M. J. Driscoll
(Arnamagnæan Institute, Copenhagen)

THE VIEW FROM THE NORTH:
SOME SCANDINAVIAN DIGITISATION PROJECTS

Abstract. This paper presents a personal view of a number of Scandinavian projects which have sought to digitize medieval manuscript materials, focusing on their goals, the methods they employed, and how successful they were, and, where not wholly successful, trying to assess what went wrong or how they might have been better.

SagaNet

In 1997 The National and University Library of Iceland and the Fiske Icelandic Collection at Cornell University, with the association of the Árni Magnússon Institute in Iceland, began a cooperative project, with funding principally from the Andrew W. Mellon Foundation, known as SagaNet (http://sagnanet.is/ or http://saga.library.cornell.edu/).

The National and University Library of Iceland (Landsbókasafn Íslands-Háskólabókasafn) houses the largest collection of Icelandic manuscripts and papers extant, some 15,000 items, almost entirely post-medieval. They also have a large collection of printed books, having been Iceland's deposit library since the late 19th century. The Fiske Icelandic Collection at Cornell University Library is based on a collection of 8600 titles bequeathed in 1905 by Willard Fiske, a noted linguist and teacher who began collecting Icelandic books as a student in Sweden in the 1850s. The collection now comprises some 38,000 volumes, including some rare early printed books of which even the National Library of Iceland does not have copy.

The early plans for collaboration between these two collections, which between them have virtually every publication printed in Iceland or in Icelandic elsewhere, focused not unnaturally on these printed books. The Fiske Collection had already begun microfilming its collection, and the National Library was interested in doing the same. When approached, the Mellon Foundation suggested they think about digitization instead of microfilming, and thought they should concentrate instead on manuscripts, in particular manuscripts of the so-called 'Sagas of Icelanders' (Íslendingasögur), the best-known and most original of the several medieval Icelandic saga-genres. The problem with this was that, although the collection of manuscripts at the National Library is large (by Icelandic standards), the manuscripts are chiefly paper, and regarded by scholars as of only secondary importance, being either post-medieval copies of medieval works, or examples of younger, 'spurious' works (I hasten to add that I do not approve of this division between 'real' and 'spurious' sagas, but it is impossible to deny its existence in

the scholarly community generally). The Árni Magnússon Institute (Stofnun Árna Magnússonar), which houses those manuscripts from the Arnamagnæan Collection in Copenhagen which have been returned to Iceland (discussed further below), including most of the important medieval vellum manuscripts of the Íslendingasögur, was then brought into the picture. A new proposal was made to the Mellon Foundation in January 1997 to digitize a 'collection of record', viz. manuscripts and printed materials, including both editions and translations as well as secondary works, of and pertaining to the 'Sagas of Icelanders' and derived works. Shortly thereafter Mellon approved a grant of 600,000 USD, for which matching funds were secured in Iceland, and the project began in July 1997.

For image capture it was decided to use a Kontron Eyelike Digital Camera from Jenoptik in Germany, 6000 x 6000/24-bit colour, which would allow for resolutions of 300 dpi for larger formats (50 x 50 cm) and 600 dpi for smaller formats (25 x 25 cm), although in the end these too were captured at 300 dpi, allowing two pages (one opening) to be captured in a single shot in order to save time and increase production (indeed, one of the reasons for choosing the Kontron was the relatively short scanning time of about 45 seconds, which of course does not include time required for focusing, setting up the book in its cradle and so on). From 1999 the manuscripts at the Árni Magnússon Institute were photographed at 600 dpi using a Dicomed 7000 Digital Camera. All images were captured and stored as TIFFs, but distributed as JPEGs (thumbnails ca. 4 Kb; 120 dpi, 100–150 Kb). The original plan was to digitize 380,000 manuscript pages from the National Library and 40,000 pages from the Árni Magnússon Institute (altogether about 1000 manuscripts), scanned from the originals, and 145,000 printed pages (ca. 450 titles) from the Fiske Collection, and scanned principally from microfilm. This was always optimistic. Given that the project period was three years, the 380,000 pages at the National Library alone would have to be captured at a rate of 88 per hour, or roughly one page-image every 40 seconds, and this is assuming everything would be up and running on day one. As it happened, there were various problems with the Kontron both in the initial phase, which meant that actual digitization did not begin until March 1998, and for a period of six months during 2000, when the camera was out of action completely. Even so, by the end of the project period, 30 June 2001, some 200,000 manuscript pages from 850 manuscripts had been digitized, still an impressive feat. Work has continued after the end of the project, and there are now images from some 2,200 manuscripts available on SagaNet.

