situated at their base”, cf. Fig. 3. We said already that Pashupati Aryaghat is also used as a cremation ground with high social significance and thus, we also introduce CremationGround to the ontology, where the pre-existing Cemetery “burial ground” is not accurate. So far, with currently 57 individuals, twelve classes have been added [effective 03-12-2022].

![Figure 3: Newly added class Ghat (Protégé).](image)

**Concepts behind Classes.** For the classification of each individual, we closely look at a pre-existing class that seems suitable at first sight. E.g., for the modeling of the toponym Bhadrakāli, an open air shrine of the goddess Bhadrakāli east of Tundikhel in Kathmandu, we turn to the class Shrine, in the DBpedia ontology labeled as Engl. “shrine” and placed within the hierarchy < ReligiousBuilding < Building.

A closer look at the concept behind Shrine reveals a problem: Engl. shrine [lt. SCRINIUM “case or chest for books or papers”, (von Wartburg, since 1922, 11,337b)] is attested since ca. 1000 (Ælfric of Eynsham) and seems strongly connected to the Christian Occident, typically suggesting a container with physical remains of a religiously venerated person, cf. the senses registered by OED SHRINE n. with sense n¹ “A box, coffer […]”, n² “The box […] in which the relics of a saint are preserved […]”, etc., and sense n⁵a defines a more generic concept that, however, still refers to an architectural structure: “A place where worship is offered or devotions are paid to a saint or deity; a temple, church”. It becomes clear that the class Shrine is not adequate for a mapping of Nepalese shrines that are often open-air places of worship. Thus, we include a generic concept represented by ReligiousSite that is not related to ArchitecturalStructure, accompanied by the comment “A natural site where religious rituals and activities, e.g., prayer and sacrifice, are carried out.”

**Class and Property Labeling.** When defining the classes and properties, our aim is to include labels in English, German, and Nepali (alongside English comments). There, we experience a noticeable impact on the ontology engineering by the translation process of the labeling, unveiling shortcomings within the class structure adopted from DBpedia with respect to Nepalese reality. E.g., there is no adequate Nepali term for a monastery in a generic sense: the Nepalese language differentiates bahala / vihara, a Buddhist monastery, from matia, a monastery of Hindu ascetics, and from gīrlaṭha, a Christian monastery. This results in adding three sub-classes to the pre-existing Monastery: MonasteryChristian, MonasteryBuddhist and MonasteryHindu, cf. the individual Bīnchē Bāhālā as an example:

```xml
Listing 4: The Buddhist monastery Bīnchē Bāhālā.

Note the linking (through rdfs:seeAlso, l. 8-10) to the respective entry in the Nepal Heritage Documentation Project (NHDP; Heidelberg Centre for Transcultural Studies, Heidelberg University).
```

22 Cf. http://mappings.dbpedia.org/server/ontology/classes/Shrine
23 Oxford English Dictionary – OED Online. March 2022. Oxford University Press, www.oed.com
24 This term being rarely used and replaced by Nep. carca designating the church.
The examples described reveal a Western-centric cultural bias within the encyclopedic resource of DBpedia. This results in an information imbalance in the direction of reality of non-western cultures such as the Nepalese. This flaw can be counteracted, e.g., by ontological work such as NEPALPLACES, and we hope that our efforts foster more endeavors in this regard.

4.4.2. Ambiguity within the Entries of the Documenta Nepalica

The entries of the named entities register show ambiguities that need to be resolved while integrating them into the ontology. See, e.g., Fig. 4 where the entry misses the distinction between the historical district of Deukhuri and the valley of Deukhuri. The entry for the place name of Dang is yet more complex: “a valley located in the Inner Tarai in midwest Nepal (north of Deukhuri valley) that was prior to the conquest by the Gorkha troops in 1786 divided among small principalities belonging to the Bāse Rāja: Dang and Chilli in the valley itself, and Salyan and Phalabang in the hills […]” a district in Western Nepal, part of the Lumbini Province; until 1961 together with Deukhuri one of the 32 districts formed in the Rana period¹⁹, https://nepalica.hadw-bw.de/nepal/ontologies/viewitem/389. The entry describes both the Dang valley and the Dang district that, in turn, must be differentiated into a modern administrative unit and a historic administrative unit of the Rana period. Thus, for the content of these entries, different instances must be introduced to the ontology (e.g., Deukhuri_District and Deukhuri_Valley for the entry Deukhuri), explicitly defined as different individuals. This has so far been done manually. For future automatic processing, ways of an automatized information identification and entity recognition must be evaluated.

