

THE SKALDIC PROJECT AND LEXICON POETICUM*

Tarrin Wills

University of Copenhagen, University of Sydney

tarrin@hum.ku.dk

1. Introduction

The Skaldic Project is in the process of producing a print and electronic edition, published online, of a corpus of early Scandinavian poetry, with five of the planned eight volumes completed at the time of writing, representing 75% of the total corpus. The project is an example of particular problems in digital editing and publication, some of which required custom solutions. One of these problems was the need for parallel development of the editions along with the resources that support the editing and resources that build on the edition. The project, which at the time of writing is in its 23rd year, is simply too long-term to require that each stage of publication to be fully completed for the whole corpus before the next stage is begun. The asynchronous development of the digital resource, where digital materials are available for further use before the editions have been finalised, requires a complex user interface. Some aspects of this interface and the methodology it represents are described here.

The project also brings together the texts in a number of different contexts: their original composition, preservation in prose works, manuscript record, as well as their interpretation. The project links all of these contexts digitally to the poetry so that editors and scholars outside the project can understand

* Acknowledgements: The research that underlies this paper has been supported in particular by the Australian Research Council, Horizon 2020, Rannis and Deutsche Forschungsgemeinschaft. The digital resource would be of very little value without the enormous amount of high quality editions and commentaries that it contains, and that work is largely supported by individual editors in their own time, and extremely thoroughly checked by the general editors (Margaret Clunies Ross, Kari Ellen Gade, Edith Marold, Diana Whaley, Guðrún Nordal, and, to a much lesser extent, the present author) also in their own time.

the poetry in all its contexts. The methodology of the project attempts to preserve the connections between all the contexts, in particular, maintaining a digital connection between the physical record of the corpus (manuscripts and inscriptions) and all aspects of the poetry and its interpretation.

The poetry itself presents particular challenges because of the complexity of its diction, the variation between manuscripts, and its lexicon. This complexity is also represented digitally, and the relevant digital structures and methods are also described in this article.

The solution to these issues has been to manage the data in a relational database with web applications for entering, editing, publishing and exporting the edition for print. This paper outlines the problems peculiar to this corpus, but which may be applicable to other corpora, and the solutions to those problems that have been developed in the course of the project.

2. Background

The poetry broadly termed skaldic (from the Old Norse *skáld* «poet») was composed in Northern Europe (particularly Norway and Iceland) and its outposts, such as settlements in the British Isles, from the ninth century until the fourteenth century. The corpus as currently understood was edited from its manuscripts by Finnur Jónsson together with a normalised and heavily emended text and simplified Danish translations in a separate volume¹. Finnur Jónsson also produced a dictionary based on the corpus² and Rudolf Meissner an index of the complex periphrases known as *kenningar* or kenning³, also based on this corpus, but with some revision and in many cases expurgation. E. A. Kock comprehensively revised Finnur's emended text and interpretations⁴ but did not attempt to go back to the manuscript evidence beyond Finnur's text, nor to address the complex problem of attribution and classification for this material. It was therefore increasingly clear that the corpus of skaldic poetry, as well as the reference tools that depended on it, was in

¹ Finnur Jónsson, *Den norsk-islandske skjaldedigtning. A: Tekst efter håndskrifterne; B: Rettet tekst*, Copenhagen: Rosenkilde & Bagger, 1912-1915. Volumes A I-II contain the manuscript text; B I-II contain the normalised and emended texts and translations.

² Finnur Jónsson, *Lexicon poeticum antiquæ linguae septentrionalis: Ordbog over det norsk-islandske skjaldesprog oprindelig forfattet af Sveinbjörn Egilsson*, Copenhagen, Møller, 1913-1916 (second revised edition 1931).

³ Rudolf Meissner, *Die Kenningar der Skalden: Ein Beitrag zur skaldischen Poetik* [Rheinische Beiträge und Hilfsbücher zur germanischen Philologie und Volkskunde 1], Bonn and Leipzig, Schroeder, 1921.

⁴ Ernst Albin Kock, *Notationes Norræna: Anteckningar till Edda och skaldediktning* [Lunds Universitets årsskrift new ser. 1. 28 vols], Lund, Gleerup, 1923-1944; Ernst Albin Kock, *Den norsk-isländska skaldediktningen*, Lund, Gleerup, 1946-1950.

need of a thorough revision. Jón Helgason had begun preparatory work on a comprehensive new edition, including overseeing a card index with detailed references to the poetry's manuscripts. It was only in 1997 that five scholars (Kari Ellen Gade, Guðrún Nordal, Diana Whaley, Edith Marold and Margaret Clunies Ross, based respectively in the US, Iceland, England, Germany and Australia) began planning a new edition of the corpus of Old Norse skaldic poetry. Shortly afterwards the present author became involved in the project, developing a digital approach to the edition.

In the early stages, a trial digital edition of *Ragnarsdrápa* based on Margaret Clunies Ross's unpublished BLitt thesis and new readings of the manuscripts was produced as an interactive web page. In 2001 Clunies Ross gained a large grant to develop the project further and it has since been supported by funding from various bodies, lead by the various main editors of the project.

From a digital point of view, there were a number of challenges with this project. Some of these had to do with the project itself and its structure. Firstly, there was a large number of people contributing to the project spread internationally, with an increasing number of contributors and assistants beyond the then five main editors (now around 50): the project would need to use internet-based resources to coordinate its work, and as editing began, resources that could be updated internationally. Secondly, there was an enormous amount of material that needed to be provided to editors to support their work, particularly information about the manuscripts of the poetry and images of them. Thirdly, editors would have to work concurrently and much of the work would have to be pursued in parallel, without requiring that all stages to be complete for the whole corpus before proceeding to the next. This differs from many other digital editing projects, for example, the 1641 Depositions Projects, where a large XML corpus was completed before being processed in later stages⁵. In more recent years new resources have been developed out of the project which has only been possible with an architecture that allows for parallel and asynchronous editing of the texts.

There are also a number of challenges that stem from the nature of the corpus itself: as previously mentioned, the poetry is recorded in a number of different contexts that have different structures. It is normally recorded piecemeal within manuscripts of prose works, but often has a different manuscript transmission to the prose works. The prose context is important to understanding the poetry and needs to be preserved alongside the reconstruction of

⁵ See Mark S. Sweetnam and Barbara A. Fennell, «Natural language processing and early-modern dirty data: applying IBM Languageware to the 1641 depositions», in *Literary and Linguistic Computing*, 27:1 (2012), pp. 39-54 <<https://doi.org/10.1093/llic/fqr050>>.

the poems themselves, which is recorded separately. And the previous reconstructions of the poetry established by Finnur and referenced in almost all the scholarship since also needed to be accessible digitally along with this other information. Each stanza therefore belongs, as it were, to a series of manuscripts, normally one or more prose works, Finnur's reconstruction, and the new edition's often revised reconstruction. Importantly, the digital resource is designed to maintain the connection between the text and its material record (manuscripts and inscriptions) as the foundation of the text and of research on the text.

