



Universiteit
Leiden
The Netherlands

Knowledge extraction from archives of natural history collections

Stork, L.

Citation

Stork, L. (2021, July 1). *Knowledge extraction from archives of natural history collections*. Retrieved from <https://hdl.handle.net/1887/3192382>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3192382>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/3192382> holds various files of this Leiden University dissertation.

Author: Stork, L.

Title: Knowledge extraction from archives of natural history collections

Issue date: 2021-07-01

Propositions

accompanying the thesis

Knowledge Extraction from Archives of Natural History Collections

by Lise Stork

1. Transcription of handwritten field notes is a labour-intensive process for domain experts. An alternative is extraction of knowledge through semantic annotation, which intrinsically motivates domain experts to produce high-quality data, and formalises minimal information required for sufficient digitisation (Chapter 3 of this thesis).
2. Storing provenance of taxonomic referencing efforts—linking verbatim names in natural history collections to current ones—as findable, accessible, interoperable, and reusable (FAIR) data on the Semantic Web, stimulates scientific discourse and knowledge discovery (Chapter 4 of this thesis).
3. Learning from real world data, where labels are difficult to obtain, samples are small and class boundaries are fuzzy, cannot rely on general machine learning, but requires strategies that rely on contextual priors, e.g., auxiliary data, the inherent structure of data, as well as domain expert knowledge (Chapter 5 and 6 of this thesis).
4. There is a need for research that analyses the performance and robustness of predictive models on complex real-world data where obtaining labels for training is costly or infeasible (Chapter 6 of this thesis).
5. Inter- and intra-collection interoperability of multimodal natural history collection objects and contemporary digital sources increases, by extension of their available knowledge, the scientific worth of singular collection objects.
6. The Semantic Web facilitates federation of semantic queries across distributed collections and thereby paves the way for the concept of *the global natural history collection*.
7. In addition to education and public outreach, natural history collections are invaluable for measuring long-term effects of human-related climate change.
8. Symbolic or neuro-symbolic methodologies are inherently more suitable for human-machine cooperation and interpretation than purely statistical methods.
9. Perhaps more difficult than formulating an appropriate solution, is formulating and solving the right task. Therefore, research spent on the application of computer science methods to a specific domain needs to be well calibrated with the demands of that domain.
10. Artificial intelligence techniques should be developed to support domain experts, rather than to replace them.
11. Scientific transparency and collaboration should be central to scientific research, and should, because of the inherently competitive nature of scientific research, be stimulated through top-down policy making.