



Legacy data in a digital age

Ellert Thor Johannsson, Simonetta Battista, Tarrin Wills

A Dictionary of Old Norse Prose, Department of Nordic Studies and Linguistics
E-mail: ellert@hum.ku.dk, sb@hum.ku.dk, tarrin@hum.ku.dk

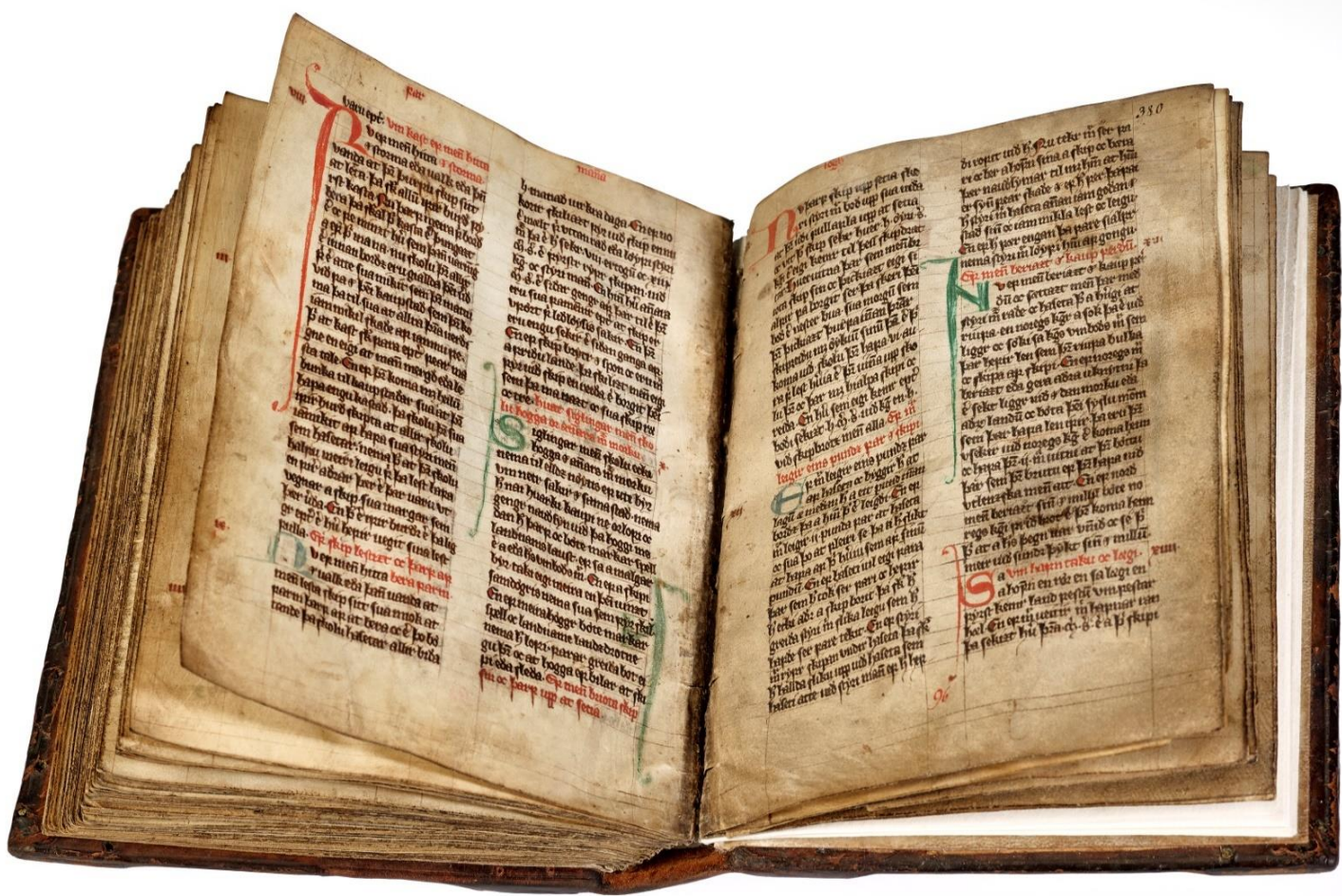
Digital Humanities in the Nordic Countries