Perhaps the most complex challenge for the digital project is the poetry itself. The most common metre, *dróttkvætt*, is highly intricate. It extends the common Germanic verse form which, in simplified terms, contains four metrical positions per line (sometimes referred to as a half-line or verse). Each line has two stressed syllables and two or more unstressed syllables, with the first stressed syllable of an even line alliterating with one or two of the stressed syllables in the previous (odd) line. Extending this basic form, *dróttkvætt* restricts the unstressed syllables to two; it requires two alliterating syllables in each odd line and only the first syllable of an even line can alliterate. It adds a trochée to the end of each line, the stressed part of which must rhyme with one of the two stressed syllables in the same line; in odd lines this must be half-rhyme (different vowel followed by the same consonants) and in even lines it must be full rhyme, although these internal rhymes are less strict in the very earliest poetry. Old Norse poetry, unlike other early Germanic poetic forms, is normally stanzaic, with the normal stanza consisting of 8 lines, with each 4-line unit syntactically separate. Probably in order to meet these complex metrical requirements, the word order of the half-stanzas tends to be very convoluted⁶, with intercalary clauses frequently inserted in the middle.

The following is an example of a half-stanza, with alliterating letters in bold and internal rhymes underlined:

Áðr **djúphugaðr** **dræpi**
dólga ramr með hamri
gegn á græðis vagna
gagnsæll faðir Magna⁷.

⁶Tarrin Wills, «The Development of Skaldic Language» in *Á austrvega: Saga and East Scandinavia*, ed. by Agneta Ney, Henrik Williams and Fredrik Charpentier Ljungqvist, Gävle, Gävle University Press, 2009, pp. 1032-1038.

⁷Tarrin Wills, «Anonymous Lausavísur: Stanzas from the Third Grammatical Treatise 27» in *Poetry from Treatises on Poetics* [Skaldic Poetry of the Scandinavian Middle Ages 3], ed. by Kari Ellen Gade and Edith Marold, Turnhout, Brepols, 2017, p. 555.

Editors reorder the verse into a prose-like syntax for further analysis, in this case: «Áðr djúphugaðr, ramr, gegn, gagnsæll {faðir Magna} dræpi á {dólga {græðis vagna}} með hamri» (the brackets are explained below). The editor also provides a translation, with glosses for the complex diction: «Before the deep-minded, mighty, reliable, victory-blessed father of Magni <god> [= Þórr] struck at the enemies of the sea of wagons [LAND > GIANTS] with his hammer.»

The diction of the poetry is also very rich and intricate, with a range of poetic synonyms (*heiti*, e.g. «Magni» as a name for a god above) used to represent common concepts. Most striking is the use of the expressions known as *kenningar* or kennings, which are marked up in the prose text with curly brackets prior to data entry. These consist, in their simple form, of two nouns which substitute for a third noun: a base-word (normally in a metonymic or metaphoric relationship to the referent) and a determinant (which normally defines the semantic domain of the referent); hence «sea of wagons» in the example above is «land». The determinant can be substituted itself for a kenning (e.g. «enemies of the sea of wagons» is «enemies of land», i.e. «giants»), sometimes several times within the same construction. Kennings largely follow a limited set of patterns and a relatively constrained number of referents, but the actual lexical variation is enormous, with thousands of combinations evidenced in the corpus. In the example above, with the referents of the kennings resolved and adjectives removed, the text means simply, «Before Thor struck the giants with (his) hammer».

From a digital point of view, it was desirable that, when the words of the text are reordered into the putative prose order, the connection between each word in both versions was maintained. The prose ordering assists the reader in understanding the syntax and kennings – Finnur Jónsson’s edition includes such a reconstruction for the more complex metres, as do many other editions of this poetry, such as in the notes to the Íslenzk fornrit editions. The Skaldic Project’s method links the words in the text, the prose order and the translation, each with its own ordering. Because the kennings are marked up and referents are inserted into the English translation in the edition, maintaining the link between the text and translation at the level of the word also means that the kennings could effectively be indexed at the point of digitisation. This method would therefore also build a resource that is very close to what is required to update Meissner’s index of kennings.

3. Stage 1: A resource for editors

The project turned Jón Helgason's card index of the stanzas and their manuscripts into a database of poets, poems, stanzas and their manuscripts. This information was used to determine which manuscript pages should be scanned, mainly using the analogue photograph library at the Arnamagnæan Institute, University of Copenhagen. To make this available to editors, the present author wrote a web application in the Perl programming language, with the data stored as a perl data structure and editable through the web interface. This method, however, was not very scalable as the amount of data increased, as the entire dataset had to be loaded into the web server memory each time a page was accessed. Soon the data was moved to a MySQL database server (2002-3), and the application eventually rewritten in PHP (2003-4) for performance reasons.

The content included a database of poets («skalds»); poems or groups of stanzas; and the stanzas themselves. These were linked together in two hierarchies: the original classification used by Finnur Jónsson, Kock, Meissner and others, and the revised classifications being developed by the new edition, which have continued to develop as the project has progressed. These stanzas were also linked to prose works using data supplied by the Dictionary of Old Norse Prose (ONP)⁸, and to manuscripts based on Jón Helgason's card index, but with detailed manuscript data also from ONP. This meant that, by browsing the web interface to the database, the corpus could be understood in the context of the previous editions, the emerging new edition, the prose context of the poetry and its manuscript preservation. As the database largely recorded the page on which stanzas occurred, one could also access scanned images of the relevant pages where available. Further information was added such as the editor responsible and project volume. The integration of data about manuscripts (including images) and prose works means that the resource is used heavily by those working even outside the area of skaldic poetry to find materials relating to Old Norse text in general.

The text of each stanza was encoded at this stage as TEI fragments, with the text, prose and translation recorded separately and stored in the same database table row as the other information about the stanza. This format was used to export the first published volume of the project (*Poetry on Christian Subjects*), but was later transformed into a relational format, as is discussed in «Stage 3» below.

⁸ Helle Degenbol et al., *A Dictionary of Old Norse Prose / Ordbog over det norrøne prosasprog*, Copenhagen, The Arnamagnæan Commission, 1989. Link: <<https://onp.ku.dk>> [accessed: 20/11/2019].

4. Stage 2: An interface for editing and publication

The initial resource for editors was in practice a fully functional web application for managing and publishing poetic data. The architecture which emerged for this application is known as LAMP (Linux operating system, Apache web server, MySQL or MariaDB SQL relational database management system, and PHP (or Perl/Python) application programming language). This open source suite is frequently supported by research institutions. Installations with separate web and database servers were available, for example, to researchers at all three institutions where the project has been hosted (University of Sydney, University of Aberdeen and University of Copenhagen). It can also be easily set up on a single Linux virtual server, as was also implemented at one stage for this project using the Australian Nectar service for researchers. LAMP configurations are also highly scalable, suitable for small applications up to global user-contributed resources such as Wikipedia, Twitter and Facebook, which use effectively the same software. The project consequently has not changed its architecture in fifteen years, despite the enormous expansion of the project and other projects which use the data and/or share its architecture.