The quality of the images, unfortunately, is not always of the best, owing chiefly to the demand for speed; little time was allotted to setting-up, and post-processing was kept to an absolute minimum, which means that many of the images are, if not out of focus, then at least not completely in focus either, and the colour balance (notoriously difficult with manuscripts anyway) is often clearly wrong. But the images are certainly usable; indeed, I use them all the time.

The manuscripts were catalogued using MARC, which is not really suitable and presented many problems for the cataloguers in the early stages. To be fair there was little alternative at the time, as the MASTER project (an EU-funded project whose goal was to define and implement a general purpose SGML/XML-based standard for the description of manuscript materials) had not yet got under way. And the result, it has to be said, is, if not ideal, certainly not unworkable. One could hope, however, that in future the records will be converted to a more suitable format.

On the whole, SagaNet has proved a useful tool for scholars and the general public. One criticism which might be leveled against it is that once it became clear that the emphasis was to be on medieval Icelandic saga manuscripts more effort might have been made to involve the other institutions in possession of such manuscripts: the Royal Library (Kungliga Biblioteket) in Stockholm, the Royal Library (Det Kongelige Bibliotek) in Copenhagen, and the Arnamagnæan Institute (Det Arnamagnæanske Institut) in Copenhagen; with these institutions on board one could truly have spoken of a 'collection of record'. The reason here too appears to be that there was simply not time, as the National Library and the Fiske Collection were under pressure to complete their negotiations with the Mellon Foundation.

Most of us are dependent on external funding in order to pursue digitization (and other) projects. Unfortunately, this means that we generally have to tailor our projects to the requirements of the funding agencies or particular call. The Mellon Foundation indicated that it would be willing to fund some kind of project involving the National Library and the Fiske Collection, just not the one the applicants had had in mind. The result was a series of compromises; in order to secure the funding the nature of the project was completely altered in a very short time (since Mellon obviously wouldn't wait), and once the funding was secured the main job was to make sure it was spent, which meant that a very large amount of material had to be digitized and catalogued in a very short time, no matter what. One could, of course, say that whatever its shortcomings, SagaNet is 'better than nothing', but to do so would be both unfair and untrue — it is a million times better than nothing, and Þorsteinn Hallgrímsson and his team deserve full marks for what they've done. But it is frustrating to think how much better it could have been, had the participants had even a little more time.

Lund University Library:
The St Laurentius Digital Manuscript Library

The collection of medieval manuscripts at Lund University Library (Lunds universitetsbibliotek) in Sweden is small by international standards, numbering only 67 items, but among them several extremely important manuscripts, such as the so-called Necrologium Lundense, thought to be one of the two earliest Scandinavian manuscripts extant. About a third of the items are Scandinavian, principally Danish, while the remainders are French, Italian, German, Dutch, Flemish, Russian and Byzantine. There are also several thousand medieval fragments.

In 1999 Lund University Library received funding for a major digitization project, Medieval Manuscripts at Lund University Library [http://laurentius.lub.lu.se/], which sought to digitize and catalogue all these medieval manuscripts and to make the integrated database (images and catalogue entries) accessible on-line. The funding was generous: some four million Swedish crowns (nearly a half a million Euros), especially considering the size of the collection. The original application was for far less, and covered only the cataloguing work. When approached, the Bank of Sweden's Tercentenary Foundation (Riksbankens Jubileumsfond) indicated its interest in funding such a project but said it was not innovative enough and that they would rather see something that included digitization and the construction of a search engine — even if such a project were far more expensive. The application was then rewritten and the money secured. In fact, the climate in Sweden was for a long time very much in favour

of digitization (cf. the huge Waller project at Uppsala University Library [http://www.ub.uu.se/arv/waller/eprojekt.cfm], as well as the Vadstena projects mentioned below), something that came as a blessing to many manuscript departments and archives. Unfortunately this period seems now to be coming to an end, and it is again becoming difficult to get funding for such projects.

