

**RCE (2013).** RADAR, a Relational Archaeobotanical Database for Advanced Research. Rijksdienst voor het Cultureel Erfgoed, Ministerie van Onderwijs, Cultuur en Wetenschap. Available online at: <https://archeologieinederland.nl/bronnen-en-kaarten/radar>

**van Reenen, G. (2007).** Snippendaalcatalogus database. Hortus Botanicus Amsterdam. Available online at: <http://dehortus.nl/en/Snippendaal-Catalogue>

Schooneveld-Oosterling, J., Knaap, G., Karskens, N., Smit-Maarschalkerweerd, D., Tetteroo, S., van den Tol, J., Nijhuis, H., van Wijk, K., Kunst, A., Buijs, J., Jongma, M., Boer, R. (2013). Boekhouder-Generaal Batavia. Huygens ING. Available online at: <http://resources.huygens.knaw.nl/boekhoudergeneraalbatavia>

**van der Sijs, N. (2001).** Chronologisch Woordenboek. Available online at: [http://dbnl.org/tekst/sijs002chro01\\_01/](http://dbnl.org/tekst/sijs002chro01_01/)

## 2. A Linked Data Approach to Disclose Handwritten Biodiversity Heritage Collections

Lise Stork, Leiden Institute of Advanced Computer Science (LIACS), Leiden University, Niels Bohrweg 1, 2333 CA Leiden, The Netherlands  
[l.stork@liacs.leidenuniv.nl](mailto:l.stork@liacs.leidenuniv.nl)

Andreas Weber, Department of Science, Technology and Policy Studies (STePS), University of Twente, PO Box 217, 7500 AE Enschede, The Netherlands  
[a.weber@utwente.nl](mailto:a.weber@utwente.nl)

Over the last decade, natural history museums in and beyond the Netherlands have heavily invested in digitizing and extracting biodiversity information from manuscript and specimen collections (Heerliën et al. 2015; Pethers and Huertas, 2015; Svensson, 2015). In particular handwritten fieldnotes describing occurrences of species in nature (see illustration) form an important but often neglected starting point for researchers interested in long-term habitat developments of a specific area and the history of scientific ordering, writing and collecting practices (Blair 2010; Bourget 2010; Eddy 2016). In order to disclose handwritten descriptions of flora and fauna and related specimen and drawings collections, natural history museums usually resort to manual enrichment methods such as full text transcription or keyword tagging (Ridge 2014; Franzoni et al. 2014). Often these methods rely on crowdsourcing, where online volunteers annotate pages with unstructured textual labels (Field Book Project 2016). More recently, curators of archives, data scientists and historians have started to experiment with semi-automatic annotation systems for historical manuscript collections such as the MONK system (Schomaker et al. 2016). Since MONK is a supervised learning system, a large amount of properly recognized textual labels is necessary to safeguard the system's recognition abilities.

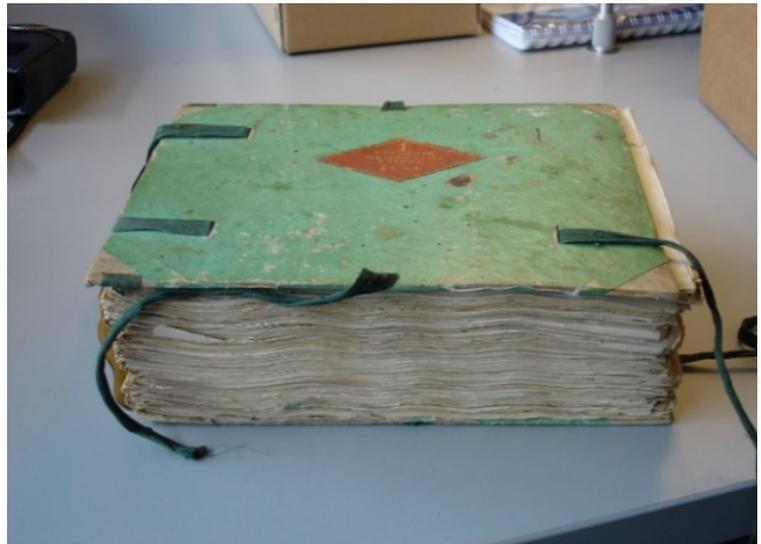


Thus, although such practices have the potential to yield high quality data, merely annotating pages with unstructured textual labels raises two problems: First, without suggestions driven by semantic

knowledge, it will be hard for volunteers or a machine to start annotating handwritten pages. Not only in the context of our case study, which deals with fieldnotes written in early nineteenth century insular Southeast Asia, but also in the context of other manuscript collections, one needs a thorough knowledge of paleography, and historical and taxonomic background information (Causer and Terras 2014). Semantics can aid the annotation process when dealing with ambiguity or provide suggestions in cases where words are hard to read and too little example instances are available. For instance, when a fieldnote describes an expedition in East- Java, a species of frogs of West-Celebes can be ruled out. Second, unstructured textual annotation will eventually result in an inefficient search process on the side of the user. Traditional keyword- based search leads to many irrelevant results or requires specific prior knowledge regarding the content. To answer more general and expressive queries, semantic relations between annotations need to be considered as well (Elbassuoni, et al. 2010).