The architecture, showing how the user interacts with the resource, is shown in Figure 1. Importantly, the application can be developed not only to provide an end-user interface, so that scholars and others can access the poetic corpus in all its complexity, but it can also be used to build an editing interface, where members of the project can add and edit the editions and related data.

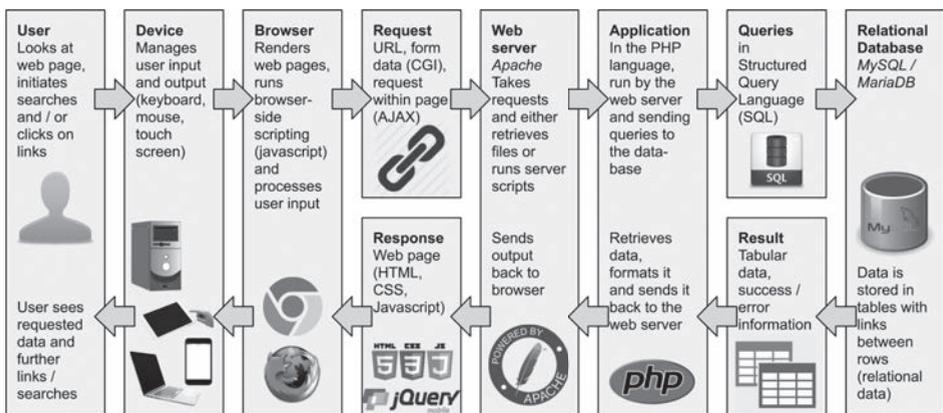


FIGURE 1. LAMP web application architecture

Developing websites in the LAMP architecture requires at least basic knowledge of a number of technologies including the database query language (SQL), a web server programming (API) language (PHP, Python, Perl or others) and the technologies used to render web pages (HTML, CSS and possibly Javascript). Programming of the application can use these technologies in varying amounts and ways to manipulate the data for the end-user, but most projects will focus on development on the web server. This project has had a single developer (the present author) with support and advice from central IT support, who have also looked after server administration. As the developer was also responsible for other parts of the project and has rebuilt large parts of the application over the years, it is difficult to say what resources would be required to get such a project to completion. With good central IT support and one or two dedicated developers we nevertheless would estimate 1-2 person years development time including training if such a complex project were to start again. Development has become far simpler since the project started thanks in part to integrated frameworks and tools that assist in the different components, particularly web layout (the project has switched to JQuery Mobile to format web pages, for example). This project has, perhaps unusually, concentrated on using the query language itself (SQL) to generate web page content with only generic manipulation on the web server. This allows for fast development of new page views, as development only needs to be coded in one part of the system.

4.1. Editing

Having developed an application for both updating and publishing the contextual information about the poetry, the next step was to extend the application so that it could be used to edit the poetry and its apparatus directly. The LAMP web application was then extended in various ways to allow project members to add the editions and translations, textual and critical apparatus to the database via the web interface.

A simple plain text edition, with the verse in lines, is sufficient to generate an encoding of the words and lines in the text for further linking and manipulation. In the first stage of data entry, the edited text is inserted and the application generates a digital representation of the lines and words, with automatic numbering of each. In the second stage the text is reordered to form a digital representation of the prose word order, linked to the verse text – the prose order clarifies the syntax for the user and provides a bridge to

the translation. Each word is saved in the prose order with separate sorting information. This means that the text and prose order is intrinsically linked, allowing for words, for example, to be searched in their syntactic as well as metrical context, or reordered interactively by an animation (see Figure 5 below). The translation is likewise based on the prose order, with the web application allowing for reordering into a more idiomatic English word order using a drag-and-drop interface, and English translations inserted for each word (Figure 3). While literal word-for-word translation is neither possible nor desirable for the edition, it has nevertheless proved very useful to have contextual translations linked to words in this way: these translations help to build the lexicographic resource, for example.

update: < Anon (TG7) 27th >

text | prose order | translation | kennings | lemmata | grammar | reorder text | realign prose | realign trans | split words | join words | delete words | lexical variants | kenning variants | kenning analysis
metrical analysis | index | reorder prose | reorder trans | split text

(view: QC format: stanza | poem | export for book: stanza | poem)

This form uses *Reykhoft*Times. To enter text using Unicode, change the settings in *preferences*.

translation: use the fields below to enter the translation aligned with the words of the prose word order; move the words into the order that produces the closest match between the original text and translation; add punctuation in the intervening boxes (the underscore represents a space, or delete the underscore if there should be no space); use the punctuation fields to add brackets for kennings (if you need a bracket in a compound word, you should enter the word with an underscore in the first stage to split it into two parts); italics can be represented with angled brackets; add hestl references (<->) in the same box following the translation (not in the punctuation box).

current translation (this will be preserved if no update is made):
Before (the deep-minded, mighty, reliable, victory-blessed father of Magni <god-> [= Þórr] struck at (the enemies (of the sea of wagons)) [LAND > GIANTS] with his hammer.

next stages: (new interface) *variants* and *add notes* — or use the notes and variants forms above

Current state of form (experimental): Before (the deep-minded, mighty, reliable, victory-bl

(initial punctuation)	Aðr	djúp	hugaðr	hamr	reliable	gagn	sæll	
	Before	the deep	minded	mighty	reliable	victory	blessed	
	faðir	Magna	dreppi	á	dólga	græðis	vagna	með
	father	of Magni <gd>	struck	at	the enemies	of the sea	of wagons	with his
	hamr							
	hammer							

click on the update button to update words and move to the next stage

update

last change: 24/01/17 12:22 (view_full.log)

FIGURE 2. Drag-and-drop manipulation of prose order and translation

The kennings are marked up in the prose order and translation by brackets marking the beginning and end of the kenning, as well as embedded kennings, as shown in the example above. The application then processes these brackets, creating entries in a kenning index table with links to the words, as well as prompting the user to insert the kenning referents.

The textual apparatus is likewise built using links to firstly the words in the text and the manuscripts in which each variant occurs. This allows not only for a consistent representation of the apparatus, but can be used to reconstruct particular versions of a stanza as represented in each manuscript (see Figure 4 below).

Critical and explanatory notes to the text are also linked to the relevant words in the text. The content of the notes are saved as formatted text using an embedded text editor in the web application.

This process has largely been performed by hired assistants who take the information from Word documents and combine it with the data already in the database. Almost all editors are nevertheless able to perform this task themselves, and to edit the information once entered by others.

Once entered, the information can be modified relatively easily by editing the relevant data tables via a series of forms that are linked to each stanza. The links between tables use arbitrary database keys and therefore the links are not broken by changes to any of the information in the table: changing the spelling of a particular word in the text, for example, means that the text is automatically updated for the variants which link to it, rather than losing the connection to the word. Some information is harder to modify, such as deleting or inserting words into the text, because the word order information is effectively linked to many other parts of the stanza. For these changes, there are additional forms that update the relevant information (such as word sorting) throughout the linked data.