The person behind the project, Eva Nylander, a classicist who had spent eight years cataloguing manuscripts in the Vatican Library, set about her task with all the rigour one would expect of someone with her background.

For image capture the best equipment available was secured from Sinar of Switzerland (details are available on the project's website [http://laurentius.lub.lu.se/info/Engelska/data_imag.shtml]), and the results, by the photographer Bengt Melliander, are exceptionally good. The manuscripts were catalogued in XML according to the MASTER standard (the present writer acting as consultant), with various experts being brought in to deal with particular areas, either linguistic or thematic (e.g. a medieval art historian wrote the description of the illumination); a set of stylesheets transforms these catalogue records to HTML for delivery over the web, and PDF or Postscript files for printing.

The search engine for the project was created by Sigfrid Lundberg, who has described his methodology in detail in the paper 'St Laurentius Digital Manuscript Library: An excursion along the border between resource discovery and resource description', available in PDF format [http://www.hum.ku.dk/ami/mjd/laurentius.lub.lu.se/search/presentation/laurentius.pdf].

All in all, the Lund project has been the most successful of those reviewed here, although this is obviously in part due to its relatively small size. Still, it shows what can be accomplished when one has adequate funding and takes the time to ensure that the highest possible standards are maintained. Sadly, few projects are so fortunate.

The Medieval Nordic Text Archive (Menota)

The Medieval Nordic Text Archive (Menota) hopes to collect transcriptions of medieval manuscript materials encoded in accordance with the Text Encoding Initiative's Guidelines for Electronic Text Encoding and Interchange [http://www.tei-c.org/]. Menota grew out of another project, begun in 1994 when a group of textual scholars from the two Arnamagnæan Institutes in Copenhagen and Reykjavík and the Universities of Gothenburg, Oslo and Bergen formed a 'Network for the electronic processing of medieval Nordic manuscripts' (in Danish Netværk til elektronisk behandling af nordiske middelalderhåndskrifter) with the idea of investigating the various aspects of digitization and their application to the field of Old Norse/Icelandic studies. A grant from the Nordic Academy for Advanced Study (Nordisk Forskerutdanningsakademi — NorFA) facilitated a series of meetings in the period 1995–98, but there was no concrete result apart from a determination to continue. It had become clear in the course of our meetings that the TEI Guidelines provided a means for dealing with most aspects of marking-up medieval texts, but a number of problems encountered in Old Norse/Icelandic materials remained to be addressed, the very extensive use of abbreviations, for example, as well as the presence of a large number of 'exotic' (i.e., non-English) characters. It was decided that effort should be concentrated on the production of a handbook for the encoding of Old Norse/Icelandic primary

sources using TEI-conformant SGML/XML. The present writer undertook the job of general editor. This handbook was unfortunately never completed, chiefly owing to a disagreement between the general editor and the other members of the editorial committee regarding various aspects of mark-up, among them the treatment of abbreviations, although an article detailing some of the suggestions to be made in the handbook, entitled 'Encoding Old Norse/Icelandic primary sources using TEI-conformant SGML', was published in *Literary and Linguistic Computing*, Vol. 15, No. 1 (2000), pp. 81–91.

In 2000–2001, a team of scholars, not including the present writer, spent a year at the Centre for Advanced Study (Senter for høyere studier) in Oslo, discussing issues pertaining to the encoding of medieval Scandinavian manuscripts. The result of this year was a set of guidelines (in Norwegian and Swedish), Håndbok for koding av nordiske middelaldertekster, available on the web at http://www.hit.uib.no/menota/handbok/innhold.html.

