

FRAGMENTA MEMBRANEA PROJECT
2009–2012
Final Report

Juha Hakala, Jesse Keskiaho, Jukka Kervinen, Heidi Törrönen, Marleena Vihakara

Kansalliskirjasto, Helsinki 2014

ISBN 978-951-51-0226-3 (pdf)

DESCRIPTION PAGE

Publisher	The National Library of Finland
Date of publication	31.1.2014
Authors	Juha Hakala, Jesse Keskiäho, Jukka Kervinen, Heidi Törrönen, Marleena Vihakara
Name of publication	Fragmenta Membranea Project 2009 – 2012 Final Report
Parts	
Name and number of series	Reports and Studies 1 / 2014
ISSN	2242-8119
ISBN	978-951-51-0226-3
URL	http://urn.fi/URN:ISBN:978-952-10-9312-8
Number of pages	41
Language	English
Keywords	manuscripts, medieval, digitization information storage and retrieval systems Finland
Abstract	<p>The National Library of Finland's Fragmenta membranea project was launched in 2009 and concluded in March of 2012. Its objective was to describe, conserve, and digitise all medieval manuscript fragments in the Library's collection and generate an information system facilitating the research and other utilisation of this material.</p> <p>As a result of the project, the entire collection can be accessed online free of charge at the address: http://fragmenta.kansalliskirjasto.fi. The online use of the manuscript fragments has facilitated the intensified use of the fragments as well as the protection of the original materials. Without digitisation, these ostensibly mutually contradictory objectives would have been impossible to achieve.</p> <p>The Helsingin Sanomat Foundation, the National Library of Finland, and the Ministry of Education and Culture funded the project.</p>
Futher information	

Contents

GENERAL	7
1 PROJECT DIRECTION	9
2 CATALOGUING OF THE MANUSCRIPTS	10
3 CONSERVATION	12
3.1 Documentation and photography	14
3.2 Surface cleaning	15
3.3 Moistening (humidification chamber treatment)	15
3.4 Straightening/stretching and pressing	15
4 SCANNING	18
5 DIGITISATION	20
6 ACCESS TO THE MATERIALS	22

7 APPENDIXES	24
7.1 Members of the Steering Goup	
Members of the Project Group	25
7.2 Medieval manuscripts' information retrieval system	26
7.3 A Dublin Core Application Profile for the description of medieval manuscripts:	27
7.3.1 General	27
7.3.2 Data elements	28
7.4 Customer interface appearance: examples	37
7.4.1 Collection's front page	37
7.4.2 Search	37
7.4.3 Metadata page	38
7.4.4 Preview	38

GENERAL

Fragmenta membranea, the National Library of Finland's collection of manuscript fragments, consists of texts that were prepared and used in the medieval Swedish Realm. The collection contains approximately 9,300 parchment leaves representing the first 500 years of Finland's literary culture.

The objective of the Fragmenta membranea project was to describe, conserve, and digitise all manuscript fragments in the Library's collection and create an information system facilitating the research and other utilisation of this material. A dedicated user interface was developed to encourage the intensified use of the fragments and protect the original materials.

The project was launched in 2009 and concluded in March of 2012. The entire collection can be accessed online free of charge at the address:
<http://fragmenta.kansalliskirjasto.fi>

Funding the project were the Helsingin Sanomat Foundation (EUR 180,000), the National Library of Finland (EUR 105,000 budget allocation), and the Ministry of Education and Culture (EUR 85,000 separate appropriation). Project financing was also obtained from the National Library of Finland's funds. The total cost of the project was EUR 390,000, which exceeded the project plan's original estimate by EUR 105,000. The overrun resulted primarily from the fact the resources required for the cataloguing of the manuscript fragments were significantly greater compared to the original estimate.

The project benefited substantially from co-operation with Docent, Ph.D. Tuomas Heikkilä's Literary Culture in Medieval Finland research project. Its researchers were responsible for cataloguing the materials and contributed expert assistance to the project in connection with the development of the information system's user interface.

1 Project direction

The project had a Project Group and Steering Group (see Appendix 1.) The Project Group convened when necessary, primarily during the project's initial stages, while the Steering Group met regularly 3-4 times a year.

Juha Hakala from the National Library of Finland was the Project Manager. Liisa Savolainen, Director of the National Library of Finland's Research Library Services, chaired the Steering Group.

The project's preparations began with the formulation of a project proposal in 2008. Following the Helsingin Sanomat Foundation's approval, the preliminary project proposal was developed into an actual project plan jointly with the Finnish medieval literary culture research project and the National Library of Finland's own experts.

It was determined that the National Library of Finland's existing software applications and equipment could be successfully exploited to digitise the fragments and facilitate their availability. For this reason, the project proposal contained no allocations for equipment and software procurements. The reliance on existing technology was also economical because the Library's personnel were already skilled in the systems' use.

2 Cataloguing of manuscripts

During the initial stages of the project, the metadata elements that would be used to describe the manuscript fragments were specified in the Project Group. The specification was based on the international Dublin Core metadata standard (<http://dublincore.org/documents/dces/>), whose application facilitated the semantic interoperability of the parchments' bibliographic descriptions with the metadata of more ordinary library materials.

Because a clarification undertaken by the Project Group determined that Dublin Core had not been used previously to describe manuscript fragments, there were no clear precedents for the development work. Subsequently, however, a similar format solution has been implemented in a medieval manuscript description project at Western Michigan University: (<http://dcpapers.dublincore.org/index.php/pubs/article/view/1006>).

Thanks to the Dublin Core's applicability, the national and international usability of the metadata created in the project is excellent and provides the prerequisites for national and international further projects. It is also possible to load the metadata into the FINNA central index and portal (<http://www.finna.fi>) created by the National Digital Library (NDL) project.

The cataloguers initially recorded the metadata as Excel files. The Excel data was transferred to the docWORKS application that was used for digitisation. In docWORKS the source metadata was migrated into the Dublin Core format.

An overview of the implemented metadata solution, including the local extensions the project made to the Dublin Core format, is provided in Appendix 3.

A large part of the Fragmenta membranea collection had already been analysed and described, considerably facilitating the cataloguing process. Available to the cataloguers were:

- Manuscript descriptions contained in existing printed catalogues (Haapanen and Taitto), 5,325 folios.
- Manuscript descriptions created by Anja Inkeri Lehtinen, 1,573 folios.
- Descriptions created by Tuomas Heikkilä and his research team, 912 folios.

The researcher-cataloguers hired for the project first processed the materials that were included in the existing printed catalogues (Haapanen and Taitto). Where applicable, the metadata has been harmonised; for example the cataloguing language is English and all the names are in Latin, although Haapanen used the German name forms. After processing the materials contained in the printed catalogues, the descriptive focus shifted to the folios included in Anja Inkeri Lehtinen's catalogues. They have not yet been published, but the author graciously made the manuscripts available to the cataloguers.