The project attempts to use consistent in-text references for both internal and bibliographic references. The database also contains an extensive bibliography (edited by Hannah Burrows) which can be used to automatically index almost all in-text references entered into the discursive text in the database. This indexing is two-way, meaning that the user can find all references to a particular poem or bibliographic resource elsewhere in the corpus. This type of indexing, however, does not use the database's internal linking mechanisms («foreign key constraints») and needs to be periodically updated.

4.2. Digital Publication

The digital publication of the poetry can in principle occur in real time, when any changes are made to the database. The application writes changes to the database with any update operation, and it always reads the data directly from the database in generating and updating pages. However, many parts of the edition – effectively those published in the print version – need to go through peer review and correction processes, and are subject to the copyright agreement with the print publisher (Brepols). The digital publication therefore includes various levels of access. Editions and their apparatus that are still under review are available only to project members, who can see in real time how the edition will appear as they are working on it, and refer to

others' in-progress editions when preparing and checking their own or other material. Editions that are published but still within the embargo period are available to individuals and institutions who have purchased the print volumes and applied for access, showing the text and translation; textual and critical apparatus; and introductory materials. All this material is made public once the publication is out of embargo. (At the time of writing around 50% of the corpus is publicly available.) The contextual and metadata for the whole corpus is public, including the manuscript and prose contexts for the poetry, literature references, and the normalised text from Finnur Jónsson's edition.

The user can search and browse the poetry according to its poets, poems, prose works, manuscripts, and the contents of the printed volumes. Most of the information about the poetry is provided when a user finds or navigates to a particular stanza. The stanza data shown in the web application includes links to lines and words of the verse text, plus manuscript references, textual apparatus entries, notes, the kenning index, and automatically-generated links based on in-text references. This wealth of information is made available in different ways.

In the latest version of the interface, a HTML5 framework (jQuery Mobile) is used to render pages using Adaptive Web Design principles, that is, allowing the same content to adapt to different devices (keyboard and mouse desktop browsers, as well as touch-screen tablets and phones), as shown in Figure 3. (The interface is likely to be updated in the coming period with a better-supported framework such as Bootstrap.)

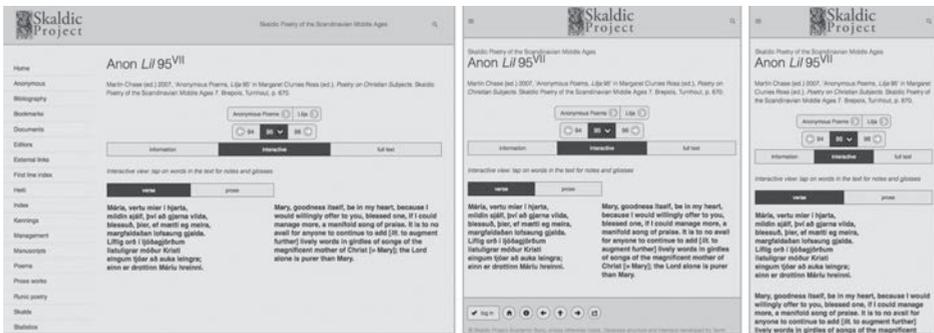


FIGURE 3. A stanza in the Skaldic Database adapting to different screens

The default view gives the text, prose order and translation, together with linked lists of variant readings, manuscript sources, editions and prose works. This view can be used to navigate between the stanzas in a poem or other grouping, as well as the prose sources. The manuscript list shows

transcriptions where available, a reconstructed normalised text for each manuscript, based on the edited text and variant apparatus, as well as (in most cases) a link to images of the relevant manuscript pages on which the stanza occurs.

Figure 4 shows the default view, with text (left), manuscript versions (centre, including showing a popup manuscript image which links to a full, zoomable, image of the manuscript page), and edition and prose context information (right, including a popup bibliographic reference which opens when the user clicks on underlined references)⁹.

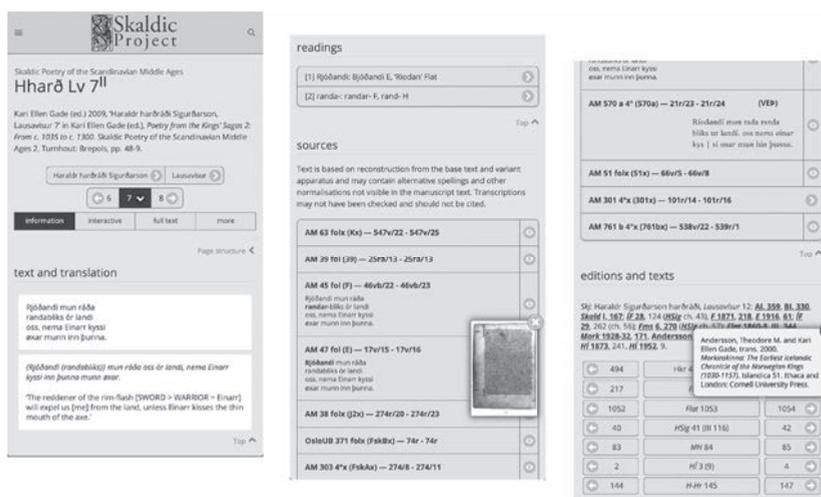


FIGURE 4. Information shown with each stanza (phone view)

Figure 3 above shows the interactive view of the stanza text. This view gives an overview of the stanza itself and the information connected to each word. In this view, clicking on a word will show the relevant translation, kenning(s), variants and notes as a popup window. Further, the alternative verse and prose ordering of the stanza is represented (via a javascript plugin) as an animated reordering of the text: when the user clicks on the «verse» / «prose» tabs the words move in relation to each other to show the different orderings (Figure 5). This gives the reader a visual sense of the extent to which the verse form have required the poet to rearrange the normal syntactic ordering of words and phrases in the stanza's text.

⁹The edition shown is Kari Ellen Gade, 'Haraldr harðráði Sigurðarson, *Lausavísur*' in *Poetry from the Kings' Sagas 2: From c. 1035 to c. 1300* [Skaldic Poetry of the Scandinavian Middle Ages 2], ed. by Kari Ellen Gade, Turnhout, Brepols, 2009, pp. 42-46.

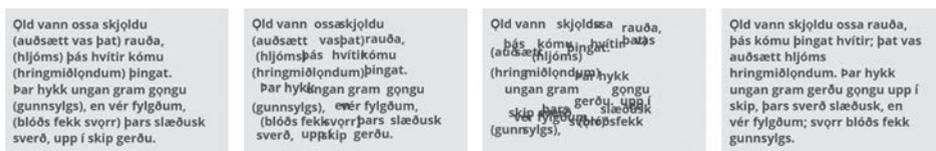


FIGURE 5. Animated reordering of verse into putative prose word order

Lastly, the edition can be viewed online in a format equivalent to how it appears in the print format, showing much of the information about the stanza in unlinked text form. The print format is described in the next section.