Menota was formally established at a meeting in Oslo on 10 September 2001. At present, twelve institutions are represented, including all the major manuscript-holding institutions in Scandinavia. The board consists of one member from each of the Scandinavian countries (Denmark, Norway, Sweden and Iceland — there is currently no representative from Finland, although there is no reason why there should not be). It was decided at this first meeting that there should be a single standard to which texts in the archive should conform, in order to ensure compatibility. The archive would, in other words, more closely resemble something like the British National Corpus, which contains 100 million words of current British English, written and spoken, marked up in TEI-conformant SGML, than the Oxford Text Archive, which, while 'strongly advocat[ing] the use of TEI-conformant SGML', accepts texts in a wide variety of formats. The major task facing Menota since its founding has been the production of a set of guidelines, incorporating work done in the original 'Network' with the results of the Oslo project. Version 1 of the Menota handbook was released in May of this year (2003) and is available at http://www.hit.uib.no/menota/guidelines/index.html.

Initially, the idea behind the handbook was to produce something simpler and more 'user-friendly' than the TEI Guidelines, for example by recommending one among the several possibilities offered by the Guidelines, or, in a few cases, proposing alternative methods of encoding where those offered seemed unnecessarily complex. In the end, however, our handbook has grown to be greater and more complex than those portions of the TEI Guidelines it set out to simplify. I am concerned too that the minimum requirements for tagging may be beyond the capabilities of many of the institutions/individuals who would be in a position to supply texts. I think what is needed is a kind of 'Menota-Lite', which would present in a clear, concise way a relatively low-level standard of mark-up which could be the basis for further mark-up as desired or required. At the time of writing the Medieval Nordic Text Archive does not actually contain any texts. Sometimes I fear that given the complexity of the current recommendations for mark-up, it may remain empty for some time to come.

Despite these reservations, there is reason to be optimistic. Professor Odd Einar Haugen of the University of Bergen in Norway, who has done more than anyone else to make Menota a reality, received funding from the Research Council of Norway (Norges forskningsråd) for a three-year project, beginning in January 2003. This project is known as MENOTA TVB [http://helmer.aksis.uib.no/menota/TVB/index.html] ('Tekst-og verktøybase', i.e. 'text and tools base'), the aim of which is to develop further the

standard and produce tools for data input, search and display. And last year Karl G. Johansson, formerly of Växjö University in Sweden but now attached to the Centre for Viking and Medieval Studies (Senter for studier i vikingtid og nordisk middelalder) at the University of Oslo and one of Menota's 'founding fathers', secured a very large grant, altogether some twelve and a half million Swedish crowns, for two digitization projects relating to manuscripts produced at the Abbey of Vadstena in Sweden, motherhouse of the Brigittine Order. The texts generated by this project will be encoded according to the Menota standard, and will be made available through the archive.

The virtual reunification
of the Arnamagnæan Manuscript Collection

The Arnamagnæan Collection, named after the Icelandic scholar and antiquarian Árni Magnússon (1663–1730), secretary of the Danish Royal Archives and professor of Danish Antiquities at the University of Copenhagen, comprises some 3000 items. The majority of these are Icelandic, but the collection also contains important Norwegian, Danish and Swedish manuscripts, along with approximately 100 of continental provenance. Bequeathed by Árni Magnússon to the University of Copenhagen upon his death, the collection also represents a milestone in the debate on the return of cultural treasures: negotiations between the governments of Denmark and Iceland (until 1944 a part of the Danish realm) resulted in the transfer of the majority of the Icelandic manuscripts in the collection (just over half of the collection as a whole) to Iceland, a process begun in 1971, immediately following the ratification of the treaty between the two countries, and finally completed in June 1997. The manuscripts retain their original shelf marks, and the two halves are to be seen as constituting a single collection, albeit one housed in two different countries.