4.4.3. Geographical Names Entangled with Offices and People

The Documenta Nepalica include a great amount of information on administrative, religious and social offices and roles together with persons holding the office or role during a particular time and at a particular place. This empirical information is very valuable for a nuanced interpretation of responsibilities, dependencies, interacting, and structural developments within Nepalese society from a diachronic and a spatial perspective. To model the manifold relations of a place presents a challenge to the development of the ontology. The following gives an example of such relations: The Degutalejyū Temple located at the Hanumāṇḍhokā palace in Kathmandu is the place of activity of a particular nagarcī, a drummer who is a musician playing an important social and religious-ritual role. This role is connected to land endowments for financing the religious and charitable functions of the nagarcī, as witnessed in different documents. One such document is a royal deed from King Gīrvāṇayuddha in Vikram Samvat 1864 (= CE 1807). The deed grants plots of land (with rice fields, gardens, lots, etc.) to a person called Bandhuvā Damā (Bandhuvā Nagārcī, resp.) for conducting rituals (providing the necessary material, e.g., buffalo, goats, cloth, etc.), for the upkeep of two large religious banners and three long trumpets offered to Degutalejyū, and, also, for military services²⁰.

Since we want to keep NEPALPLACES focused on toponymy and not populate it with aspects related to anthroponymy and prosopography, we establish a connection of NEPALPLACES to a sister ontology called NEPALPEOPLE. This sister ontology models persons within the Documenta Nepalica, including their proper names, life events and genealogical aspects, professions, societal roles, and more. In NEPALPEOPLE, Bandhuvā Damā and all other persons with the same position (of nagarcī) are modeled as individuals connected (via a specific NEPALPEOPLE property hasPosition) to a position defined through the class Nagarcī; this position is, in turn, enriched with temporal information and, through a property hasPlace, with the individual ‘Degutalejyū Temple in Kathmandu’ of NEPALPLACES²¹.

4.5. Integration of chronological aspects

The ontology provides classes representing historical concepts: HistoricalPlace, HistoricalAreaOfAuthority, HistoricalSettlement, and HistoricalDistrict, the latter classifying, e.g., former districts of the Rana period until 1961 such as said Deukhuri and Dang districts. However, these classes do not inform us about a precise date and, thus, the dimension of time needs to be integrated: When was, e.g., a district created, and when did the existence come to an end or merge into another administrative area? E.g., Tanahun is a modern district in Western Nepal and part of the Gandaki Province but it has also been a kingdom, i.e., one of the Caubèsī Rāja (twenty-four sovereign and intermittently allied petty kingdoms in South Asia ruled by the Khas people and unified between 1744 [by Prthvīnārāyaṇa Sāha, king of Gorkha Kingdom, cp. List. 2] and 1816 to what is now present-day Nepal (van Driem, 2001, 1107). The Time Ontology in OWL (Cox and Little, 2020) provides a wide range of classes, object properties and data types to model time instances and time spans that are of particular interest for Nepalese temporal entities entangled with toponyms: Alongside the Gregorian calendar as standard time, one can define alternative calendars through the property time:hasTRS with the object being, e.g., a DBpedia entry. The following example

¹⁹The Nepalese documents show different calendars: Vikram Samvat (VS), Shaka Era (SŚ), and Nepal Sambat (NS); when not further specified, we use CE (Common Era).
²⁰Document DNA_0013_0031, https://nepalica.hadw-bw.de/nepal/editions/show/194
²¹Details of NEPALPEOPLE are not subject of this paper.
Figure 4: Deukhuri in the named entities register of the Documenta Nepalica.

Figure 5: Historical concepts of NEPALPLACES.

5. Conclusion

NEPALPLACES presents the first step towards the creation of LINKEDOPENNEPAL. It shows very promising results but still is a work in progress: By incrementally populating the ontology with all the toponyms registered in the Documenta Nepalica text corpus, the ability of NEPALPLACES to facilitate modeling of all relevant information will be put to the test. This is a process that will most likely necessitate repeatedly revisiting its structure. To populate NEPALPLACES in an exhaustive way is, thus, a task that cannot easily be integrated into the research project’s time frame and workflow. Therefore, we consider executing this task either in the form of a satellite or, more likely, a follow-up project.
6. Bibliographical References

Acheson, E., De Sabbata, S., and Purves, R. (2017). A quantitative analysis of global gazetteers: Patterns of coverage for common feature types. *Computers, Environment and Urban Systems*, 64:309–320.