4.3 Print Publication

The project was originally envisaged in 1997 to be primarily a traditional printed edition. Although it is now based on a digital resource, the edition is nevertheless published as a series of large volumes of over 1000 pages each. This has some drawbacks: in particular additional time has to be spent in making the database structure and content produce the correct output for print publication; there are compromises in terms of copyright, particularly a two-year embargo on open publication of the edition online; and the digital publication is only made public in large tranches corresponding to the printed volumes, meaning that delays in a small part of a volume lead to all other editions in the volume being delayed not only in print but online as well.

However, the advantages of dual print and online publication have outweighed these problems. Printed publications, or digital publications which correspond to a printed one, tend to be better recognised by the many different research assessment and monitoring systems used by different countries and institutions. Print publication therefore can help support the careers and departments which contributed to the edition. Traditionally-minded scholars in our experience are more willing to use and reference the editions if they also appear in print – even if they primarily access the editions online. The printed volumes provide a potentially more permanent record of the project's work (effectively physically archived in libraries internationally), as well as a reference point should the digital resource at some later point require maintenance or reconstruction due to some major infrastructure failure.

The print publication is generated directly from the database. Each volume is exported in a single operation, with the database putting together the structure according to a defined table of contents. The content for each stanza is assembled by queries to the database which generate formatted (HTML)

text for the poetry, prose word order, translation, manuscript list, textual apparatus and notes. The resulting file is imported into a word processor template which adds headers and dividers. The remaining processes of checking and (in some cases) reformatting the content for print have taken a few hour's work for recent volumes.

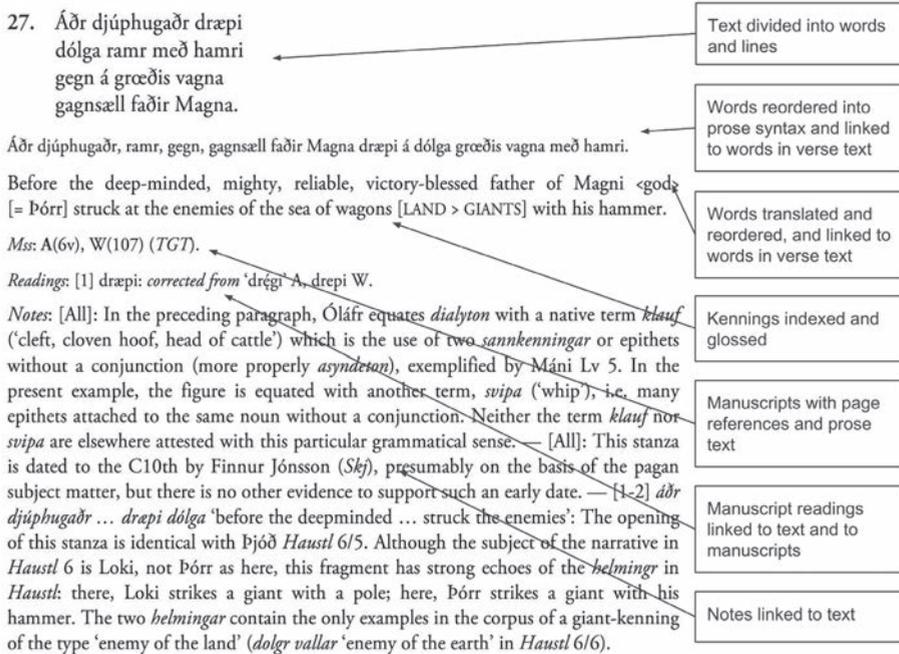


FIGURE 6. Print version of a stanza showing its relationship to the database

The resulting print output is shown in Figure 6¹⁰, with labels showing how each part relates to the underlying data in the database.

The Skaldic Project's application is thus able to produce both print and real-time electronic versions of the corpus using the same underlying data and digital methods.

5. Stage 3: A relational model for the text

The project's data is managed by a relational database management system (RDBMS, in this case MySQL / MariaDB). As noted above, such systems are

¹⁰Wills 2017, *loc. cit.*

highly scalable. In particular, there are very popular and well-supported tools for building applications using this software. A RDBMS has many advantages: data can be checked and values constrained in various consistent ways; links between all types of data are two way, making it easy and very fast to query, for example, all stanzas in a manuscript as well as all manuscripts for a stanza; the application can manage simultaneous access to updating all data, making it simple to manage and log multiple users' access and changes.

For the first published volume of the project (*Poetry on Christian Subjects*), the text of the edition, prose order and translation was stored in the database as Text Encoding Initiative (TEI) XML fragments using the tagset for verse texts. TEI has been the *de facto* standard for digital textual editing for the last two decades. It was originally an SGML implementation which was at first modeled on the structure of printed text editions and geared towards print publication for output. From version P4 it became an XML implementation. XML as a technology is now perhaps more commonly used for interchange and display of data, rather than as a data storage format and not normally a web application storage format. SGML/XML has two key features which make it very easy to apply to text encoding. Firstly, it allows for the complex mixing of markup and strings (in SGML terms «character data», that is, snippets of text), although the resulting data must still be part of the tree structure. Secondly, unlike databases, where the order of the data is not intrinsic to the data storage format, XML preserves the text intrinsically in the correct order. Given that the order of the text – its syntagmatic structure – is fundamental to all language, this is a requirement for any text editing¹¹. The mixing of text and markup is in some ways a legacy of the SGML text markup days, and a great many digital text projects have limited this feature in order to ease processing and minimise the potential variation in encoding practices for those using TEI. The Menota project and handbook, for example, defines a subset of TEI which is used to mark up texts down to the level of each word. This allows for the parallel processing of manuscript and linguistic information. There is very little mixing of elements and character data above the level of the word.

A good example of this method for a poetic corpus is the Codex Regius project, which has produced a very rich and detailed Menota/TEI edition of the majority of pre-1400, non-skaldic Norse poetry. The lack of a suitable web publication system for manuscript-based poetry encoded in this format,

¹¹ for more detailed discussion see Tarrin Wills, «Relational Data Modelling of Textual corpora: The Skaldic Project and its Extensions», in *Literary and Linguistic Computing* 30:2 (2015), pp. 294-313 <doi:10.1093/lc/ftt045>.

however, has meant that this rich resource has not been published on the web, despite being finalised in 2012. (The first poem, *Völuspá*, however, is available in the Menota archive¹².)

An alternative encoding type, often used in corpus linguistics, is «vertical text». In this format, each word occupies a line and effectively forms a data table, with other information such as word lemma and morphosyntactic information provided alongside the word in each line. The syntagmatic structure is preserved as the information is saved in a text file in the text order. Any changes to the text, however, normally require recompiling the whole text in the application, and the format is not always suitable for linking external information to the words in the text.