In 1995 I joined staff of the Arnamagnæan Institute in Copenhagen, a teaching and research institute within the Faculty of Humanities of the University of Copenhagen, as the process of transfer was coming to an end. I had received my training and cut my philological teeth at its sister institute, the Árni Magnússon Institute in Reykjavík, similarly a department of the University of Iceland, to which the manuscripts from Copenhagen had been transferred, and could therefore see the matter from both sides as it were. It was at about the same time that I first heard of digitization, having become involved in the Network mentioned previously, and it wasn't long before a thought struck me — it didn't matter where the manuscripts resided physically, as the collection could be reunited virtually through digital technology. An idea was born: the virtual reunification of the Arnamagnæan manuscript collection. I discussed the matter with my former colleagues in Reykjavík, in particular Sigurgeir Steingrímsson, assistant director of the Institute, and my new ones in Copenhagen, and there was universal agreement that it was indeed a wonderful idea. Sigurgeir and I began looking into file formats and compression ratios, read the specifications of dozens of digital cameras, became involved in the MASTER project and the TEI, and then, at a ceremony in Reykjavík in 1997 when the last two manuscripts were handed over, I gave a paper in which the virtual reunification project was presented to the world. Following that I gave similar papers, in a variety of languages, at a number of conferences, some of which were even published ('The virtual reunification of the Arnamagnæan Manuscript Collection', The Digital Demotic: A selection of papers from Digital

Resources in the Humanities 1997, ed. Lou Burnard et al. (London, 1998), pp. 55–64, or 'The digitization of manuscripts in the Arnamagnæan Collection', Care and conservation of manuscripts 4: Proceedings of the fourth international seminar held at the University of Copenhagen 13–14 October 1997 (Copenhagen, 1999), pp. 8–11). The idea was simple, if ambitious. We would produce a new electronic catalogue, with links to digitized images and, eventually, transcriptions, of the entire collection, all of which would be made available freely over the Web for use by scholars, in schools, and by the general public. After the initial euphoria, it began to dawn on us just how ambitious this was, as we began assessing the sheer size of the project — although not huge (compared to, say, the Bodleian or Vatican libraries), the Arnamagnæan Collection is fairly big. Just cataloguing the 3000 manuscript items would require a good few man-years, probably somewhere between eight and ten, image capture and post-processing presumably something similar, and what it would take to produce marked-up transcriptions — reliable transcriptions, i.e. proof-read and checked — of all the manuscripts in the collection is anybody's guess, but on the basis of the production rate of individual editions certainly in the hundreds. So the first problem was obviously going to be money — who was going to pay for all this? The next problem was, even assuming that the funds could be got, who was going to do the work? Old Norse is a small field, and there simply aren't that many people available with the necessary training, certainly not in Copenhagen. Still, I reasoned, if you could get the money you could probably find graduate students to do the work, but the checking would have to be done by at least reasonably senior scholars, most of whom have far too much on the their plates as it is. Another question which we began to ask ourselves was, even if all this could be done, did the world really need 600 dpi images and fully-marked up transcriptions of every manuscript in the collection, given that there were a fair number of manuscripts in the collection which had limited scholarly interest — which, to repeat what I said above, in no way implies that they are entirely without interest. So, we thought, let's limit ourselves on the one hand to manuscripts of 'primary' importance (perhaps 500–600), and on the other to catalogue records and digital images, i.e. forget the marked-up transcriptions.

In 1998 we applied for, and received, a small grant from the European Union's Raphael Programme for a series of workshops, organized by the two Arnamagnæan Institutes in conjunction with Oxford University's Libraries Automation Service and the Bodleian Library, the aim of which was to explore the various possibilities offered by digital technology. Three three-day workshops were held, one at each of the sites (Oxford, Reykjavík, Copenhagen), focusing, respectively, on electronic cataloguing, text mark-up and image capture. Altogether some fifty people from half-a-dozen countries participated.

Major funding for the project proved elusive, however. The EU, we assumed, was unlikely to be interested in something so peripheral (and only Denmark is a member). Nordic funding agencies generally require the participation of at least three Nordic countries; we were only two, and didn't really want to involve any others. Most national funding agencies, on the other hand, want projects the focus of which is limited to one country. The Danish government had just finished paying for the transfer of half the collection to Iceland (the manuscripts were conserved and photographed before transfer), and were unlikely, we thought, to want to finance its virtual return. The Icelanders were somewhat successful, receiving funding both for the purchase of a digital camera (through its association with SagaNet) and the hiring of extra staff to

undertake the cataloguing, but little extra funding was received by the institute in Copenhagen, for the simple reason that no real effort was ever made to obtain any. An important factor here, it has to be said, has been the, if not outright opposition to the project, at least a decided lack of enthusiasm for it on the part of the Institute's leadership. It took me far longer than it should have to realize that whatever people might say nothing was actually going to be done — apart from what I was able to do myself, which was been limited.

