Kodikologie und Paläographie im digitalen Zeitalter 4

Codicology and Palaeography in the Digital Age 4

herausgegeben von | edited by

Hannah Busch, Franz Fischer, Patrick Sahle

unter Mitarbeit von | in collaboration with

Bernhard Assmann, Philipp Hegel, Celia Krause

Digitale Parallelfassung der gedruckten Publikation zur Archivierung im Kölner Universitäts-Publikations-Server (KUPS). Stand 4. September 2017.

The Legendary Legacy: Crunching 600 Years of Saga Manuscript Data

Matthew Driscoll

Abstract

The research project "Stories for all time", which ran from 2011 to 2014, had as its aim to survey the entire transmission history of the *Fornaldarsögur Norðurlanda*, a group of Icelandic sagas often referred to in English as "mythical-heroic" or "legendary" sagas. Although the sagas themselves are thought to date from the 13th or 14th century, they are preserved mostly in post-medieval paper manuscripts. We set out therefore to locate and catalogue all the manuscripts containing texts of the 35 sagas which make up the corpus. In the end we found over 1000 manuscripts – containing nearly 2000 texts – the earliest from the beginning of the 14th century, the latest from the beginning of the 20th. About a quarter of these were not previously known to scholarship. We catalogued all of these manuscripts using a very restrictive subset of the TEI manuscript description module, which allows us to compare codicological and other features of the manuscripts in a way hitherto impossible. The article presents the schema and some of the results of our analysis of the encoded data.

Zusammenfassung

Im Rahmen des Forschungsprojekts »Stories for all time« wurde in den Jahren 2011-2014 die Überlieferungsgeschichte aller unter dem Namen *Fornaldarsögur Norður-landa* gefassten und als »mythisch-heroisch« oder »legendenhaft« bezeichneten isländischen Sagen untersucht. Ihre Entstehung wird für gewöhnlich in das 13. und 14. Jahrhundert datiert; überliefert sind sie gleichwohl vor allem in neuzeitlichen Papierhandschriften. Ziel des Projekts war es, sämtliche Textzeugen des 35 Sagen umfassenden Corpus aufzufinden und zu katalogisieren. 1000 Handschriften mit etwa 2000 Texten konnten identifiziert werden, die älteste vom frühen 14. Jahrhundert, die jüngste vom frühen 20. Jahrhundert. Ein Viertel aller Textzeugen waren der Forschung zuvor noch unbekannt. Die Handschriften wurden unter Verwendung eines sehr strikten TEI-Schemas katalogisiert, das einen bis dato nicht möglichen Vergleich kodikologischer und anderer Eigenschaften erlaubt. Dieser Artikel stellt sowohl das Schema selbst als auch die Ergebnisse einer Analyse der mit diesem Schema erfassten Daten vor.

The project "Stories for all time: The Icelandic fornaldarsögur", based at the University of Copenhagen, has as its chief focus the transmission history of the *Fornaldarsögur Norðurlanda* – literally "Stories of the ancient men of the northern lands" but generally known in English as Legendary or Mythical-heroic sagas – a group of Icelandic prose narratives dealing with the early history of mainland Scandinavia, before the unification of Norway under Haraldr *hárfagri* (fair-hair) and the settlement of Iceland in the late 9th century. Although many of them demonstrably have older roots, the sagas as we have them were first written down in the 14th century. They remained popular throughout the late medieval and early-modern period, even into the 18th and 19th centuries and the first decades of the 20th.[1]

The project's chief deliverable is an electronic catalogue of all the manuscripts in which *fornaldarsaga* texts are found, containing information on their format and layout, the other texts they preserve and when, where and by and/or for whom they were written. Ancillary to this is a fully searchable bibliography of editions, translations and secondary material pertaining to the *fornaldarsögur*. Both the manuscript catalogue and the bibliography were produced in TEI-conformant XML. Both are regularly updated and available on the project website.