The project's database software, MySQL, introduced in version 4.1 a function for joining text from multiple grouped rows in a single query (GROUP_CONCAT, the equivalent of LISTAGG in Oracle). This function allowed a rethink of the way in which the text could be encoded within the stanza in the Skaldic Project. The database could potentially store words as individual rows in a separate table, and retrieve them, joined together in the correct order, to reconstruct a stanza's text. Having the words encoded in a database table would make it easier to link other information to the words, such as the notes and variants, but also kennings and dictionary headwords. For the first volume, some solutions were found for some of these problems, but they required the web application to check links with every update. In a relational database, the links can be defined so that they do not need further checking.

The order of the text could be encoded in a database format if each row contains information about the relative ordering of the row. Traditional stanzaic poetry is highly amenable to this approach because it is in its nature divided into smaller units (stanzas, lines) which are very often numbered in conventional editorial practices. This information is therefore already largely implicit and can be used in a database-based editing system to store and retrieve the text in the correct order, although additional information is required to order the words within lines.

6. Stage 4: Related resources

Using the relational model at all levels has made possible the creation of resources which link to the database.

¹²Link: <<http://clarino.uib.no/menota/catalogue>> [accessed on: 6/5/2019].

6.1. The Kenning Lexicon

As previously noted, the data entry process for the edition involves basic markup for kennings which are then processed and indexed by the application. Kennings are stored in a separate table with references to the words contained in the kennings and references to their analysis, in the form of text referents. Users are able to search the referents as they are entered, but referents are only partially normalised. For example, if the kenning is grammatically plural the referent might be MEN, and such kennings would not therefore be integrated with the singular (MAN) referents.

The goal is eventually to replace the current reference work on the kenning corpus, Meissner's *Die Kenningar der Skalden*¹³. While this can only be completed once all editions are finalised, the process of preparing a new kenning lexicon has started alongside the main project, headed by Jana Krüger (Kiel)¹⁴.

The first stage is to normalise the current referent list according to a semantic model (ontology). Figure 7 shows the table of kenning classes linked to the table of kennings¹⁵, so that all shield-kennings, for example, can be found through the resulting resource, regardless of whether their referent is labelled SHIELD or SHIELDS.

Some of the kenning classes contain hundreds of examples which are more usefully organised into particular patterns where the semantic values of the two parts of the kenning may be represented by various words or sub-kennings. In the example in Figure 7, shields are represented by the patterns sky/heaven of shield part and cloud of war/battle. These are established patterns that are represented by a number of kennings (e.g. those translated as «cloud of battle», «strife-clouds», «war-cloud»).

It should be noted that Figure 7 greatly simplifies the structure of the database, including by showing only one column for each data type. The black text represents data used by the Skaldic Project itself, with the grey data used by other projects, in particular *Lexicon Poeticum* (the lemmas and the variant words and compounds) and the *Kenning Lexicon* (kenning referent and word classes).

¹³ Meissner, *op. cit.*

¹⁴ The in-progress resource is available at <<https://skaldic.abdn.ac.uk/m.php?p=kenninglexicon>> [accessed on: 6/5/2019].

¹⁵ Based on the edition of Gade, *op. cit.*. Kenning class and lemma links were added by the present author.

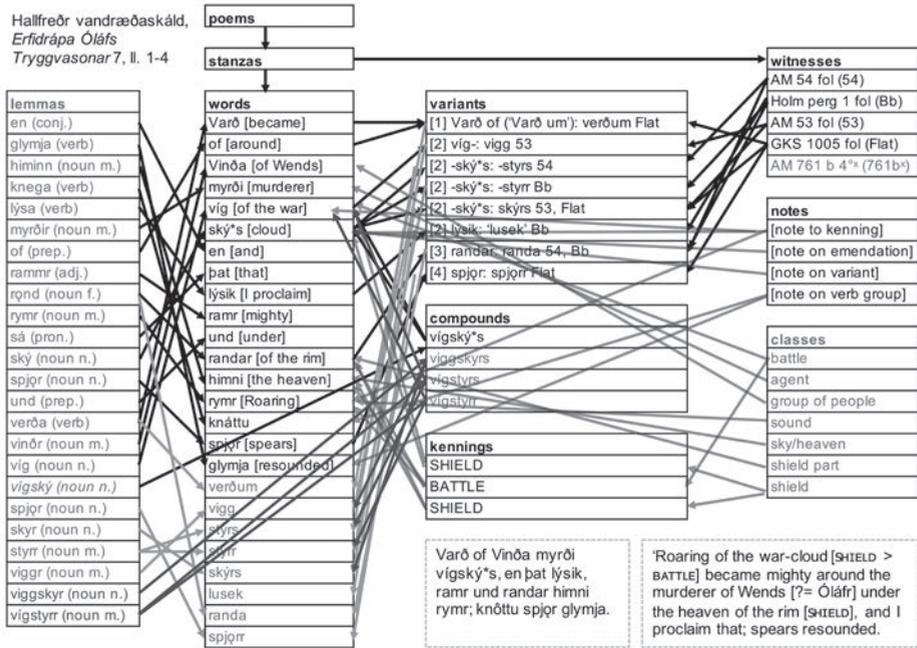


FIGURE 7. Extensions to the Skaldic Project's lexical data

In order to establish these kenning patterns, the *Kenning Lexicon* project classifies all substantive kenning parts according to the same table of classes (ontology), shown in Figure 7 as links between the classes and the words in the text. This process is assisted by prompting the lexicon editor with previous classifications of the same word lemma, which also creates a more consistent classification scheme. The classes are in the process of being arranged into a tree-structured ontology, so that users can search related classes. Eventually the lexicon will include prose descriptions of the classes and patterns contained within them.

6.2. The Lexicon Poeticum

The *Kenning Lexicon* relies on the linking of words in kennings to dictionary headwords (lemmas) which has been done under a related project, the *Lexicon Poeticum*. This project borrows its name from the previous dictionaries

of Old Norse poetry, namely that by Sveinbjörn Egilsson (1860¹⁶) and revised and rewritten by Finnur Jónsson (1913-1916 and 1931¹⁷).

The first stage of this project involves linking all words in the edition (around 150,000) to a dictionary headword list. This list was originally exported from ONP's word list, which included words only found in poetry as well as all words in prose. This list has subsequently been linked back to ONP's current wordlist. Compounds, which for metrical reasons almost never consist of more than two parts, are linked both as a whole to the main word list, with the individual parts in the main word list. This helps in the analysis of kennings which are represented by a compound, for example.

The lemmatising is performed through the web application. Each word's lemma is chosen by the lexicon editor, but the user is prompted by the application with likely choices based on the word's particular form and the frequency of linking to a particular lemma. This process has been assisted by importing large lemmatised prose corpora from Menota and Málföng¹⁸. The user goes through the text of the stanza in its prose word order selecting from the list of suggested lemmas in more than 90% of cases, or looking up a lemma using a search form in the other cases. In some instances the new edition's text includes a word not previously identified by other dictionaries. In particular, a great many compounds representing kennings have been added as new lemmas to the word list. There is also an integrated system for searching for updates from ONP's word list and importing them if needed. At the time of writing almost all the editions in the database have been lemmatised, although these have not been reviewed.