Agudo de Cea, G., Montiel-Ponsada, E., Kerneman, I., and Ordan, N. (2016). From Dictionaries to Cross-lingual Lexical Resources. *Kerneman DICTIONARY News*, 24:25–31.

Bajracharya, M. and Michaels, A. (2022). Introduction. In Manik Bajracharya, editor, *Slavery and Unfree Labour in Nepal*. Documents from the 18th to Early 20th Century, pages 6–36. Heidelberg University Publishing, Heidelberg.

Bechhofer, S., van Harmelen, F., Hendler, J., Horrocks, I., McGuinness, D., Patel-Schneider, P., and Noy, N. (2004). OWL Web Ontology Language. Reference. W3C Recommendation 10 February 2004. URL: https://www.w3.org/TR/2004/REC-owl-ref-20040210/ [accessed: 03-10-2022].

Bellandi, A., Giovannetti, E., and Weingart, A. (2018). Multilingual and Multitword Phenomena in a lemon Old Occitan Medico-Botanical Lexicon. *Information*, 9 (3), 52.

Berners-Lee, T. (2009). Linked Data. URL: https://www.w3.org/DesignIssues/LinkedData.html [accessed: 03-15-2022].

Bizer, C., Heath, T., and Berners-Lee, T. (2009). Linked Data – The Story So Far. *International Journal on Semantic Web and Information Systems*, 5:1–22.

Bosque-Gil, J., Gracia, J., Montiel-Ponsoda, E., and Gómez-Pérez, A. (2018). Models to represent linguistic linked data. *Natural Language Engineering*, 24(6):811–859.

Brickley, D. and Guha, R. (2014). XML Schema Part 0: Primer Second Edition. URL: https://www.w3.org/TR/rdf-schema/ [accessed: 03-14-2022].

Chiarcos, C., McCrae, J., Cimiano, P., and Fellbaum, C. (2013). Towards Open Data for Linguistics: Lexical Linked Data. In Alessandro Oltramari, et al., editors, *New Trends of Research in Ontologies and Lexical Resources: Ideas, Projects, Systems*, pages 7–25. Springer, Berlin, Heidelberg.

Cimiano, P., McCrae, J. P., and Buitelaar, P. (2016). Lexicon Model for Ontologies: Community Report. Final Community Group Report, 10 May 2016. URL: https://www.w3.org/2016/05/ontolex/ [accessed: 03-15-2022].

Cimiano, P., Chiarcos, C., McCrae, J., and Gracia, J. (2020). *Linguistic Linked Data: Representation, Generation and Applications*. Springer, Cham.

Ciotti, F. and Tomasi, F. (2016–2017). Formal Ontologies, Linked Data, and TEI Semantics. *Journal of the Text Encoding Initiative* [Online], Issue 9.

Cox, S. and Little, C. (2020). Time Ontology in OWL. W3C Candidate Recommendation 26 March 2020. URL: https://www.w3.org/TR/owl-time/ [accessed: 03-10-2022].

Crofts, N., Doerr, M., and Gill, T. (2003). The CIDOC Conceptual Reference Model. A Standard for Communicating Cultural Contents. *Cultivate Interactive*, 9.

Cubelic, S., Michaels, A., and Zotter, A. (2018). Studying Documents of South Asia: An Introduction. In Simon Cubelic, et al., editors, *Studies in Historical Documents from Nepal and India*, pages 1–33. Heidelberg University Publishing, Heidelberg.

Cyganiak, R., Wood, D., and Lanthaler, M. (2014). RDF 1.1. concepts and abstract syntax: W3C recommendation 25 February 2014. URL: https://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/ [accessed: 03-14-2022].

Declerck, T., Wandel-Vogt, E., and Mörth, K. (2015). Towards a Pan European Lexicography by Means of Linked (Open) Data. In Iztok Kosem, et al., editors, *Electronic Lexicography in the 21st Century: Linking Lexical Data in the Digital Age*. Proceedings of the eLex 2015 Conference, 11-13 August 2015, Herstmonceux Castle, United Kingdom, pages 342–355. Ljubljana/Brighton. Trojina, Institute for Applied Slovene Studies/Lexical Computing Ltd.

Diwasa, T., Bandhu, C., and Nepal, B. (2007). *The Intangible Cultural Heritage of Nepal: Future Directions*. UNESCO Kathmandu Series of Monographs and Working Papers No 14, Kathmandu.