So far, 817 manuscripts have been identified as containing *fornaldarsaga* texts; about a quarter of these were not previously known to scholarship.[2] Of these, 82 are composite manuscripts, i.e. are made up of parts (two or more) of originally separate manuscripts. If the parts are counted separately, the total number of manuscripts is 1049 (a typical Fornaldarsaga manuscript is shown in fig. 1).

Most are from Iceland, but some were written, generally by or for scholars, in Sweden or Denmark. And although most are in Icelandic, about 150 are, or contain alongside the Icelandic text, translations into Swedish, Danish, French or Latin. Only around a quarter of the manuscripts only contain *fornaldarsögur*; the rest contain material belonging to other genres, principally *riddarasögur* (chivalric romances, both translated and original) and *Íslendingasögur* (Icelandic family sagas), but all sorts of other things as well (see further below).

For each manuscript there is a catalogue record produced using a restrictive subset of the TEI P5 module for manuscript description.[3] Among other things, the number of elements available for use was greatly reduced, many elements and attributes which are optional in the TEI were made mandatory, and many attribute value lists were 'hard-wired' into the schema. This was done both to make data-input easier for the cataloguers and reduce the risk of error, and also to make the data more easily searchable. We have for the same reason deliberately tried to put as much information

---

[1] For a definition of the *fornaldarsögur*, see Driscoll 2003 and Driscoll 2009.
[2] This number will certainly increase as more manuscripts in private ownership are discovered and catalogued.
[3] The module for manuscript description is presented in chapter 10 of the *TEI Guidelines*.
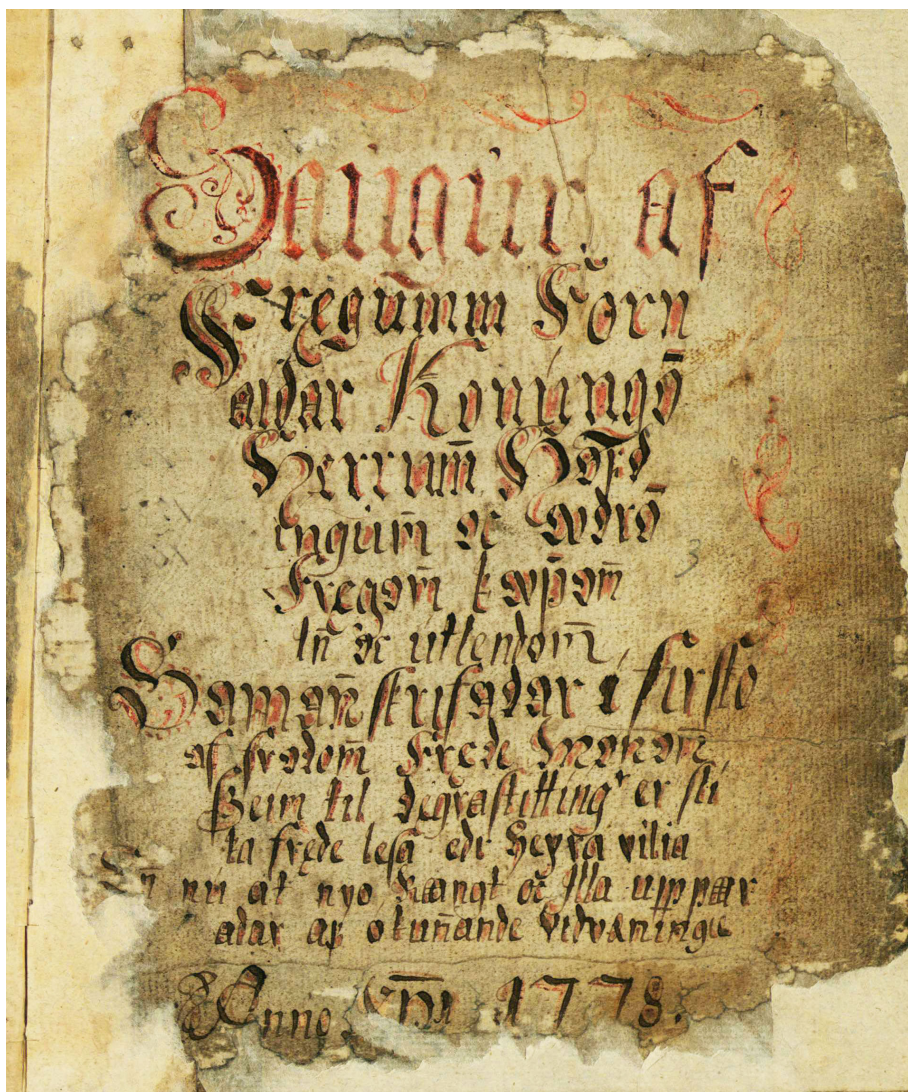
Figure 1: Reykjavík, Landsbókasafn Íslands – Háskólabókasafn, ÍB 165 4to, a large collection of Apostles'
vitæ and *fornaldarsögur* written in 1778 in Selárdalur by an unknown scribe, who identifies
himself only as "P. J.son". According to the title-page, shown here, the sagas were "Saman*n*
skrifadar í firstu af fródum*m* fręde møn*num*m, þeím til dęgrastittíng*ar* er slíka fręde lesa edr heyra
vilia, en nú at nyo rængt oc jlla upppær_adar af ókun*n*ande vidvaningie" (originally compiled by
wise men of learning for the enjoyment of those who wish to read or hear such lore, but now
once again badly and inaccurately scrawled by an ignorant amateur).