The metadata records based on printed catalogues do not contain all the data from the catalogues. This would limit the query options the database is able to support. To assist more demanding information seekers, the full text of Haapanen's and Taitto's catalogues has been digitised and indexed as full text.

The Fragmenta membranea collection contains a total of 9,319 folios, of which 1,509 folios are still uncatalogued, but even this material has been conserved and digitised, and can be accessed from the database with a temporary signum. There were three reasons for postponing the cataloguing:

- The fragment's content information is lacking and requires further study (77 folios).
- There are gaps in signum designations, the folios' order is uncertain, and the dating and origin are still unstudied (226 folios).
- The work the manuscript fragment is a part of has not been identified, nor is there sufficient content or descriptive information (1,206 folios).

Since the project unfortunately lacked the resources required for the analyses of these fragments, the Fragmenta membranea collection will thus remain as a research subject even after the posting of the database; the technical solutions developed in the project will also facilitate the inclusion of future research results in the database.

Haapanen's and Taitto's printed catalogues will be published digitally. An agreement for the assignation of the catalogues' user rights was concluded at the National Library of Finland, and the owners of the rights, Ilkka Taitto and Tuomas Haapanen (representing the estate of Toivo Haapanen) have signed it. The catalogues were digitised in 2011.

3 Conservation

Although each parchment leaf was handled individually, the entire collection was treated as an aggregate as effectively and flexibly as possible, keeping in mind, however, the ethical rules of conservation. Achieving a high-quality digitising result was the starting point.

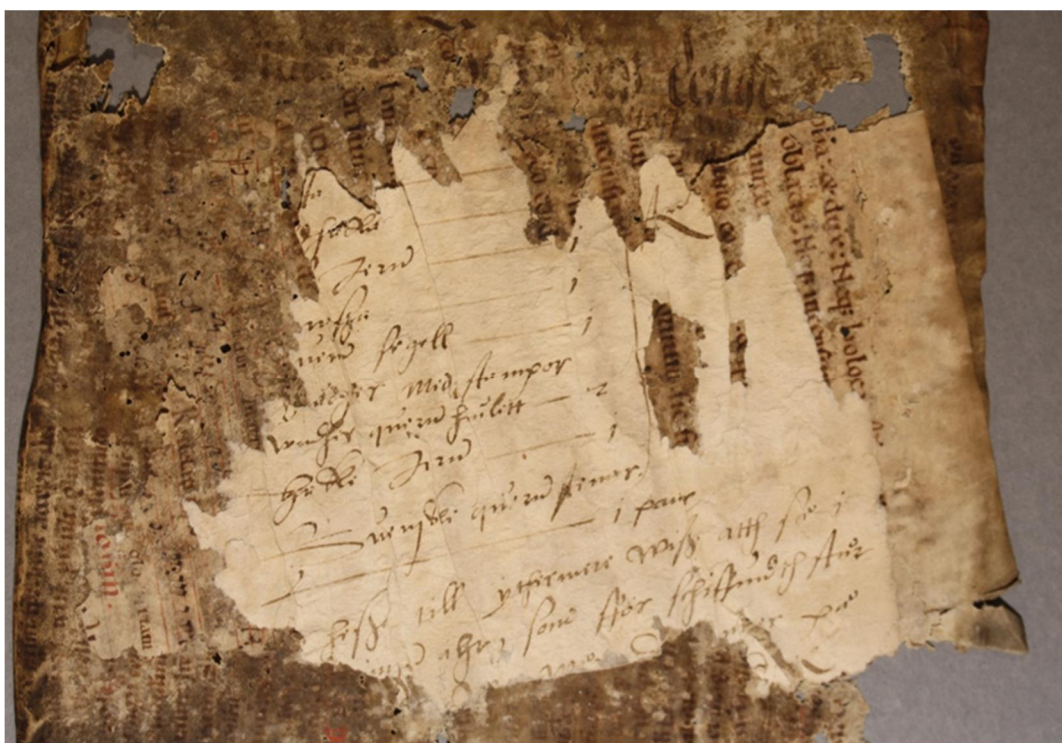
In the spring of 2009, a test batch of 35 parchment leaves, a representative sample of the manuscript fragments found in the manuscript collection, was processed in the national library's Centre for Preservation and Digitisation in the city of Mikkeli, 250 kilometres from Helsinki. The test batch was used to determine the most effective conservation process and analyse the challenges arising in connection with the fragments' digitisation. At the same time, various types of dry cleaning, humidification, and straightening methods were studied and tested. The digitising process generated information regarding the level to which conservation measures were to be carried out.

To ensure conservation quality, a separate area, in which every stage of the process had its own work station, was "constructed" at the facility. A humidification chamber facilitating the simultaneous humidification of the large number of parchment leaves was constructed in the working area. At the same time, suitable equipment required for the post-humidification straightening and stretching processes was procured and configured. The conservation of the materials began in August of 2009. Thanks to careful preliminary planning, the work progressed for the most part as had been anticipated. The National Library paid particular attention to, for example, the safe transport of materials from Helsinki to Mikkeli and back, self-financing the acquisition of special shipping crates designed for the transport of extremely valuable materials; none of the irreplaceable materials were damaged during the transfers. '

During the conservation process it was necessary to carry out various additional clarifications regarding the damage sustained by the parchments as well as the applicability of conservation techniques to different materials. For example, the effect of humidification treatments on texts (writings) that had been made in gold was studied. Paper fragments that had been glued to the parchment leaves resulted in unforeseen additional work. These fragments were only detected during the cataloguing process. The wish, however, was to preserve those paper fragments that contained text (see picture below). The detachment of paper fragments from parchment leaves safely and intact was extremely slow and laborious.



1. Conservation facilities featured a work station for every stage of the process



2. Paper fragment glued to parchment leaf

The most common damage to the leaves resulted from grime, soot, stains, folds, fire damage, moisture damage, ink corrosion, mould, tears, missing pieces, and “cloth tape”.

The conservation methods included the materials’ documentation, photography, surface cleaning, humidification chamber treatment, straightening/stretching, finishing.

All measures absolutely necessary from the standpoint of digitisation were utilised, and the damage sustained by the materials determined how these conservation methods were implemented in practice.

1.



2.



3.



4.



5.



6.



3. Types of damage:

1. moisture damage, ink corrosion 2. folds, fire damage 3. “cloth tapes”
4. missing pieces 5. mould, tears 6. grime, soot, stains

3.1 Documentation and photography

Early in the process, the printed conservation documentation form originally created for the project was replaced by an electronic form. The documentation was made by manuscript (signum), not by leaf. Photographs were used to supplement written documentation and elucidate visual observations. Due to the tight time scheduling, the decision was made to photograph only the parchment leaves requiring humidification and straightening.