This process is relatively straightforward for the main text, where the editor has interpreted all words and there is normally only one word to each lemma. There is, however, a large amount of variation between manuscripts of a given stanza, and ideally the lexical variants present in all independent manuscripts will form part of the resulting *Lexicon Poeticum*.

The current textual apparatus, shown in Figure 7, represents very complex relationships between the words in the text and the manuscript witnesses. The variants are not put into the main word table as part of the process of editing the poetry. The variant might consist of one or more words in one or more manuscripts, or it may show that a word is missing in one or more

¹⁶ Sveinbjörn Egilsson, *Lexicon poeticum antiquae linguae septentrionalis*, Copenhagen, Societas Regia antiquariorum septentrionalium, 1860.

¹⁷ Finnur Jónsson, *Lexicon Poeticum*.

¹⁸ Eiríkur Rögnvaldsson and Sigrún Helgadóttir, «Morphosyntactic Tagging of Old Icelandic Texts and Its Use in Studying Syntactic Variation and Change», in *Language Technology for Cultural Heritage: Selected Papers from the LaTeCH Workshop Series*, ed. by Caroline Sporleder et al, Berlin, Springer, pp. 63-76.

manuscripts. The number of words in the variant may not be equal to the number of words of which it is a variant. The interpretation of variants is not linked in the database in the way that the main text is (as a translation).

In order to treat the variants as part of the lexicon, the relevant words need to be imported into the main word list (without affecting the normal display and editing of the edition) and linked to the lemma list. This process is simplified by the use of a form which extends the lemmatising form described above to deal with variants.

Figure 8 shows the web form for adding and lemmatising lexical variants for a stanza. This form goes through the variant apparatus and creates a row in the form for every combination of word in the variant with a word in the text. In most cases there is just one word in the variant and one word in the text. The form also allows for classifying the type of variant. In order to help the user, the main text's translation and lemma is shown. Combinations of variant words into new variant compounds are also available through the form.

When the form is submitted, the variant words are inserted into the main word table together with links to the main text's words and to the entry in the variant apparatus. This allows the database to reconstruct the material preservation of the variant, and to show variations on main text words in the lexicon.

The screenshot shows a web browser window with the URL skaldic.abdn.ac.uk/m.php?i=3830&f=425&d=1&v=. The page title is "Skaldic Poetry of the Scandinavians". The form is titled "Form for adding lexical variants from the edition to the Lexicon Poeticum".

The form is organized into rows for different stanzas. Each row contains a variant word, its translation, and a search field for the main text word. The form also includes options for adding variants and lemmatising.

Row 1: Variant word "göfug" (the noble), translation "göfugr (adj.) 'noble, glorious'", main text word "göfugr - 3m", ONP: göfugr - 3m. Options: "optional translation of variant", "add variant 'göfugr - 3m'", "Look up lemma or choose from list", "Note: word uncertain".

Row 2: Variant word "röðr", translation "röðr 'enraged'", main text word "1. röða (verb) - 444 'ride'", "2. röða (verb) - 4". Options: "optional translation of variant", "add variant 'röðr'", "Look up lemma or choose from list", "Note: word uncertain".

Row 3: Variant word "þeyg", translation "þeyg 'not then'", main text word "ONP: þeyg - 3". Options: "optional translation of variant", "add variant 'þeyg - 3'", "Look up lemma or choose from list", "Note: word uncertain".

Row 4: Variant word "strengja", translation "strengjar 'of the rope'", main text word "strengr (noun m.) - 120 'string, rope, ...'", "strengja (verb) - 94 'best, made'". Options: "optional translation of variant", "add variant 'strengja'", "Look up lemma or choose from list", "Note: word uncertain".

Row 5: Variant word "heim", translation "heimð 'ye'", main text word "heimr (noun m.) - 221 'home, abode, ...'", "heimr (adv.) - 123 'home, back'", "heimskyllr (noun n.) - 29", "heimr (pron.) x". Options: "optional translation of variant", "add variant 'heim'", "Look up lemma or choose from list", "Note: word uncertain".

Row 6: Compound "heimþegum 'retainers'", main text word "heimþeggi (noun m.)". Options: "add 'heimþegum' (J1, J2, 75a)", "Look up lemma", "add 'heimþegum' (325V1, 73a, 78a)", "Look up lemma", "add 'heimþegum' (J1, J2, 325V1, 75a, 73a, 78a)", "Look up lemma".

FIGURE 8. Form for adding lexical variants from the edition to the *Lexicon Poeticum*

This process is in its early stages, but nevertheless almost 6000 words representing lexical variants to the main editions have been entered and lemmatised. It is expected that around 50,000 lexical variants will be added to the 150,000 words of the main text of the corpus.

The screenshot displays the Lexicon Poeticum interface for the word 'ský'. The main entry shows 'ský (noun n.) [°-s; -] 'cloud...'. Below this, there are sections for 'ONP: ský sb. n.', 'semantic categories' (cloud), and 'citations (undefined)'. A search filter is set to 'Filter items...'. The 'variants' section lists several entries with their source editions and folio numbers, such as 'Skjunn* tree (of the cloud) — HSI Rst 1¹' and 'skýljóðs 'offener (of the cloud) — HSI Rst 20¹'. The 'kennings' section lists several entries with their source editions and folio numbers, such as 'skýljóðung gjalts skýja' and 'skýljóðungr veldis skýja'. The 'skaldic manuscripts' section is expanded to show a list of manuscripts with their folio numbers and dates, such as 'AM 325 VII 4* (325VII) (1)' and 'GKS 1009 fol (Mork) (3)'. The interface is designed to provide a comprehensive view of the word's usage and variants in Old Norse poetry.

FIGURE 9. *Lexicon Poeticum* entry for *ský* (cloud)

The information available for a particular word consequently extends through the links to the poetry, variants, manuscripts, kennings, compounds and more. Figure 9 shows the information available through the current *Lexicon Poeticum* interface for a particular word, including a link to the ONP, the semantic categories applied to the word in the *Kenning Lexicon*, and a full concordance of instances of the word in the skaldic corpus, including variants and manuscript text, if available. This view also links directly to any annotations on the word in the edition (the «i» button next to the second citation in this view). It also generates a full concordance of the word as it appears in any compounds in the corpus, as well as all kennings in which it appears. There is also a complete list of manuscripts where the word is found, based on the variant apparatus and manuscript links. Clicking on one of these gives a full concordance of the word in poetry in a given manuscript.

6.3 The Dictionary of Old Norse Prose

The *Lexicon Poeticum* provides an interface to the words of the project but it does not at this stage semantically structure the citations in the way a conventional dictionary does. Rather than independently editing the lexicon separately from the *Dictionary of Old Norse Prose* (ONP), the present author is developing ways in which the two dictionaries can work together with a single semantic tree for each word. While both projects are built on a RD-BMS (ONP uses Oracle), it is not feasible to integrate the two data structures. Instead the two projects link their data externally.