into the encoding as we can, avoiding wherever possible the use of natural language. So instead of indicating the language of the text by using the word "Icelandic" or "Swedish", for example, like this:

```
<textLang>Icelandic, with some Swedish</textLang>
```

we would put this information in attribute values, like this:

```
<textLang mainLang="is" otherLangs="sv"/>
```

The way this is displayed in the online database is then determined by the stylesheet. One added advantage of this method is that content can then be generated in any language, should one want to have the option of multiple interface languages.

As in standard TEI, the `<msDesc>` (manuscript description) element contains a description of a single identifiable manuscript. In our schema it must have the attributes `@xml:id`, which provides a unique identifier for the element, and `@xml:lang`, which indicates the language of the element content. The sub-elements of `<msDesc>` are `<msIdentifier>`, `<msContents>`, `<physDesc>`, `<history>` and `<additional>`, all of which should be present. Two further elements, `<msPart>` and `<msFrag>`, are also available within `<msDesc>`; the former is used for composite manuscripts, i.e. manuscripts comprising two or more originally distinct manuscript parts now kept together as a unit, and the latter for scattered manuscripts, i.e. manuscripts one or more parts of which have become separated from the original codex and may now be kept in different repositories.

Each of the child elements of `<msDesc>` contains a number of sub-elements, many of which have also been made mandatory. `<msContents>`, for example, must contain at least one `<msItem>` element, on which the attributes `@class` and `@n` must be present. Each `<msItem>`, in turn, must contain the elements `<locus>`, `<title>` and `<textLang>`, each with their required attributes.

```
<msContents>
 <msItem class="#fas" n="1">
 <locus from="1r:1" to="8v:17"/>
 <title type="uniform" ref="#snfdsv">
  Sögubrot af nokkrum fornkonungum í Dana ok Svía veldi
 </title>
 <textLang mainLang="is"/>
 </msItem>
</msContents>
```

One recurrent topic of debate within *fornaldarsaga* studies has been that of genre: to which extent can or should they be considered to represent a corpus?[4] Apart from the criteria of the time and geographical space in which the stories are set, do they share any features which may be said specifically to characterise them and distinguish them from other types of sagas. And, more importantly, is there any evidence in the

---

4   One of the more recent contributions to this debate is Quinn et al. 2006.

manuscripts themselves to suggest that those who copied and read them regarded them as constituting a genre?