During the documentation stage, the leaves were divided into two groups: leaves requiring surface cleaning only, and leaves requiring surface cleaning and humidification. The materials’ documentation information was supplemented after conservation.

3.2 Surface cleaning

The purpose of surface cleaning was to remove loose grime from the parchments' surfaces by using various surface cleaning techniques. The measure is absolutely necessary before the humidification chamber treatment; any grime absorbed into the parchment during moistening could cause extremely severe damage.

3.3 Moistening (humidification chamber treatment)

The purpose of moistening was to make the parchments more flexible for straightening. The humidification chamber was generally used for moistening, but local moistening was used for leaves with ink corrosion or gold.



4. Moistening

3.4 Straightening/stretching and pressing

Straightening was accomplished by placing the leaves either under weights or in stretching frames, after which the leaves were placed between felts and cardboard sheets for final straightening. The straightening method selected took into account the parchments' damages and quality.



5. Straightening under weights.



6. Straightening in a stretching frame.

After conservation measures, the straightened parchment leaves were classified by signum and placed in protective casings to await digitisation, after which the materials were immediately returned to the conservation facilities where they were inspected, protected, and packed for shipment back to the Helsinki collections.



7. Placement in protective casing

The total number of parchment leaves received was 9,206 leaves, all of which were documented. Surface cleaning was carried out for approximately 6,800 leaves. A portion of the parchments (approximately 2,500) had been conserved previously and thus required no other measures than documentation and protection; 3,137 leaves (34%) were moistened and straightened. Paper fragments were detached from 416 parchments (approximately 4.5%).

Processing the parchment fragments was thus more challenging than had been expected. The manuscripts were only in protective casings and they had been numbered by leaf; the numbering of every manuscript began from the number 1. It took more time than expected to simultaneously handle the large quantity of single leaves whose numbering was identical (1, 2, 3 etc.). The leaves from different manuscripts had to be kept separate from each other to avoid mix-ups, which in turn slowed the scanning process.

The National Library of Finland's staff and the Literary Culture in Medieval Finland project's researchers developed the general principles governing conservation in close co-operation.

4 Scanning

The National Library of Finland's existing equipment and applications were used in scanning processes; scanners acquired by the Centre for Preservation and Digitisation in December of 2009 generated an extremely high image quality. The objective was that the scans would never have to be made again.

The scanning proceeding for the most part as expected, although the work was slowed by the need to install, for example, leaves consisting of multiple small pieces in scanners. The opening of fragments consisting of four or more leaves was also occasionally difficult and required the use of separate supports.

A total of 9,290 parchment leaves were scanned during the project.



8. Pellet weight and metal spatulas used for support in scanning.



9. Support of double page.



10. Cardboard backing supporting brittle materials.



11. Removal of silk paper before scanning

5 Digitisation

The printed manuscript catalogues were scanned in the summer of 2011, and the manuscript descriptions in them were linked to the digitised versions of the manuscripts as a part of the digitisation process.

The process was designed by the library staff and researchers in the summer of 2011, and a detailed work plan was formulated to ensure smooth workflows. The docWORKS application used in the digitisation was enhanced to facilitate, among other things, the digitised materials' long-term preservation.

90% of the digitisation work was completed before the end of 2011. The remaining manuscript fragments, of which there were approximately 90, were produced during early 2012. At the end of the digitisation process there was a METS¹ container for each manuscript. These containers can be used as Submission Information Packets to send the data into the digital archive as specified in the OAIS Reference Model. METS was also used to send the data from docWORKS to DSpace. At this stage, certain METS containers exhibiting technical problems were rebuilt in January-February of 2012.

One of the aims of the project was to eliminate the need to digitise the manuscripts ever again, meaning that it must be possible, at least in principle, to preserve the digitised materials permanently. This is only possible if there is manuscripts ever again, meaning that it must be possible, at least in principle, to preserve the digitised materials permanently. This is only possible if there is adherence to the principles of the National Digital Library's Long-Term Preservation, or LTP system.

The LTP system is not yet in production (the current plan is to implement the system in 2016) but the first version of the requirements specifying the structure and content of the Submission Information Packages of the LTP system was completed in 2011². This document was most useful to the Fragmenta membranea project because the LTP system is a pioneer in the long-term preservation field in Finland, and with respect to its progressive policies regarding the long-term preservation of metadata, possibly worldwide.

Unlike traditional descriptive metadata, the administrative metadata (including technical and preservation metadata) is not, or should not be, generated manually, but automatically by programs. The docWORKS application utilised in digitisation was capable of generating OAIS Submission Information Packages based on the METS standard as required by the NDL project.

¹ <http://www.loc.gov/standards/mets/>

² http://www.kdk.fi/images/tiedostot/KDK_metatiedot_ja_aineiston_paketointi_v1.3.pdf

The project was the National Library of Finland's first initiative to exploit the work done by the National Digital Library. The pioneer status came at a price; since 2011, LTP requirements have developed, as a result of which the data packages containing the manuscripts must be updated before they can be transferred to the future long-term preservation system alongside the National Library of Finland's other digitised materials. But every cloud has a silver lining; the experience gained in the *Fragmenta membranea* project related to the mechanised generation of administrative metadata also facilitated the further development of the LTP system's specifications.

The advantages of strict adherence to national long-term preservation policies are undeniable: it facilitates, besides the materials' long-term preservation, the possible export of the digitised manuscripts to other organisations' information systems, as well as the downloading of the descriptive metadata to the memory institutions' shared FINNA central portal and other information retrieval systems.

Nearly all the administrative metadata required in long-term preservation can be generated programmatically in digitisation processes, at least in principle. Because docWORKS supports the NDL requirements for long term preservation, fulfilling these requirements did not have a measurable effect on the time required for the digitisation.

Although detailed explanations of the policies formulated in the project is beyond the scope of this report because they can be found in control documents related to the NDL project's administrative metadata³ and file formats⁴, the *Fragmenta membranea* project functioned as one of the LTP system's pilot projects, providing the NDL with valuable information in a situation where few memory organisations would have been capable of testing the feasibility of LTP-related requirements in practice.

Because the technical principles for long-term preservation are similar for all kinds of library materials, it has been possible to export the know-how developed in the *Fragmenta membranea* project's planning stages and production processes directly to the National Library of Finland's other digitisation initiatives, as well as indirectly – through updated NDL specifications – to other memory organisations' digitisation projects. Based on the experiences gained in NDL pilot projects carried out in 2012 and 2013, it should, however, be noted that the production of digital documents in projects meeting NDL requirements will only be efficient with versatile docWORKS-type applications facilitating a highly automated production process; the manual construction of METS containers is excessively time-consuming and susceptible to error.