20-23 Oktober 2020, Riga

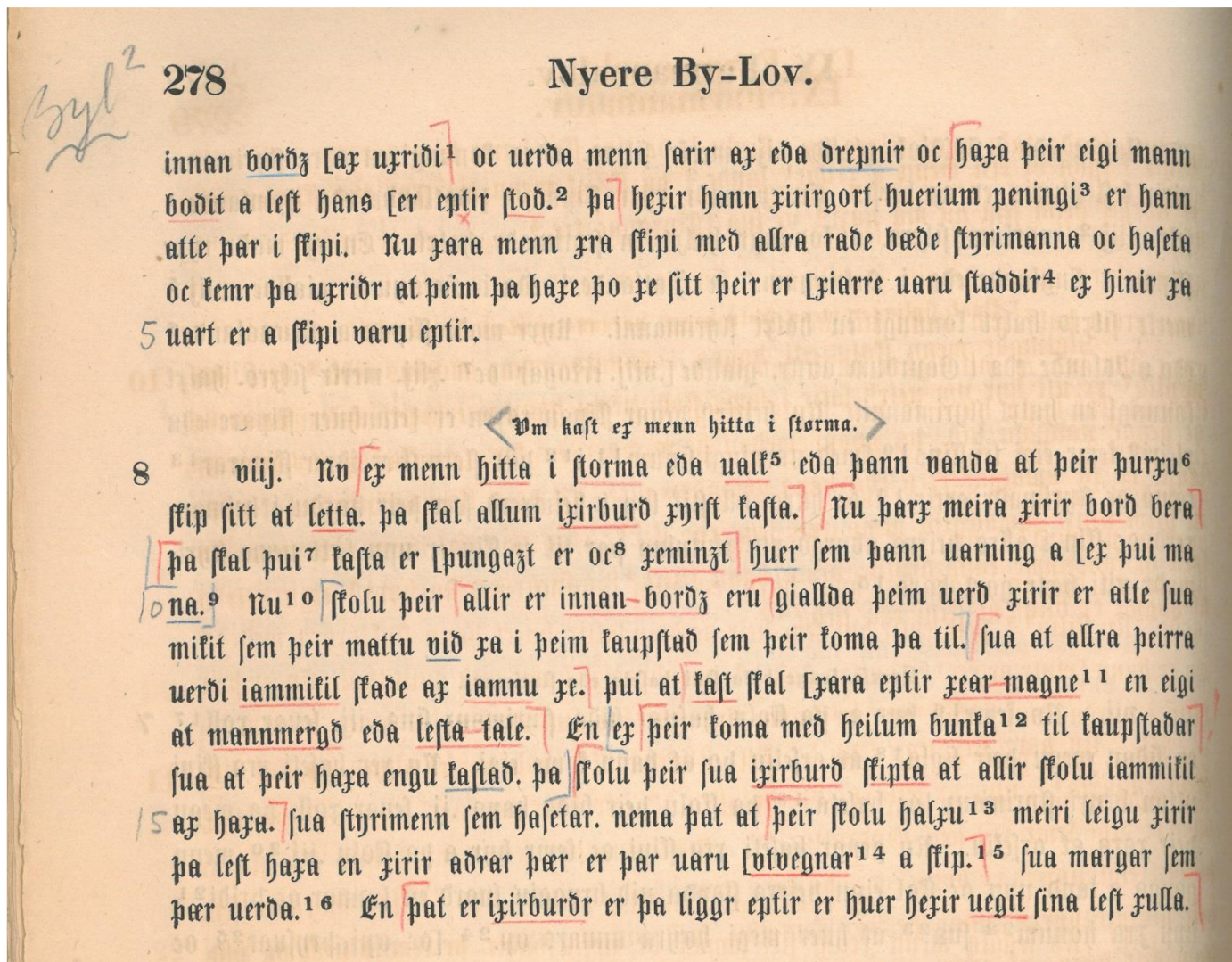
Abstract: In this poster presentation the focus is on A Dictionary of Old Norse Prose (ONP), which covers the language of Iceland and Norway in the Middle Ages. We present the data used for the making of this historical dictionary and its development during three different periods: from a rather straightforward collection of citations through a multi-volume, but incomplete, print publication to its current state as a dynamic online lexicographic tool, providing detailed information about the vocabulary of Old Norse and its textual foundation in medieval manuscripts. Even though the dictionary is not finished, its wide scope is evident by the fact that its archive of around 800,000 example citations represents an estimated 7% of the entire 11 million word corpus of Old Norse. The long history of the project gives a unique opportunity to study the development of the data and how it has been organized and used while the project has been in existence.