In order to help solve such problems this paper argues for the development and application of a semantic model for semi-automatic semantic annotation. The model aggregates existing metadata standards and ontologies, following the Linked Data principles, and prepares them for semantically annotating and interpreting the *Named Entities (NEs)* in the fieldnotes of digitized natural historical collections.<sup>10</sup>

The case study of this paper is a collection of 8000 fieldnotes gathered by the Committee for Natural History of the Netherlands Indies (*Natuurkundige Commissie voor Nederlandsch-Indië*, further referred to by the acronym NC). In the first half of the nineteenth century, naturalists of the NC charted the natural and economic state of the Indonesian Archipelago and returned a wealth of scientific observations which are now stored in the archives and depot of Naturalis Biodiversity Center in Leiden (Mees 1994; Klaver 2007). An in-depth historical analysis reveals that Heinrich Kuhl (1797-1821), Johan Coenraad van Hasselt (1797-1823) and other travelers of the NC use the following NEs to structure their fieldnotes (see illustration displaying a bundle of NC fieldnotes) while traveling in insular Southeast Asia: collecting localities, dates, collectors' names, taxonomic names, and references to other printed or handwritten sources. Kuhl and Van Hasselt, for instance, regularly use the illustrations of printed works such as the *Voyage de découvertes aux terres australes (1807-1816)* by M.F. Péron as visual point of reference for their fieldnote descriptions. While links to published resources can be easily established by linking them to domain specific repositories of digitized books such as the Biodiversity Heritage Library (BHL), collection localities, taxonomic names and collectors' names are more difficult to process.



In order to be able to identify, annotate and interlink such NEs in a semi-automatic way, this paper proposes the implementation of a Knowledge Base (KB). The KB has two goals: first, the underlying data structure of the KB enables cross-matching of resources within and across fieldnote

---

<sup>10</sup> The project Semantic Blumenbach thinks in a similar direction, but then with a focus on published material (Wettlaufer et al. 2015).

collections. In order to realize this function a lightweight application ontology written in RDF<sup>11</sup> and OWL<sup>12</sup> is suggested that serves as a schema to semantically structure the KB. It expresses species observations, ensures their provenance in relation to the digitized fieldnotes and builds on existing metadata and ontology standards. Entities in turn are described using uniform resource identifiers (URIs). This allows for an integration of the fieldnote annotations into the web of Linked Data (LD) and ensures interoperability with other digital collections (Hallo et al. 2016). Second, the logical characteristics of the properties in the ontology enable a reasoner system to suggest possible NEs. In order to provide possible labels regarding these NEs, the KB is prepopulated with lists extracted from thesauri, gazetteers, and taxonomies. As regards collection localities we, for instance, draw upon the *GEOnets Names Server (GNS)*, a large semantically structured database containing historical and present-day geographical locations in insular Southeast Asia. Biological species names can be drawn from the Linnaean taxonomy of species which was already well established at the time of the NC (Farber 2000; Beckman 2012). As regards person names we rely on the database *Cyclopedia of Malaysian Collectors* which M. J. van Steenis-Kruseman compiled in the 1960s and 1970s.<sup>13</sup> Taken together, by prompting users to annotate with terms from the KB, a semantic network of annotations is formed that is able to improve the quality of the annotations and bootstraps the annotation process. The ontology and an implementation of the KB based on our case study, together with possibilities regarding supported querying and reasoning techniques, will be discussed in more detail during the presentation.