In certain respects the project exceeded the LTP requirements. The condition assessments and information gathered during the conservation process was stored in METS containers as administrative metadata for further utilisation. There is also a substantial amount of image-related technical metadata in MIX format⁵ assembled in connection with scanning. This will facilitate the presentation of these images and their migration to new image formats.

³ http://www.kdk.fi/images/tiedostot/KDK_metatiedot_ja_aineiston_paketointi_v1.3.pdf

⁴ <http://www.kdk.fi/images/tiedostot/KDK-PAS-tiedostomuodot.pdf>

⁵ <http://www.loc.gov/standards/mix/>

6 Access to the materials

The digitised manuscripts are freely accessible at the National Library of Finland's DSpace publication archives; the address of the service is: <http://fragmenta.kansalliskirjasto.fi/>.

In July of 2011, the design of the user interface was initialised jointly with researchers in the field. Compared to conventional DSpace document collections, the medieval manuscripts pose some interesting challenges. Since researchers are the main target group, the user interface has been provided with optimised and versatile search options. The user interface's visual appearance was also modified with the researchers' needs in mind.

As was originally scheduled, the database became ready for use in March of 2012, but guidance texts and new features observed as being necessary during the testing phase were added to the service later in 2012. Appendix 4 features screenshots of the system's appearance in January 2013.

Because the implementation is the National Library's own, the appearance and functionality of the user interface can in principle be developed further within the limitations imposed by available resources.

Use statistics can be found at the address: <http://fragmenta.kansalliskirjasto.fi/simplestats/>. Between June 2012 and April 2013, approximately 88,000 manuscript pages have been downloaded by the users.

The DSpace application⁶ is open source code software with a wide-ranging and active user community. DSpace search functionality was modified slightly by the project; an enhanced version makes it possible to search manuscripts on the basis of signum, date of creation, or other metadata. The special solution was necessary because DSpace's "out of the box" search function would not have been sufficiently versatile to deal with the fragments and their somewhat exceptional metadata. Adaptive features related to appearance and the functionality of the user interface were also implemented for DSpace.

To guarantee the quality of the service, the project's researchers and other relevant parties familiarised themselves with equivalent international information retrieval systems (see Appendix 2). The National Library of Finland's DSpace-based information retrieval system already incorporated many of these systems' features, and missing capabilities were added during the adjustment phase. The project's staff also added certain features, such as the possibility to view pages of different sizes, which are rarely encountered in equivalent systems.

⁶ <http://www.dspace.org/>

A researcher for the most part approaches a digitised manuscript collection like he or she would approach an “actual” collection – by localising familiar manuscripts, in other words the kinds of manuscripts whose signums are already known. This is now easily accomplished by effortlessly browsing the manuscripts in the user interface. In this connection, being able to adjust the number of references shown on the screen while browsing is particularly practical. The search function facilitates the collection’s analytic use in ways that deviate dramatically from the techniques utilised before the existence of the digitised collection.

A search can target the entire collection or one of its sub-entities. Besides locational information, search terms were selected based on researchers’ assessments and equivalent international databases; these included a manuscript’s content, genre, language, the place where it was written (origin), the author of its presented text, the religious order with which it was associated (liturgical use location), as well as a dating function (see Appendix 3) that also facilitates searches from certain time periods. The descriptive information can also be targeted at a free text search. So far, the search mechanisms have been ascertained as being workable for research purposes, and they are also more versatile compared to those in many other international manuscript databases.

Unlike many digitised manuscript collections, the *Fragmenta membranacea* database provides users with several images of each manuscript in different sizes and in different file formats. In this respect the best basis for comparison can be found at the manuscript database of the Cologne Archdiocese (see Appendix 2). There the user can select several images of different sizes for examination, but only together and in the same file format. Its versatility enables the *Fragmenta membranacea* database to serve various functional purposes. When gathering information, often only a quick glance at digitised materials is necessary; for more thorough research, large detailed images are absolutely necessary.

The library had already software that semi-automatically converts the METS containers (metadata and documents) generated in docWORKS to a format suiting DSpace. This application had been used for e.g. monographs; it was modified to streamline the manuscripts’ processing. Further modifications, facilitating the loading of other material types, may be carried out in the future.

Because the project could utilise the National Library of Finland’s previous DSpace development work, the additional properties required by the manuscripts could be implemented quickly. When the digitisation was complete, all systems necessary for the manuscripts’ online publication were pre-installed and operable. Some changes to the DORIA service’s technical infrastructure were required, including the Dspace update to version 1.7.

Because the format of choice was Dublin Core, metadata created in the project can be extracted either directly from the METS containers or from DSpace, and transmitted to other information systems. Owing to the Dublin Core extensions made, both a comprehensive Dublin Core metadata record and a more convenient simple record have been stored in the METS containers. If the comprehensive record is used, the recipient must know how to handle the collection-specific metadata.

7 Appendixes

7.1 Appendix

Members of the Steering Group

Juha Hakala, National Library of Finland
Tuomas Heikkilä, University of Helsinki
Minna Kaukonen, National Library of Finland
Marja Pohjola, National Archives
Leena Saarinen, National Library of Finland
Heleena Savela, Helsingin Sanomat Foundation Sanomat Foundation
Liisa Savolainen, National Library of Finland (Chair)

Members of the Project Group

Harri Ahonen, National Library of Finland
Juha Hakala, National Library of Finland (Chair)
Sirkka Havu, National Library of Finland
Jukka Kervinen, National Library of Finland
Eila Kupias, National Library of Finland
Leena Saarinen, National Library of Finland
Jaakko Tahkokallio, University of Helsinki
Heidi Törrönen, National Library of Finland

7.2 Appendix

Medieval manuscripts' information retrieval systems

Cologne Archdiocese libraries' manuscripts:

<http://www.ceec.uni-koeln.de/>

Swiss libraries' manuscripts:

<http://www.e-codices.unifr.ch/>

Herzog-August Bibliothek, Wolfenbüttel:

<http://diglib.hab.de/?db=mss>

Badische Landesbibliothek, Karlsruhe:

<http://digital.blb-karlsruhe.de/Handschriften/nav/classification/20950>

Bayerische Staatsbibliothek, München:

http://www.digitale-sammlungen.de/?c=sammlungen&kategorie_sammlung=1&l=en

7.3 Appendix

A DUBLIN CORE APPLICATION PROFILE FOR THE DESCRIPTION OF MEDIEVAL MANUSCRIPTS

7.3.1 GENERAL

This document describes the metadata format used for cataloguing the manuscripts, and explains the slightly non-standard usage of some Dublin Core elements. There is also a short description of the METS packages generated in the digitisation process. For its part, METS guidelines are an essential part of the compatibilities between the National Digital Library's (NDL) project and METS policies.