1. Background

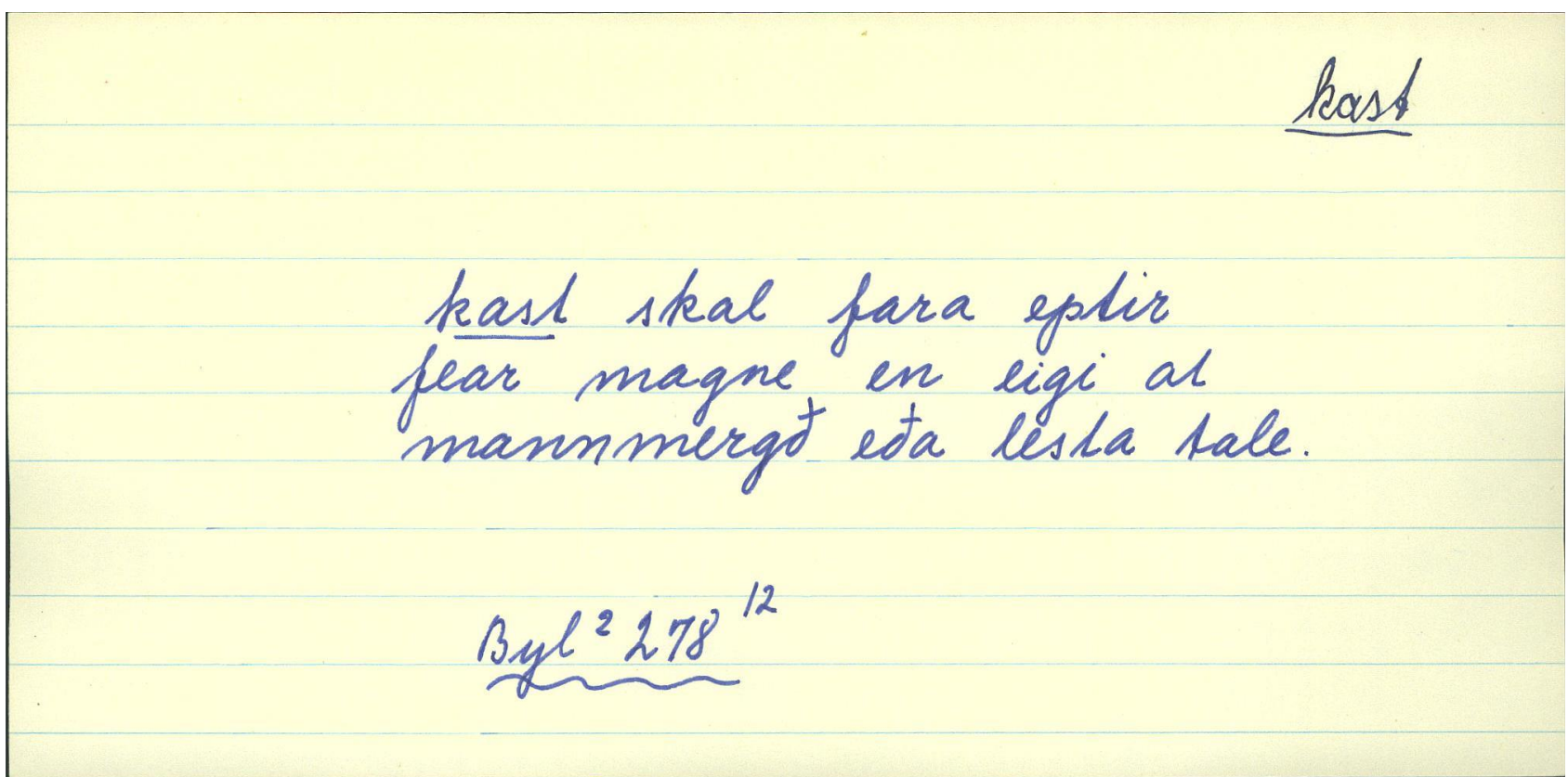
- The work began in 1939 long before computers and databases became available.
- The corpus consists of prosa texts from Medieval Iceland and Norway, which are for the most part edited in critical editions.
- The nature of the material and the editorial principles demanded a wide variety of data be collected and organized (cf. Widding 1964).
- This involved excerpting the source material for examples of word use, which were then copied onto slips and filed alphabetically in a physical archive.