Eide, Ø. (2014–2015). Ontologies, Data Modeling, and TEI. *Journal of the Text Encoding Initiative*, Issue 8.

Fallside, D. and Walmsley, P. (2004). RDF Schema 1.1. W3C Recommendation 25 February 2014. URL: https://www.w3.org/TR/rdf-schema/ [accessed: 03-14-2022].

Fallside, D. and Walmsley, P. (2004). RDF Schema 1.1. W3C Recommendation 25 February 2014. URL: https://www.w3.org/TR/rdf-schema/ [accessed: 03-14-2022].

Gandon, F., Sabou, M., and Sack, H. (2017). Weaving a Web of Linked Resources. *Semantic Web Journal*, 6:1–6.

Herman, I., Aside, B., McCarron, S., and Birbeck, M. (2015). RDFa Core 1.1 – Third Edition. URL: https://www.w3.org/TR/rdf-a-core [accessed: 03-14-2022].

Hutt, M. (1988). *Nepali, A National Language and its Literature*. Sterling Publishers, New Delhi.

Jarrar, M., Amayreh, H., and Mc CRAE, J. (2019). Representing Arabic Lexicons in Lemon – a Preliminary Study. In Thierry Declerck et al., editors, *2nd Conference on Language, Data and Knowledge (LDK 2019)*. Posters Track. Birzeit University.

Khatiwoda, R., Cubelic, S., and Michaels, A. (2021). *The Mulukı̇ Ain of 1854. Nepal’s First Legal Code*. Heidelberg University Publishing, Heidelberg.

Michaels, A. (2008). *Siva in Trouble. Festivals and Rituals at the Pasupatinath Temple of Deopatan*. OUP, New York.
Michaels, A. (2018). *Kultur und Geschichte Nepals*. Alfred Kröner Verlag, Stuttgart.

Osterhammel, J. (2011). *Die Verwandlung der Welt: Eine Geschichte des 19. Jahrhunderts*. C.H. Beck, München.

Pokharel, S., Sherif, M., and Lehmann, J. (2014). Ontology Based Data Access and Integration for Improving the Effectiveness of Farming in Nepal. In *Proceedings of the International Conference on Web Intelligence and Intelligent Agent Technologies (WI-IAT ’14)*, pages 319–326.

Prud’hommeaux, E. and Carothers, G. (2014). RDF 1.1 Turtle: Terse RDF Triple Language. W3C Recommendation, 25 February 2014. URL: http://www.w3.org/TR/turtle/ [accessed: 03-10-2022].

Rabinowitz, A., Shaw, R., Buchanan, S., Golden, P., and Kansa, E. (2016). Making Sense of the Ways We Make Sense of the Past: The Periodo Project. *Bulletin of the Institute of Classical Studies*, 59(2):42–55, 12.

Riccardi, T. (2003). Nepali. In George Cardona et al., editors, *The Indo-Aryan Languages*, pages 538–580, New York. Routledge.

TEI Consortium. (2017). TEI P5: Guidelines for Electronic Text Encoding and Interchange. Version 3.2.0. Last updated on 10th July 2017. URL: http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ [accessed: 03-16-2022].

Tittel, S., Bermúdez-Sabel, H., and Chiarcos, C. (2018). Using RDFa to Link Text and Dictionary Data for Medieval French. In John P. McCrae, et al., editors, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. 6th Workshop on Linked Data in Linguistics (LDL-2018), 12 May 2018, Miyazaki, Japan, pages 30–38, Paris. ELRA.

van Driem, G. (2001). *Languages of the Himalayas*. Brill, Leiden/Boston/Köln.

von Wartburg, W. (since 1922). *Französisches Etymologisches Wörterbuch. Eine darstellung des galloromanischen sprachschatzes – FEW*. ATILF. [continued by O. Jänicke, C.T. Gossen, J.-P. Chambon, J.-P. Chauveau, and Yan Greub].

Wood, D., Zaidman, M., Ruth, L., and Hausenblas, M. (2014). *Linked data: structured data on the web*. Manning Publications Co., New York.

Zotter, A. (2018). Papier ist geduldig. Digital editierte Dokumente geben Aufschluss zur Nationalstaatsbildung in Nepals “langem 19. Jahrhundert” (1768–1951). In Union der deutschen Akademien der Wissenschaften, editor, *Die Wissenschaftsakademien – Wissensspeicher für die Zukunft Forschungsprojekte im Akademienprogramm*, Berlin.