From the standpoint of information retrieval, it is essential that:

- the references' descriptive language is English, but names are in Latin according to the HELKA⁷ database's general practice. The descriptive languages of published manuscript catalogues vary; in Haapanen's catalogues, it is German according to the practice prevailing at the time of publication. In Taitto's catalogue, as well as the materials of Tuomas Heikkilä's research team, it is English, and in Lehtinen's catalogues, not yet released, it is Latin.
- the subject of the description is a manuscript fragment from the Fragmenta membranea collection. It has still not been possible to match all the fragments with the works from which they originate. Analyses are further hindered because a few of the fragments are extremely small, while the more extensive fragments may consist of dozens of leaves.

A Dublin Core Application Profile was developed specifically to describe the fragments, but during the planning stage, being able to dumb down the metadata to the basic Dublin Core format was considered essential. This dumb down process is an approximation; an equivalent basic Dublin Core metadata element can always be found, but in certain cases the information content does not fully match the instructions for the use of the metadata element in question.

It should be possible, however, to load the fragments' metadata to other systems such as FINNA or Europeana.

The manuscript descriptions from printed catalogues were embedded in the Dublin Core records as a separate dc:description field that was indexed for full text searches. There is a link from the database record to the description of the manuscript in the digitised catalogue.

⁷ The National Library's OPAC, available at <https://helka.linneanet.fi/>

7.3.2 DATA ELEMENTS:

1. IDENTIFIER

A manuscript's signum, its identifier, is stored in the Dublin Core *Identifier* element. For researchers, the signum is the primary search element. Although often volatile, in this collection it can also be assumed that the signum will remain fairly permanent. The only exceptions to this are temporary signums, which may change after the future research has revealed the work and / or manuscript from which the fragment originated.

The signum expresses the name of the collection (F.m. = Fragmenta membranea), a sub-collection in Roman numerals, as well as a sub-collection-level sequential number, for example F.m. I.157, F.m. II.1, F.m. III.1, F.m. IV.1, F.m. V.BI.1.

In the light of current knowledge, signums are fully searchable in DSpace. The project did not investigate whether this also applies to other information retrieval systems such as VuFind. If necessary, signums and other metadata can be altered in connection with conversion or indexing taking place in the target system.

The manuscripts have an URN identifier, giving them a unique Internet-wide identifier that can be used when creating links to references. To be on the safe side, a fragment's signum is not used as its URN identifier; instead it is a non-semantic sequential number, an NBN (National Bibliography Number) from the *fd* sub-domain used normally in digitising. Thus the URN identifier will not change when, for example, a fragment currently assigned a temporary signum is eventually given a permanent one.

In Dspace indexing, the signum and URN must be separated from each other because the data is processed differently. For that reason they are entered according to the following model:

```
<dc:identifier type="signum">F.m.I.24</dc:identifier>
```

```
<dc:identifier type="urn">URN:NBN:fi-fd2011-1200075</dc:identifier>
```

Note 1: type="signum" is a local DC extension that is not necessarily supported in all information retrieval systems. If the support is lacking, the metadata element, based on the rules of DC's *dumb down* rule, should be handled in the search system like the normal *dc:identifier* field.

Note 2: the ability to separate the signum and URN from each other in indexing and information searches with the above coding is an absolutely necessary property in information retrieval systems where manuscripts are set as the primary search targets.

Page-level identifiers

The most concise way to enter a page-level URN identifier into the METS information structure is to use the physical StructMap's div-element's CONTENTIDS attribute. In that case the identifier refers to the page as a concept, not as the page of a certain file.

```
<mets:structMap LABEL="Physical Structure" TYPE="PHYSICAL">
<mets:div ID="DIVP1" DMDID="dmd-dc" ADMID="main-amd"
LABEL="Missal" TYPE="METAe_Monograph">

<mets:div ID="img0001" ORDER="1" TYPE="PAGE" CONTENTIDS="URN:NBN:fi-fd2011-
1200075">
<mets:fptr>
<mets:par>
<mets:area FILEID="img0001-master"/>
<mets:area FILEID="img0001-access"/>
<mets:area FILEID="img0001-thumb"/>
<mets:area FILEID="img0001-alto" BETYPE="IDREF" BEGIN="P1"/>
</mets:par>
</mets:fptr>
</mets:div>
```

A second method is to refer the page's div to a separate page-level metadata section. The metadata can be in a MODS or DC format.

```
<mets:dmdSec ID="img0001-dmd">
<mets:mdWrap MIMETYPE="text/xml" MDTYPE="MODS" LABEL="Bibliographic meta-data of
chapter '1'">
<mets:xmlData>
<MODS:mods>
<MODS:identifier type="urn">URN:NBN:fi-fd2011-1200075</MODS:identifier>
</MODS:mods>
</mets:xmlData>
</mets:mdWrap>
</mets:dmdSec>

<mets:structMap LABEL="Physical Structure" TYPE="PHYSICAL">
<mets:div ID="DIVP1" DMDID="dmd-dc" ADMID="main-amd"
LABEL="Missal" TYPE="METAe_Monograph">

<mets:div ID="img0001" ORDER="1" TYPE="PAGE" DMDID="img0001-dmd">
<mets:fptr>
<mets:par>
<mets:area FILEID="img0001-master"/>
<mets:area FILEID="img0001-access"/>
```

```

<mets:area FILEID="img0001-thumb"/>
<mets:area FILEID="img0001-alto" BETYPE="IDREF" BEGIN="P1"/>
</mets:par>
</mets:fptr>
</mets:div>

```

2. *GENRE*

There are only a few Genre terms, for example Liturgy or Theology.

The *dc:type* element was used to express Genre.

```
<dc:type>Liturgy</dc:type>
```

Usually the Dublin Core Genre terms are selected from the list at <http://dublincore.org/documents/dcmi-type-vocabulary/>. Because the list does not contain the Genre terms needed for medieval manuscripts, the project had to specify its own term list.

3. *TITLE*

The title is provided in the Dublin Core *dc:title* field.

```
<dc:title>Missal</dc:title>
```

Although Dublin Core has no obligatory metadata elements, records without title may cause problems with certain applications. With manuscripts, specifying the title is a problem. Even if the title can be given, it is often generic, such as "Missal". There were 338 manuscripts where no title could be given. A marking [s.n.] indicating the lack of a title was already added to them at Mikkeli to ensure that the data records' DSpace load would not fail due to title checks, and that references without titles are easily and separately searchable.