As the *Lexicon Poeticum* word list is largely linked to ONP's wordlist by use of its numeric keys, the two dictionaries can easily search linked information from each other. At the ONP end, this means that the dictionary can query the Skaldic Project in order to supplement its list of prose citations with the poetic citations found in the latter. For words not yet edited by ONP, this means that the user has access to the Skaldic Project's translations, for example, as well as the references to the manuscripts which contain the word only in poetry. Figure 10 shows the new ONP web interface with links to the skaldic corpus fetched from the skaldic database and integrated into the page.

The screenshot shows the ONP web interface for the word 'gumi'. The main content area displays the following text:

AM 757 a 4* (10r^o)
 Aflnægir seðr auflgu iod huggari þjóðar áttu
 þann er allðri fettiz ...uð skil- l nínngar bra...ðe.
 þvi er oss með hug huóssum hrein ...riost ok tru
 fósttar glóð letr guðdóms e... guma l kyn faudur
 skynja.

Aflnægir seðr öflgu
 jóð — huggari þjóðar,
 átt, þann er aldrí fettísk,
 auð — skilningar brauði,
 þvi er oss með hug hvóssum
 hrein brjóst ok trú fósttar;
 glóð letr guðdóms eðli
 guma kyn fódur skynja.

Aflnægir seðr jóð öflgu brauði skilningar — huggari
 þjóðar, átt auð, þann er aldrí fettísk — þvi er fósttar oss
 hrein brjóst með hvóssum hug ok trú; letr glóð kyn
 guma skynja guðdóms eðli fódur.

The strength-granter satisfies his children with the
 powerful bread of understanding — you, comforter of
 people, have the wealth that never diminishes — which
 fosters in us pure hearts [lit. breasts] along with keen
 thought and faith; it allows the cheerful race of men
 [MANKIND] to comprehend the divine nature of the
 Father.

Other corpora
 Translations from the Skaldic Project: men (59); of men (35); for men (9); people (7); man (7); to men (4); men's (3); the men (2); warriors (2); among men (1); Let men
 (1); the man (1); of a man (1); from men (1); a man (1); person (1); of warriors (1); gumar (1); gumar (1); them (1);

References on the right side of the interface include: GrgKof 236⁷⁷, Hrað 99⁷², SnE 136⁷⁷, Hem1005 15⁷⁴, Hem1005 15⁷⁴, and SnE 136⁷⁷.

FIGURE 10. Skaldic Project / *Lexicon Poeticum* references integrated into ONP

In the other direction, the skaldic lemmatising form now integrates links to ONP's semantic tree for each word, so that existing definitions can be used where possible for the particular sense represented by the word in the skaldic corpus. Eventually the two resources can be integrated into a single dictionary of Old Norse, containing citations from both poetry and prose and using the same semantic tree for each dictionary entry.

7. Conclusion

The three main challenges of the Skaldic Project as presented in this paper have been addressed by the development of a complex web application built on a relational database. Firstly, a relational data structure allows the project to maintain a digital link between all aspects of the resource, including the material evidence for the text and its interpretation. These links mean that the poetry, its lexicon, translation, interpretation, poets and prose contexts can all be understood in the context of their material preservation in the manuscript record and vice-versa. Secondly, the web application allows for multiple users to access and edit the data for the project in real time, allowing for asynchronous development of the project and related projects. In other words, members of the projects can see what is currently developing, link to resources as they are created, and have those connections automatically updated (or removed) as the original resources are modified and developed. And thirdly, the web application allows the complexities of the poetry to be encoded thanks to a series of interactive forms that provide intuitive ways of manipulating the data associated with the poetry, and in bring together existing information that can help the editor in making decisions about the analysis of the poetry.

The framework has been extended not only to incorporate related projects but also generalised to encompass structures that can be used by unrelated projects which bring together text and image, for example (e.g. *Pre-Christian Religions of the North: Sources Database*¹⁹, texts for teaching, and services for the Menota project). This means that new projects do not require such a system to be developed from scratch but instead can use the existing infrastructure with minimal additional resources.

Received: 6/05/2019

Accepted: 30/10/2019

¹⁹Link: <prechristianreligions.org> [accessed on: 6/5/2019].



THE SKALDIC PROJECT AND *LEXICON POETICUM*

ABSTRACT: This paper describes a digital project to edit the Old Norse poetic corpus known as skaldic poetry, composed between the ninth and fourteenth centuries. The Skaldic Project started in 1997 with the first editions published in 2007, and 75% of the corpus is now published in print and online. The long-term nature of the project, together with the complexities of the corpus and its manuscript and textual preservation, have meant that the digital resource has had a number of challenges to address. This article focuses on three of these challenges: the need to provide a large amount of material about the context of the poetry to contributors, including manuscript images; the need to provide an integrated editing and publication system capable of simultaneous updates by multiple editors and assistants; and instant digital access, online publication and a streamlined process for export for print publication. This article describes these processes and solutions, which encompass both human workflows as well as technical solutions. It demonstrates how the method adopted and data created has provided a foundation for further research projects based on the digital resource, including a lexicographic resource (*Lexicon Poeticum*) and an index of the complex diction of the corpus (*Kenning Lexicon*).

KEYWORDS: Digital Humanities, Skaldic poetry, Old Norse, Lexicography.

SKALDIC PROJECT Y *LEXICON POETICUM*

RESUMEN: Este artículo describe un proyecto digital de edición del corpus poético en nórdico antiguo, compuesto entre los siglos IX y XIV. El Proyecto Skaldic comenzó en 1997 con las primeras ediciones publicadas en 2007, y el 75% del corpus está ahora publicado tanto impreso como en línea. La larga duración de este proyecto, junto con las complejidades del corpus y su conservación manuscrita y textual, conllevan a que el recurso digital haya tenido que enfrentarse a numerosos retos. Este artículo se centra en tres de esos retos: la necesidad de poner a disposición de los colaboradores una importante cantidad de materiales sobre el contexto de la poesía, incluyendo las imágenes de los manuscritos; la necesidad de proporcionar un sistema de edición y de publicación integrado que permita que varios editores y asistentes puedan hacer actualizaciones simultáneas; y el acceso digital inmediato, la publicación en línea y un proceso optimizado de exportación para una publicación impresa. Esta contribución describe estos procedimientos y soluciones, que incluyen flujos de trabajo humanos así como soluciones técnicas. Se demuestra como el método adoptado y los datos creados proporcionan los cimientos para otros proyectos de investigación basados en el recurso electrónico, incluyendo un recurso lexicográfico (*Lexicon Poeticum*) y un índice del complejo estilo del corpus (*Kenning Lexicon*).

PALABRAS CLAVE: Humanidades Digitales, Poesía escáldica, Nórdico antiguo, Lexicografía.