In order better to address this question we have attempted to identify the other types of texts found in manuscripts alongside the *fornaldarsögur*, which is why the @class attribute has been made mandatory on all <msItem> elements. The possible values for @class are defined in a <taxonomy> element in the TEIheader. The different class designations are based on the indexing terms used by the National Library of Iceland, but simplified greatly.

We also place a lot of emphasis on the manuscripts' codicological features, and so many of the elements within <physDesc> (physical description) are also mandatory. Describing such features can be very time-consuming, however, and since we had a large number of manuscripts to get through in a relatively short period we developed a simple 'short cut' which allows us to provide basic information on the presence or absence of a feature or its relative level or extent without the necessity of going into any further detail. Flagging the presence of a feature in this way allows us to return to the manuscript later if need be. To this end the attribute @ana (analysis) is used on a number of elements.

To take one example: title pages, which were not a feature of medieval manuscripts but developed after the invention of moveable type, are often found in younger, post-medieval, manuscripts. In order simply to register their presence, and whether they appeared to be contemporary with the manuscript or added later, we require the attribute @ana on the element <titlePage>, with possible values "no", "contemporary", "later" and "unknown". No further content is required, but sub-elements such as <titlePart> can be used, or added later.

Other elements which can (or must) also use @ana in this way include <foliation>, <watermark>, <condition>, <decoDesc> and <additions>. In the latter two cases the possible values are "no", "low", "medium" and "high"; no other information need be supplied.

It could be argued that this is misuse of the @ana attribute, which is intended to provide a pointer "to one or more elements containing interpretations of the element on which the @ana attribute appears",[5] and that if a manuscript contains no watermarks, say, the best way to indicate this is by simply not using the <watermark> element. We disagree, however; the absence of an element does not necessarily indicate the absence of the feature that element is intended to be used to describe.[6]

---

[5] TEI Guidelines, section 17.2, "Global Attributes for Simple Analyses".
[6] This matter was discussed, though with no conclusion being reached, on the TEI listserv in February 2010; see <http://tei-l.970651.n3.nabble.com/Re-Indicating-the-presence-or-absence-of-a-feature-td2349886.html#a2349891>.

Within `<support>` we use `<num>` (number) to indicate the number of the leaves, and `<dimensions>` to indicate their size (a visualisation of leaf-size over time is given in fig. 2). As mentioned above we try to put as much information into the encoding as we can, as in the following example:

```
<supportDesc material="chart">
 <support>
  <num type="front-flyleaf" value="2"/>
  <num type="book-block" value="19"/>
  <num type="back-flyleaf" value="1"/>
  <dimensions type="leaf" unit="mm" scope="all">
   <height quantity="305"/>
   <width quantity="190"/>
  </dimensions>
 </support>
 <watermark ana="#yes"/>
 <foliation ana="#later #fol"/>
 <condition ana="#average"/>
</supportDesc>
```

The description of the layout is similar, again using `<num>` to indicate the number of words per line and `<dimensions>` to indicate the size of the written area:

```
<layoutDesc>
 <layout columns="1" writtenLines="28">
  <num type="wpl" atLeast="10" atMost="12"/>
  <dimensions type="written" unit="mm" scope="all">
   <height quantity="230"/>
   <width quantity="175"/>
  </dimensions>
 </layout>
</layoutDesc>
```

On the basis of this, one can easily work out the density of the text on the page. The proportion of the page taken up by the writing, the 'text-page ratio', can be determined by simply dividing the size of the written area (height × width) by the size of the leaf (height × width). In the case of the manuscript described here this would be 69.5%. A simple way of determining text density is to divide the size of the written area (height × width) by the number of words on the page (no. of lines × no. of words per line), which gives you the average amount of space (in mm²) taken up by a single word; in this case the figure would be 130.68. The smaller the number, the greater the text density. There are, of course, other ways to measure text density, for example by the average amount of space taken up by a single sign (whether letter, abbreviation or mark of punctuation), or the number of signs per unit of space, typically dm².[7] Both of these are more time-consuming than the method outlined here, which, despite its 'quick and dirty' nature, does give a good indication of the density of the text on the page which can be used as a basis for quantitative analysis.