4. AUTHOR

The Latin-language form of the author's name was entered in the Dublin Core *dc:creator* field.

```
<dc:creator>Aristoteles</dc:creator>
```

The authors' names (for example Aegidius Romanus OESA) had to be entered according to cataloguing rules that required, for instance in the example above, the deletion of the additional information (OESA) denoting the religious order. From the standpoint of authority supervision, the corrected form of the name poses no problems because, in compliance with the cataloguing rules, the Latin forms of the names of medieval persons are used in the HELKA database. The Virtual International Authority File (<http://viaf.org>) database's authority data records generally also contain this Latin-language form, and for that reason the names can, with the VIAF's assistance, be converted automatically to the Anglo-American form if necessary.

If there were several authors, each one was entered into its own *dc:creator* field.

In at least three cases, it was not possible to determine the author with 100 % certainty. The problem was solved by giving two names; one is the most likely author of the work in question, while the other is an alternative. The latter name was entered into the Dublin Core *dc:description* element to prevent it from being shown in the author browse list.

```
<dc:creator>Bartholomeus Brixienensis</dc:creator>
<dc:description>Iohannes Teutonicus</dc:description>
```

The same method was applied when the author information was unreliable.

1. Decretum Gratiani cum Glossa Ordinaria / 2. Petrus Ilerdensis
Abbas Antiquus (Bernardus de Montemirato)
Accursius
Aegidius Romanus OESA
Albertus <Coloniensis> Magnus OP
Alexander IV Papa / Gregorius IX Papa
Alexander de Hales (Halensis) OFM
Aristoteles
Augustinus Aurelius
Azo Porcius
Bartholomaeus Anglicus de Glanville OFM
Bartholomeus Brixienensis - Iohannes Teutonicus?
Bartholomeus de Sancto Concordio Pisanus OP
Bernardus Papiensis
Bernardus Parmensis de Botone
Bernardus abbas Claraevallensis OCist
Birgitta de Suecia

Bonaventura de Balneoregio OFM
Bruno Herbipolensis
Conradus de Halberstadt OP
Defensor Logotiagensis
Durandus de Sancto Porciano OP
Formula
Franciscus de Abbatibus OFM
Goffredus Tranensis
Gregorius I papa
Gregorius IX Papa
Guido de Baysio (archidiaconus)
Guillelmus de Saliceto (de Placentinis)
Guillelmus Duranti
Guillelmus Petri de Godino Baionensis OP
Guillelmus de Melitona OFM
Guillelmus de Ware OFM
Hugo Ripelin de Argentina OP
Hugo de Sancto Caro OP
Huguccio Pisanus
Incerti Auctoris
Incerti auctoris
Incerti auctoris (Rufinus Aquileiensis)
Innocentius IV Papa (Sinibaldus Fliscus)
Innocentius IV Papa?
Iohannes Andreae
Iohannes Monachus
Iohannes de Balbis lanuensis OP
Iohannes de Friburgo OP
Jacobus de Losanna OP
Jacobus de Vitriaco
Jacobus de Voragine
Jacobus de Voragine OP
Jacobus de Voragine OP et quidam alii
Johannes Crysostomus
Johannes de Ripa OFM
Jordanus de Quedlinburg OESA
Lietbertus de Insulis OSACan
Lucius Annaeus Seneca
Mathias de Lincopia
Maximianus; Publius Papinius Statius
Nicholaus Tudeschis, abbas Panormitanus?
Nicolaus de Gorran OP
Nicolaus de Gorran OP (uel Guillelmus de Melitona OFM?)
Nicolaus de Lyra OFM
Origines
Petrus Beneventanus
Petrus Berchorius Pictaviensis OSB
Petrus Cantor Parisiensis

Petrus Capuanus
Petrus Comestor OSACan
Petrus Lombardus
Petrus de Alvernia
Petrus de Tarentasia OP
Pierre Bersuire OSB
Ps. Albertus Magnus
Raymundus de Pennafort OP
Robertus Grosseteste
Robertus Holcot OP
Stephen Langton
Thomas de Aquino OP
Vincentius Belvacensis (de Beauvais) OP
Vincentius Hispanus

5. CHRONOLOGY

Expressing chronological information in a machine readable form was a challenge, because the date format preferred by researchers was neither supported in the Dublin Core format nor computer understandable. Moreover, dating the manuscripts was often difficult, and often only a rough estimate of when the manuscript was written could be given.

The researchers' prefer date information in the form "Saec. xii" or "Saec xii–xiii". The machine readable form of the examples above is 1101–1200 and 1101–1300.

Machine readable date information was entered into the *dc:date* field according to the ISO 8601 standard's EDTF profile. The date in the form preferred by researchers is supplied in *dc:description*.

```
<dc:date>1101/1200</dc:date>
<dc:description>Saec. xii </dc:description>
```

Many information retrieval systems expect the data in the *dc:date* field to be in the ISO 8601 format; another kind of data could result in the rejection of the data record.

6. ORIGIN

The researchers wanted to describe the origin of the manuscript, that is, to define the geographical region or bishopric where the manuscript was written (for example France or the Church province of Cologne). This data was supplied in the *dc:coverage* element that is normally used to provide the geographical coverage of the resource, and is therefore relatively close to the project usage.

```
<dc:coverage>England</dc:coverage>
```

dc:coverage could not be used to express the religious order where the work has been used (for example Cistercian). Instead, the *dc:provenance* element was used.

```
<dc:provenance>Diocese of Maastricht</dc:provenance>
```

7. LANGUAGE

The language of the manuscript is expressed as a standard 3-letter language code (for example lat).

ISO 639 3-letter codes can be found at:

http://www.loc.gov/standards/iso639-2/php/English_list.php

```
<dc:language xsi:type="dcterms:ISO639-2">lat</dc:language>
```

8. ADDITIONAL INFORMATION

There was a considerable amount of additional information available that was difficult to express in any standard metadata format. For instance, researchers could refer to the previous research:

DATE: saec. xiii/xiv suggested (Niskanen).

LITURGICAL USE: Cistercian suggested (Niskanen).

ORIGIN: Nordic suggested (Niskanen).

This type of information was stored in *dc:description*. At its most complex, the metadata appeared as follows:

```
<dc:description>DATE: Probably saec. xii 2/2 (MPO); saec. xii med. (Niskanen 2010).  
ORIGIN: Probably England (Haapanen, probably according to liturgical use). Probably  
England or English influence (Niskanen 2010). OTHER NOTES:  
Also known as: MPO Fr 26644 (CCM Mi 216). From same codex: CCM Mi 216 (Stockholm,  
Riksarkivet: Fr 26639 KB A 103 b 171; Fr 26640 KB A 103 b 170; Fr 26641 KB A 103 b 157;  
Fr 26642 KB A 103 b 208; --, --: Kammararkivet Fr 9248 Fogdarnas restantieräkenskaper 2:5  
Uppland 1557; Fr 26645 Provianträkenskaper 18:16 Estland Reval 1561-1564 Engelbrekt  
Anderssons r 1562; Fr 26646 Provianträkenskaper 18:2 Estland Reval 1561-1564 Engelbrekt  
Anderssons r 1561; Fr 26647 Smålands handlingar 1579:10:4 Kvittenser; Fr 26648 Finska  
cameralia 35:1 Erik Olufssons löneregister 1562; Fr 26649 Ångermanlands handlingar  
1563:17:1 Räkenskap; Växjö, Växjö L.: Fr 26643 1529:21; 11 fr., 21 fol.).</dc:description>
```

Please note the references to manuscripts held in other collections.