The only area in which some progress has been made is in cataloguing. Preliminary work on a new electronic catalogue of the collection, based on Kristian Kålund's Katalog over Den Arnamagnæanske Håndskriftsamling (Copenhagen, 1888–1894) but supplemented by more recent scholarship, was begun at both institutes as part of the MASTER project. During the project period itself some 500 records, the majority of them minimal, were produced in Copenhagen. It was decided to concentrate on the medieval manuscripts in the collection, although post-medieval manuscripts of special importance (for example copies of medieval vellums now lost) were also described. Since the end of the period very minimal records — comprising little more than shelf mark, date and place of origin and an identification of the contents — were made for the remainder of the collection, but little more than that has been done owing to lack of funding. In Iceland basic cataloguing began in the year 2000. It was decided to include all information regarding each manuscript from the printed catalogue, but, in the initial stages, no more than that. Two full-time employees carried out most of this work. By the autumn of 2002, approximately half of the manuscripts in the Icelandic collection had been catalogued in this manner. In October 2002 'complete cataloguing' was begun; that is, the manuscripts are examined, and their contents and appearance described in detail. At the present time about 100 manuscripts have been catalogued in this way. The cataloguers in Iceland, in cooperation with the computer company Íslenska verkfræðistofan, have used semantic web technology on some of the texts that were catalogued completely. A part of this project is exhibited on the Institute's website [http://www.am.hi.is/], under the title 'Með Flateyjarbók í farteskinu'. Apart from these, none of the catalogue records produced at either institution has as yet been made publicly available, although we hope to have a demo up and running by the end of the year.

So, although not completely dead, the 'virtual reunification project' will never be what it could have been: its moment has passed. Perhaps things would have been different, if, five years ago, we had involved other institutions — as was mentioned, there are important collections of Icelandic manuscripts in Stockholm and Uppsala — but while a digital library of all Icelandic manuscripts everywhere is obviously very much a desideratum, we felt that as far as this project was concerned it was important to take the Arnamagnæan collection as a whole, warts and all, as it were (something, in fact, like what they went on to do in Lund, only on a larger scale), an idea the beautiful simplicity of which I confess I still find compelling. And I think that if we had just been more determined in our search, funding could probably have been obtained for such a project. This is hardly the case now: funding agencies are simply not willing to put money into 'content creation'. Looking at the experience of some other projects, perhaps it's just as well. As my old father (a middle-distance runner in his youth) used to say: 'Slow and steady wins the race.' It would have been nice in 1997, just as the last two manuscripts were being handed over, to flip the switch, reverse the engines, and start the process of reunification, but although that didn't happen I'd still like to think

that we'll get there in the end. It took 25 years to divide the Arnamagnæan collection in two, and there is no real reason, I suppose, why it should take any less time to put it back together again.

I am increasingly struck, I find, by the role played by fortuity in most areas of human endeavor — I say 'most' because it seems fairly obvious that although there may be some element of arbitrariness involved, one doesn't become a successful footballer, say, or violinist, without having a fair degree of talent, on the one hand, and training on the other. One has only to look around, however, to see that it is possible to have a reasonably successful academic career, for example, without very much of either.

There can be few areas in which arbitrariness is as evident as in the funding of cultural heritage projects. What projects get proposed in the first place can in and of itself be fairly arbitrary, and is dependant on a wide variety of factors, not least the nature of the calls at any given time, but even though one sometimes suspects that a proposed project has been thought up in order to fit a call, one hopes at least that the majority of the projects for which applications for funding are made are viewed by the applicants as worth doing, and that the applicants view themselves as the people most qualified to do them. What projects actually do get funding, and how much funding they get, seems, on the other hand, to be almost entirely random.

This is probably inevitable, and perhaps not evens such a bad thing: randomness is, after all, a governing principle in nature. Another, associated with a certain Charles Darwin, decrees that those individuals best outfitted for survival will survive. The same, I suspect, is true of digitization projects.

M. J. Driscoll
Arnamagnæan Institute, Copenhagen
E-mail: mjd@hum.ku.dk