---

[7]   See Maniaci 2002, esp. 101-120, and Gumbert 2010, 50-53. There are also software programs which can measure the relative amounts of ink and white space on a page and thus measure the density of the text; see e.g. Gurrado 2009.
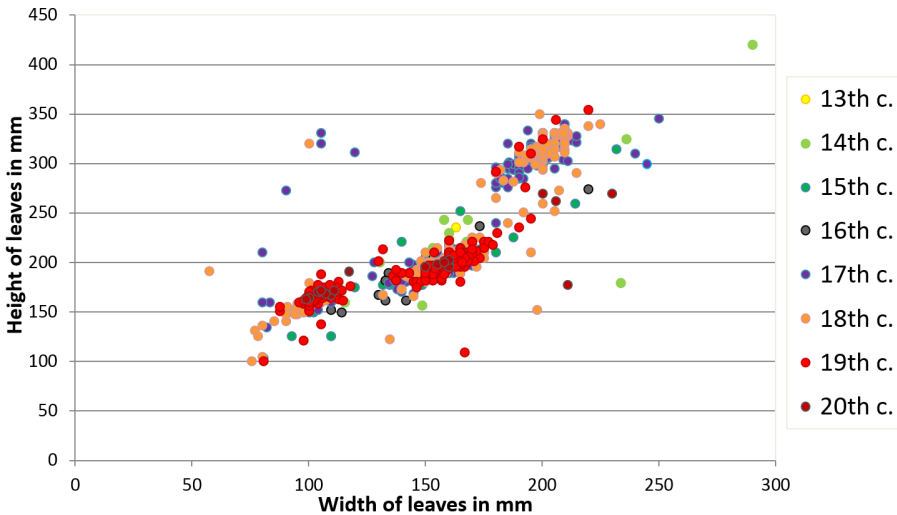
Figure 2: Manuscript leaf-size over time. Visualisation by Beeke Stegmann.

In addition to information on the manuscript's support and layout, the schema allows data on other characteristics to be supplied in a similarly data-intensive fashion. These include:

- number of hands in the manuscript and relative scope of each;
- the names of the scribes identified as corresponding to hands in the manuscript;
- the relative level of decoration of the manuscript;
- the relative level of additions (marginalia) made to the manuscript;
- the degree to which the binding is decorated and the contemporaneousness of the binding with the manuscript;
- the date and place of origin;
- the names of previous owners or other individuals known to have had a part in the manuscript's history.

We have also produced authority files for persons and places named in the manuscript descriptions using the <person> and <place> elements. Here, for example, is the <person> element for the scribe Jón Erlendsson:

```
<person sex="1" role="scribe" xml:id="JonErl001">
 <persName xml:lang="is">
  <forename sort="1">Jón</forename>
  <surname sort="2">Erlendsson</surname>
 </persName>
 <birth notBefore="1600" notAfter="1610">ca. 1605</birth>
 <death when="1672-08">August 1672</death>
```

```
<residence>
 <placeName>
  <settlement type="farm" ref="#VilVil01">Villingaholt</settlement>
 </placeName>
</residence>
<occupation xml:lang="en">Priest</occupation>
<bibl>
 <ref target="#IsAev">Íslenzkar æviskrár</ref>
 <biblScope unit="volume">III</biblScope>
 <biblScope unit="page" from="195" to="106"/>
</bibl>
</person>
```

This is then referenced in the individual manuscript records like this:

```
<name ref="#JonErl001" type="person">Jón Erlendsson</name>
```

Or within the <handDesc> element like this:

```
<handNote scope="major" scribeRef="#JonErl001" script="textualis">
 Written, apart from fol. 12, by Jón Erlendsson from Villingaholt
 in a clear, seventeenth-century Gothic book hand.
</handNote>
```