Indexing this kind of metadata is tricky. Following discussions with the researchers, the decision was made to store all additional information pertaining to a manuscript to a single instance of the *dc:description* metadata element, with no attempt to split the data into multiple *dc:description* instances. Ideally the researchers can later update the data, for example by clarifying the relationships between the manuscripts.

9. OTHER INFORMATION

During the digitisation process, information to the effect that the digitised manuscript fragment belonging to the National Library of Finland was generated to the *dc:rightsholder* field. If fragments and/or their descriptions are exchanged internationally, ownership and user rights must be embedded in the data.

<dc:rightsholder>The National Library of Finland</dc:rightsholder>

Equivalently, a text that specifies the user rights both to the manuscript fragments' and their Dublin Core-based metadata was correspondingly generated for the *dc:rights* field, for example:

<dc:rights>Creative Commons Public Domain Mark 1.0</dc:rights>

This license is authenticated as follows (see <http://creativecommons.org/publicdomain/mark/1.0/>):

This work has been identified as being free of known restrictions under copyright law, including all related and neighboring rights.

You can copy, modify, distribute and perform the work, even for commercial purposes, all without asking permission.

Open licensing is sensible in this case because it gives future research projects the possibility to continue the work done by the Fragmenta membranea project. Researchers can enrich the metadata describing the manuscript fragments, or use the digital images of the manuscripts in their own research publications. The free right of use applies to all materials in the DSpace system, including the most detailed printable images.

The exception to the presented operational model is the re-use of reference data copied from the printed catalogues to the METS packets. Their user rights are more limited compared to other metadata; data scanned from printed catalogues and OCR-modified information should not be changed, except for the deletion of OCR errors.

10. PRINTED CATALOGUES

Printed manuscript catalogues have been scanned and converted to text (without manual correction of OCR errors). Metadata records in the Fragmenta membranea database contain the description of the fragment in the dc:description element. An URN identifier has been given for the digitised catalogues as well as for each reference. An URN-based link to the fragments' digitised versions has also been embedded in the metadata records.

11. MODEL DATA RECORD

<dc:identifier type="signum">F.m.I.24</dc:identifier>

<dc:identifier type="urn">URN:NBN:fi-fd2011-1200075</dc:identifier>

<dc:type>Liturgy</dc:type>

<dc:title>Missal</dc:title>

<dc:creator>Robertus Grosseteste</dc:creator>

<dc:date>1101/1200</dc:date>

<dc:description>Saec. xii</dc:description>

<dc:coverage>England?</dc:coverage>

<dc:provenance>Diocese of Maastricht?</dc:provenance>

<dc:language xsi:type="dcterms:ISO639-2">lat</dc:language>

<dc:description>DATE: Probably saec. xii 2/2 (MPO); saec. xii med. (Niskanen 2010).
 ORIGIN: Probably England (Haapanen, probably according to liturgical use). Probably
 England or English influence (Niskanen 2010). OTHER NOTES: Also known as: MPO Fr
 26644 (CCM Mi 216). From same codex: CCM Mi 216 (Stockholm, Riksarkivet: Fr 26639 KB
 A 103 b 171; Fr 26640 KB A 103 b 170; Fr 26641 KB A 103 b 157; Fr 26642 KB A 103 b 208; --,
 --: Kammararkivet Fr 9248 Fogdarnas restantieräkenskaper 2:5 Uppland 1557; Fr 26645
 Provianträenskaper 18:16 Estland Reval 1561-1564 Engelbrekt Anderssons r 1562; Fr 26646
 Provianträenskaper 18:2 Estland Reval 1561-1564 Engelbrekt Anderssons r 1561; Fr 26647
 Smålands handlingar 1579:10:4 Kvittenser; Fr 26648 Finska cameralia 35:1 Erik Olufssons
 lönerregister 1562; Fr 26649 Ångermanlands handlingar 1563:17:1 Räkskap; Växjö, Växjö
 L.: Fr 26643 1529:21; 11 fr., 21 fol.).</dc:description>

<dc:rightsholder>The National Library of Finland</dc:rightsholder>

<dc:rights>Creative Commons Public Domain Mark 1.0</dc:rights>

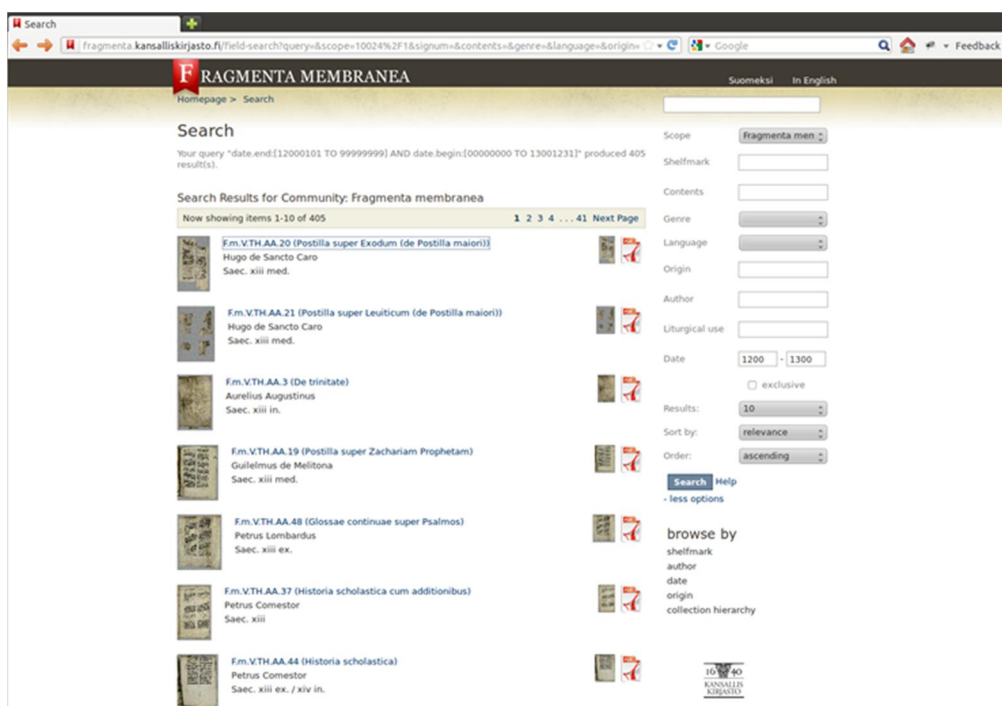
7.4 Appendix

CUSTOMER INTERFACE APPEARANCE: EXAMPLES

7.4.1 Collection's front page



7.4.2 Search



7.4.3 Metadata page

The screenshot shows the metadata page for the manuscript **F.m.V.TH.AA.20 (Postilla super Exodum (de Postilla maiori))**. The page is part of the **FRAGMENTA MEMBRANEA** database, with navigation links for **Suomeksi** and **In English**. The main content area displays the following information:

- Show full item record**
- F.m.V.TH.AA.20 (Postilla super Exodum (de Postilla maiori))**
- Permanent address:** <http://um.fi/urn:nbn:fi:42011-aa0927>
- Shelfmark:** F.m.V.TH.AA.20
- Contents:** Postilla super Exodum (de Postilla maiori)
- Genre:** Theology
- Language:** Latin
- Origin:** France
- Author:** Hugo de Sancto Caro
- Liturgical use:** Saec. xiii med.
- Date:** Saec. xiii med.
- Catalog:** Additional information
- DATE:** Saec. xiii 1/2 also suggested (MPO). OTHER NOTES: From same codex: F.m.V.TH.AA.21; Stockholm, Riksbibliotek; Kammararkivet Fr 7719 Pärmylinader 3.15 Summarium for Uppsala stift 1541; 1 f., 6 fol.
- Preview:** [Show full item record](#)
- Whole fragment**
- Individual pages**
- Download** links for **F.m.V.TH.AA.20.pdf** (1.470Mb) and **F.m.V.TH.AA.20-0001.tif** (48.92Mb).
- Download** link for **F.m.V.TH.AA.20-0002.tif** (48.95Mb).
- donia-aa@helsinki.fi**

The right sidebar contains a **Search** field, a **Help** link, and a **Subcollection by** section with links for **shelfmark**, **author**, **date**, **contents**, **origin**, and **collection hierarchy**. Logos for the **KANSALLINEN KIRJASTO** and **HELSINKIN KIRJASTO** are also visible.

7.4.4 Preview

The screenshot shows the preview page for the manuscript **F.m.V.TH.AA.20**. The page is titled **Fragmenta membranea - F.m.V.TH.AA.20 - Mozilla Firefox**. The main content area displays a large thumbnail image of the manuscript page, showing text in two columns. The left sidebar contains the following information:

- FRAGMENTA MEMBRANEA**
- F.m.V.TH.AA.20**
- Hugo de Sancto Caro**
- Saec. xiii med.**
- Item metadata view**
- Save whole document**
- Save current page**
- Contents**
- Thumbnail image of the manuscript page.**

The right sidebar contains a **Search** field, a **Help** link, and a **Subcollection by** section with links for **shelfmark**, **author**, **date**, **contents**, **origin**, and **collection hierarchy**. Logos for the **KANSALLINEN KIRJASTO** and **HELSINKIN KIRJASTO** are also visible.



F.m.1.196



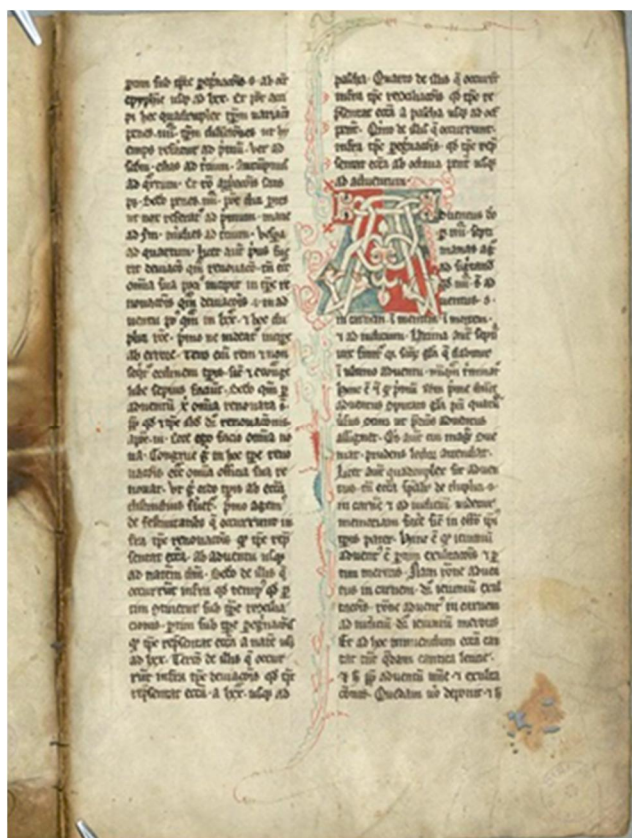
F.m.III.120-0005



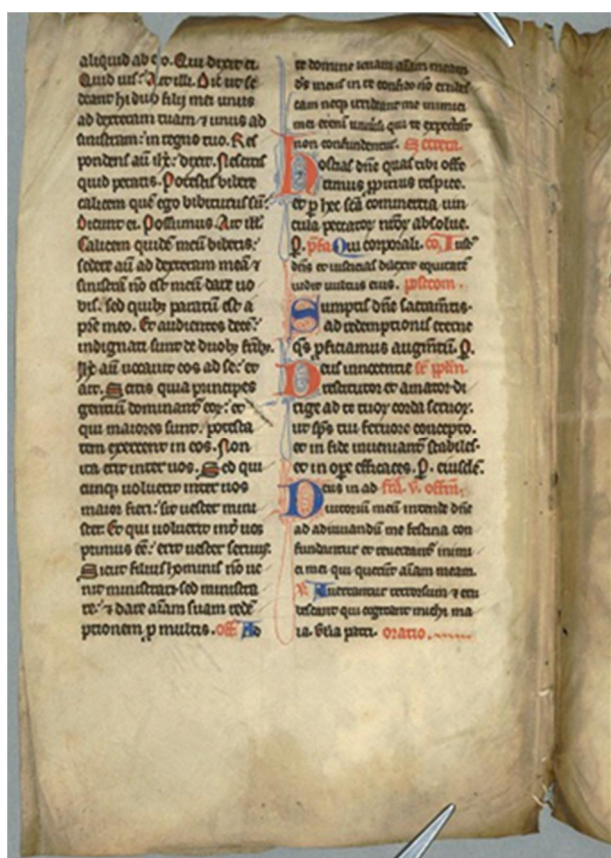
F.m.I.55.fol.1



F.m.I.231-0007



F.m.VII.18



F.m.I.115 page 22