## Bibliography

- Beckman, J. "The Swedish Taxonomy Initiative : Managing the Boundaries of 'Sweden' and 'Taxonomy'" In *Scientists and Scholars in the Field: Studies in the History of Fieldwork and Expeditions*, edited by K.H. Nielsen, H. Harbsmeier, and Ch. J. Ries, 395–414. Aarhus: Aarhus University Press, 2012.
- Bourguet, M.-N. "A Portable World: The Notebooks of European Travellers (Eighteenth to Nineteenth Centuries)." *Intellectual History Review* 20, no. 3 (2010): 377–400.
- Causser, T. and M. Terras. ""Many Hands Make Light Work. Many Hands Together Make Merry Work": Transcribe Bentham and Crowdsourcing Manuscript Collections." In *Crowdsourcing Our Cultural Heritage*, 57–88. Surrey: Ashgate, 2014.
- Eddy, M. D. "The Interactive Notebook: How Students Learned to Keep Notes during the Scottish Enlightenment." *Book History* 19, no. 1 (2016): 86–131.
- Elbassuoni, S., Ramanath, M., Schenkel, R., and Weikum, G. "Searching RDF Graphs with SPARQL and Keywords". *IEEE Data Eng. Bull.*, 33(1), (2010), 16-24.
- Farber, P.L. *Finding Order in Nature: The Naturalist Tradition from Linnaeus to E.O. Wilson*. Baltimore, Md.: Johns Hopkins University Press, 2000.
- Field Book Project, Smithsonian National Museum of Natural History: <http://naturalhistory.si.edu/fieldbooks/> [accessed 15 February 2017].
- Franzoni, Ch. and H. Sauermann, "Crowd science: The organization of scientific research in open collaborative projects," *Research policy* 43, no. 1 (2014), 1-20.

---

<sup>11</sup> <https://www.w3.org/RDF/> [accessed February 15, 2017].

<sup>12</sup> <https://www.w3.org/OWL/> [accessed February 15, 2017].

<sup>13</sup> The database is available online: <http://www.nationaalherbarium.nl/FMCollectors/> [accessed February 15, 2017]

GEONets Name Server, <http://geonames.nga.mil/gns/html/> [accessed February 15, 2017]

Hallo, M., et al. "Current state of Linked Data in digital libraries." *Journal of Information Science* 42.2 (2016): 117-127.

Heerlien, M., J. Van Leusen, S. Schnörr, S. De Jong-Kole, N. Raes, and Kirsten Van Hulsen. "The Natural History Production Line: An Industrial Approach to the Digitization of Scientific Collections." *J. Comput. Cult. Herit.* 8, no. 1 (February 2015): 3:1–3:11.

Klaver, Ch.J.J. *Inseparable Friends in Life and Death: The Life and Work of Heinrich Kuhl (1797-1821) and Johan Conrad van Hasselt (1797-1823), Students of Prof. Theodorus van Swinderen.* Groningen: Barkhuis, 2007.

Mees, G.F. and C. van Achterberg. "Vogelkundig onderzoek op Nieuw Guinea in 1828: terugblik op de ornithologische resultaten van de reis van Zr. Ms. Korvet Triton naar de zuidwest kust van Nieuw-Guinea." *Zoologische Bijdragen* 40 (1994): 3–64.

Péron, F., N. Baudin, L.C. Desaulses de Freycinet, Ch. Alexandre Lesueur, and N.-M. Petit. *Voyage de Découvertes Aux Terres Australes* (Paris : De l'Imprimerie impériale, 1807).

Pethers, H. and B. Huertas. "The Dollmann Collection: A Case Study of Linking Library and Historical Specimen Collections at the Natural History Museum, London." *The Linnean* 31, no. 2 (2015): 18–22.

Ridge, M. (ed.), *Crowdsourcing our cultural heritage* (Ashgate: Farnham, 2014).

Schomaker, L., A. Weber, M. Thijssen, M. Heerlien, A. Plaat, S. Nijssen, et al. "Making Sense of Illustrated Handwritten Archives." In *Book of Abstracts, Digital Humanities Conference 2016 Krakow*, 764–66, 2016.

Svensson, A. "Global Plants and Digital Letters: Epistemological Implications of Digitising the Directors' Correspondence at the Royal Botanic Gardens, Kew." *Environmental Humanities* 6 (2015): 73–102.

Wettlaufer, J, Ch. Johnson, M. Scholz, M. Fichtner, and S. Ganesh Thotempudi. "Semantic Blumenbach: Exploration of Text–Object Relationships with Semantic Web Technology in the History of Science." *Digital Scholarship in the Humanities* 30, Suppl. 1 (December 1, 2015): 187–98.

---

### 3. Linked cultural events: Digitizing past events and its implications for analyzing and theorizing the 'creative city'

**Harm Nijboer (Huygens ING)**

**Claartje Rasterhoff (University of Amsterdam)**

#### Introduction

This paper introduces 'linked cultural events' as a novel methodological framework that allows for the systematic analysis of cultural expressions in their urban context. The events-based approach is inspired by datasets developed in the research program *CREATE: Creative Amsterdam: An E-Humanities Perspective* (University of Amsterdam, 2014-present).<sup>14</sup> In this program, the cultural sectors of performing arts take up a particularly prominent position, as data on for instance music, theatre and cinema programming is available in various formats. In terms of methodology, the data

---

<sup>14</sup> [www.create.humanities.uva.nl](http://www.create.humanities.uva.nl).