Within the `<person>` element itself, the @ref attribute on the `<settlement>` element, indicating Jón Erlendsson's place of residence, points to the corresponding `<place>` element in the place name authority file:

```
<place xml:id="VilVil01">
 <placeName xml:lang="is">
  <settlement type="farm">Villingaholt</settlement>
  <region type="parish" ref="#Villin01"/>
  <region type="county" ref="#&#xC1;rnes01"/>
  <region type="geog" ref="#Sunnle01"/>
  <country ref="#IS"/>
 </placeName>
 <location>
  <geo>63.883997 –20.750909</geo>
 </location>
</place>
```

Note that within the `<placeName>` element, all sub-elements, apart from `<settlement>`, which provides the name of the specific place in question, are pointers to other `<place>` elements in the authority.

Although this has not yet been implemented within our project, it would be possible on the basis of this mark-up to generate maps showing manuscript origin; this could help to reveal, among other things, whether certain sagas were more popular in certain places, and whether this changed over time.

All these different types of information can be collated, revealing things like changes in the distribution of texts over time or trends in format and layout. In the graph below, for example, manuscript format is collated with period of writing. The clusters show clearly the three principal formats, folio, quarto and octavo. It is interesting that the 19th-century manuscripts, which were mostly copied by ordinary people

for their own use, tended to be in the smaller formats of octavo and quarto, while those of the 17th and 18th centuries, which were often written by or for scholars and antiquarians, tended to be in folio.

The point of this highly restrictive schema was to allow for the encoding of basic codicological data on a moderately large number of manuscripts, based, wherever possible, on first hand examination of the manuscripts in question. As these manuscripts were held by some 29 repositories in 8 different countries, we were often forced to work at some speed, without the luxury of in-depth inspection. The idea was therefore to make data input as easy as possible, to reduce the possibility of error and to allow the presence or absence of particular features to be recorded, both for statistical purposes, and to flag items potentially of interest for further investigation. Although the resulting electronic catalogue is narrowly focused on one type of late medieval Icelandic narrative, we hope that our schema, or at least our approach, could be used as a model for similar investigations of virtually any body of documents.

Bibliography

Driscoll, Matthew. "Fornaldarsögur Norðurlanda: The stories that wouldn't die." In Jakobsson, Ármann, Annette Lassen, and Agneta Ney (eds). *Fornaldarsagornas struktur och ideologi*. Uppsala: Institionen för nordiske språk, 2003. 257-67.

Driscoll, Matthew. "A new edition of the *fornaldarsögur Norðurlanda*: Some basic questions." In Bampi, Massimiliano and Fulvio Ferrari (eds). *On editing Old Scandinavian texts: Problems and perspectives*. Trento: Università degli studi di Trento, 2009. 71-84.

Gumbert, J. Peter. *Words for codices: An English codicological terminology*. Lopik: [s.n.], 2010.

Gurrado, Maria. "'Graphoskop', uno strumento informatico per l'analisi paleographica quantitative." In Rehbein, Malte, Patrick Sahle, and Torsten Schaßan. *Kodikologie und Paläographie im digitalen Zeitalter 1 — Codicology and Palaeography in the Digital Age 1*. Norderstedt: BoD, 2009. 251-59.

Fornaldarsögur Norðurlanda: *Stories for all time*. 2011-2014. <http://fasnl.ku.dk>.

Maniaci, Marilena. *Archeologia del manoscritto: Metodi, problem, bibliografia recente*. Roma: Viella, 2002.

TEI Consortium (ed.). "Manuscript Description." In *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Version 3.0.0. Last updated on 29th March 2016. <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/MS.html>.

Quinn, Judy et al. "Interrogating genre in the *Fornaldarsögur*: Round-table discussion." *Viking and Medieval Scandinavia* 2 (2006). 276-96.