The law manuscript AM 60 4to (c1320) 95v-96r



Text edition of manuscript text (Keyser & Munch 1848)



Excerpted citation from the text on a handwritten slip

2. The data

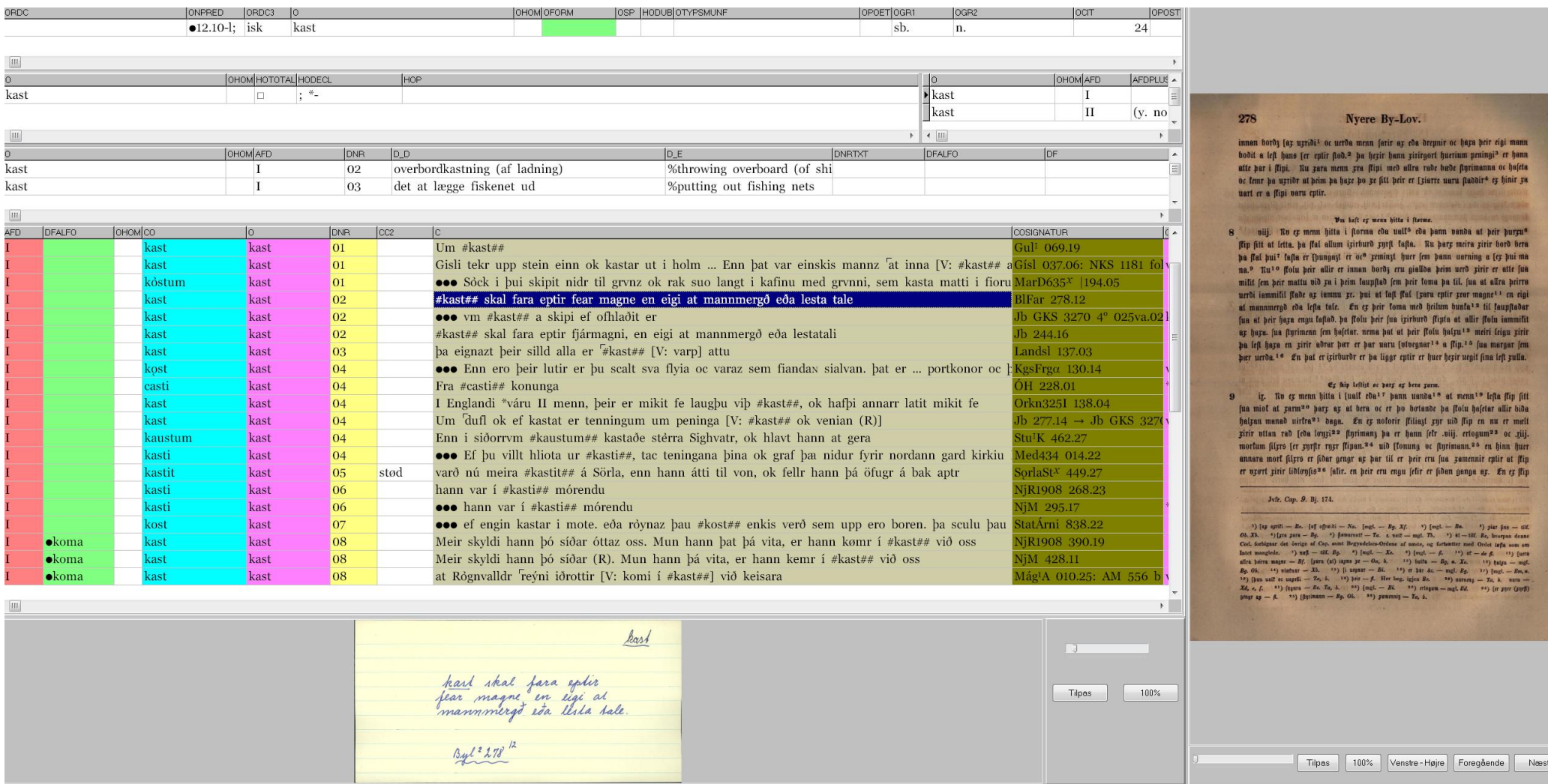
- ONP has 800.000 example citations, each of which is provided with a sentence illustrating a specific form and/or meaning of the headword, a detailed reference showing the work of origin as well as page and line number.
- Besides the dictionary citation archive, it was important to keep track of various information relating to the source material.
- The citations included a reference to scholarly editions as well as the manuscripts these were based on, so all this information had to be registered as well. This work was also done by hand.

3. The database and the printed dictionary

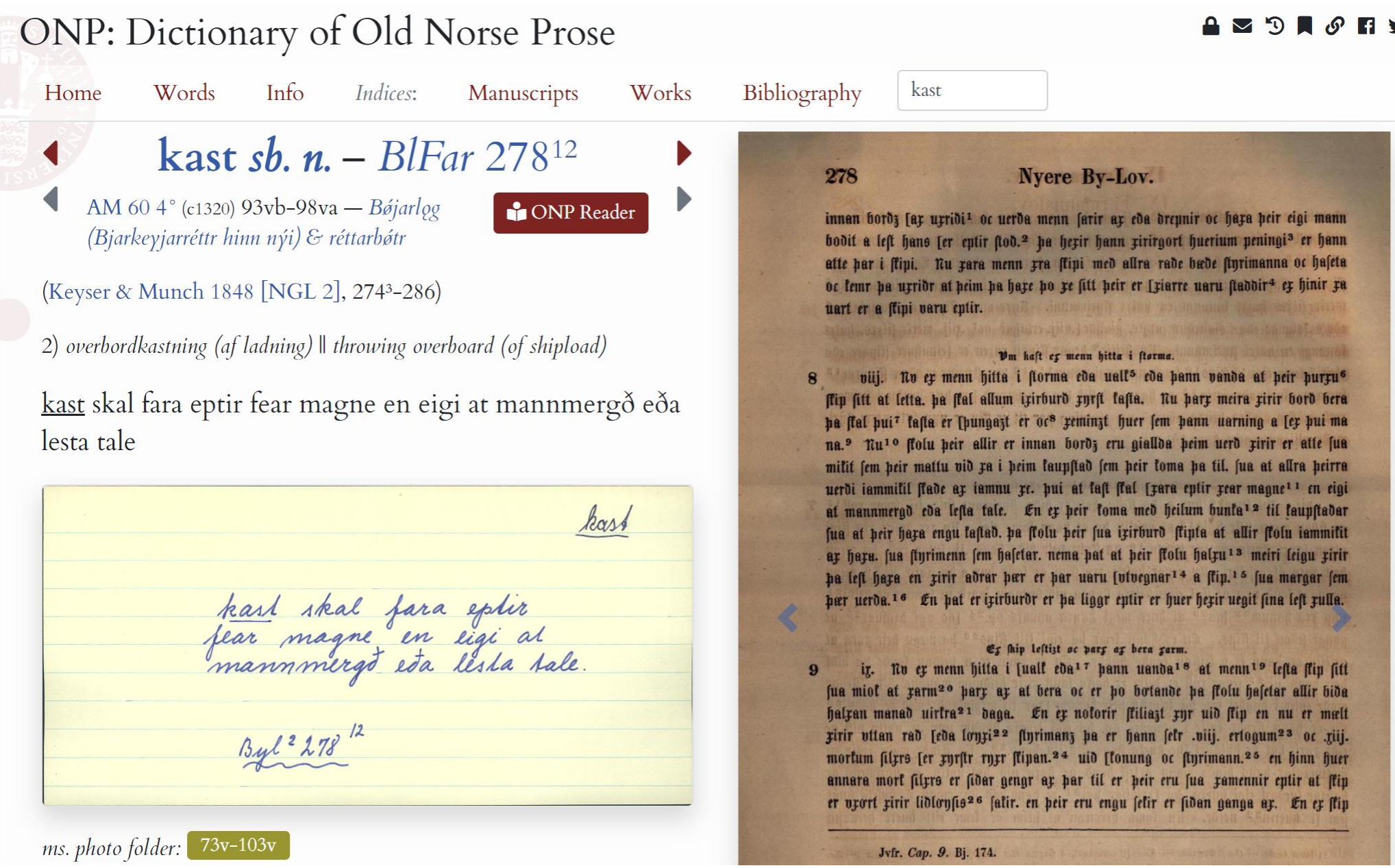
- In the 1980's: Advantages of working with dictionary data in a database structure.
- The database kept track of all the dictionary citations as well as the data related to the source material.
- This involved creating many different tables, such as wordlist table, headword table, definition table, citation table. Additional tables hold references to secondary literature and other glossaries.
- The tables were linked together through the common headword field.
- This allowed for additional information relating to both the source material and each citation, such as geographical provenience of manuscripts and grouping source material into literary genres.
- The information from the hand-written index registry was integrated into the database and made available as a printed volume in 1989 (ONP Registre).
- The database work facilitated the editing and publishing of dictionary entries. The process involved supplying extra grammatical information as well as information about collocations and syntactic relations.
- Three printed volumes of the dictionary were published from 1995 to 2004, covering the alphabet from a-em.

4. Restructuring of ONP as an online dictionary

- In 2004 print publication was halted and the project underwent a restructuring process, which resulted in an online version made available in 2010.
- The conversion to a digital online dictionary involved the scanning of ca. 500.000 non-typed paper slips, which were integrated into the database (cf. Johannsson 2019).
- Critical editions were also scanned and linked to the citations, in order to show every citation in context.
- These database enhancements allowed for a more efficient editing process and online users could query the database directly being no longer limited by the printed alphabetical list of dictionary entries (cf. Johannsson & Battista 2016).



Database interface for editing



Citation example from the current version of ONP Online

5. Current version of ONP Online - onp.ku.dk

- The online version has gradually been expanded with new edited entries and improved with additional options.
- The current platform makes it possible to search the data by several criteria, e.g. the dating of the original manuscript, country of provenance, work, literary genre.
- The data have been enhanced by linking them internally and to other digital resources. There are now links to other dictionaries as well as digital editions of Old Norse texts (cf. Wills et al. 2018).



A new ONP reader feature provides glossaries to scholarly text editions:



12 **kast** = **kast** sb. n. (27): 2) *overbordkastning (af ladning)* // *throwing overboard (of shipload)*
12 **fear magne** = **fjarmegin** sb. n. (27)
12 **var. magne** = **megin** sb. n. (134): 1) *kraft, force, styrke, voldsomhed, (voldsom) entusiasme*
12 **mikil** = **mikill** adj. (861): 5) [af e-u] (*om grad/styrke*) *stærk, heftig*

6. Conclusions

Legacy data, originally only organized in a paper filing system, have been structured in a database and improved in various ways throughout the project's history. The original data still provide the basis of the dictionary but the database structure and linking of the data has facilitated the use of the information in many different ways. Search capabilities made possible in the digital version of the dictionary and later linking of the data to external digital sources further enhance the value of the legacy data and provide innovative ways to bring together and take advantage of all the information gathered by ONP during its existence.

References

Johannsson, E. (2019). "Integrating analog citations into an online dictionary", in C. Navarretta, M. Agirrezabal, B. Maegaard (eds.): *Proceedings of the Digital Humanities in the Nordic Countries 4th Conference*, pp. 250-258.

Johannsson, Ellert Thor & Simonetta Battista (2016). "Editing and Presenting Complex Source Material in an Online Dictionary: The Case of ONP", in Tinatin Margalitadze & Georg Meladze (eds.): *Proceedings of the XVII EURALEX International Congress: Lexicography and Linguistic Diversity, 6-10 September 2016*, Tbilisi, 117-128.

Keyser, R. & P.A. Munch (eds.) (1848): *Norges Gamle Love indtil 1387*, Vol. II. Christiania: Chr. Gröndahl.

ONP Online. Accessed at: onp.ku.dk [January 2020].

ONP = Degnbol, H., Jacobsen, B.C., Knirk, J.E., Rode, E., Sanders, C. & Helgadóttir, Þ. (eds.): *Ordbog over det norrøne prosasprog / A Dictionary of Old Norse Prose*. ONP Registre (1989). ONP 1: a-bam (1994). ONP 2: ban-da (2000). ONP 3: de-em (2004). Copenhagen: Den Arnamagnæanske Kommission.

Widding, Ole (1964): *Den Arnamagnæanske Kommissions Ordbog, 1939-1964: Rapport og plan*, Copenhagen: G.E.C.GADS Forlag.

Wills, T., Johannsson, E., & Battista, S. (2018). Linking Corpus Data to an Excerpt-based Historical Dictionary. In J. Čibej, V. Gorjanc, I. Kosem, S. Krek (eds.): *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts*. Ljubljana: Ljubljana University Press, Faculty of Arts, pp. 979-987.