Music Retrieval Tools

Optical Music Recognition (OMR) and Music Analysis

Martha E. Thomae
CESEM-IN2PAST,
NOVA University of Lisbon

Antoine Phan
CESEM-IN2PAST,
NOVA University of Lisbon
and McGill University

Elsa De Luca
CESEM-IN2PAST,
NOVA University of Lisbon

Research Seminars Series
"Musicologists and Their Sources: The Alchemy of Turning Parchment into Pixels"

NOVA University of Lisbon,
Portugal, 24 June 2024

Music Information Retrieval Tools

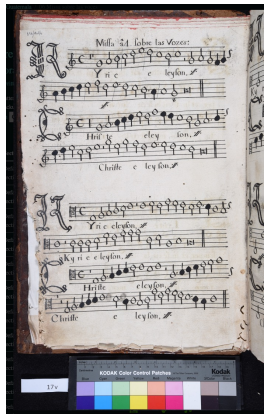
1. Optical music recognition (OMR) *Martha E. Thomae*
2. Music search & analysis *Antoine Phan*

Focus: Early music written in neumatic and mensural scripts

Optical Music Recognition (OMR)

- Similar to Optical **Character** Recognition (O**C**R)
  - Where the characters of a digital text document are readable by the computer
  - It allows to perform content-based searches
- For music: Optical **Music** Recognition (O**M**R)
  - The **music** characters (and text) are readable by the computer
  - Therefore, in addition to search music pieces by their metadata (e.g., composer, title of the piece, and date)
  - The music content is searchable (e.g., ask the computer to look for a melody → melodic search)
  - Moreover, perform computational music analysis

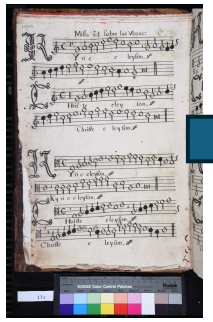
The Traditional OMR Pipeline



Optical Music Recognition

```
<mei meiversion="5.0">
   <meiHead>
      <fileDesc>
         <titleStmt>
            <title>06 Missa sobre las voces</title>
            <composer>Cristobal de Morales</composer>
         </titleStmt>
         <pubStmt/>
      </fileDesc>
   </meiHead>
   <music>
      <body>
         <mdiv>
            <score>
               <scoreDef>
                  <staffGrp>
                     <staffDef n="1" lines="5" label="superius" notationtype="mensural">
                        <clef line="2" shape="G"/>
                        <mensur sign="C" tempus="2" prolatio="2"/>
                     </staffDef>
                  </staffGrp>
               </scoreDef>
               <section>
                  <staff n="1">
                     <layer n="1">
                        <rest dur="brevis" loc="4"/>
                        <rest dur="semibrevis" loc="4"/>
                        <note dur="semibrevis" pname="g" oct="4"/>
                        <note dur="minima" pname="a" oct="4"/>
                        <note dur="minima" pname="b" oct="4"/>
                        <note dur="minima" pname="c" oct="5"/>
                        <note dur="minima" pname="d" oct="5"/>
```

The Traditional OMR Pipeline



Preprocessing
or
Document Analysis
or
Document Segmentation

Music Recognition

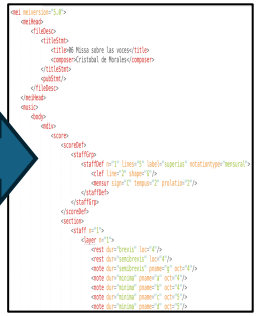
(a) Staff processing

(b) Music symbol processing

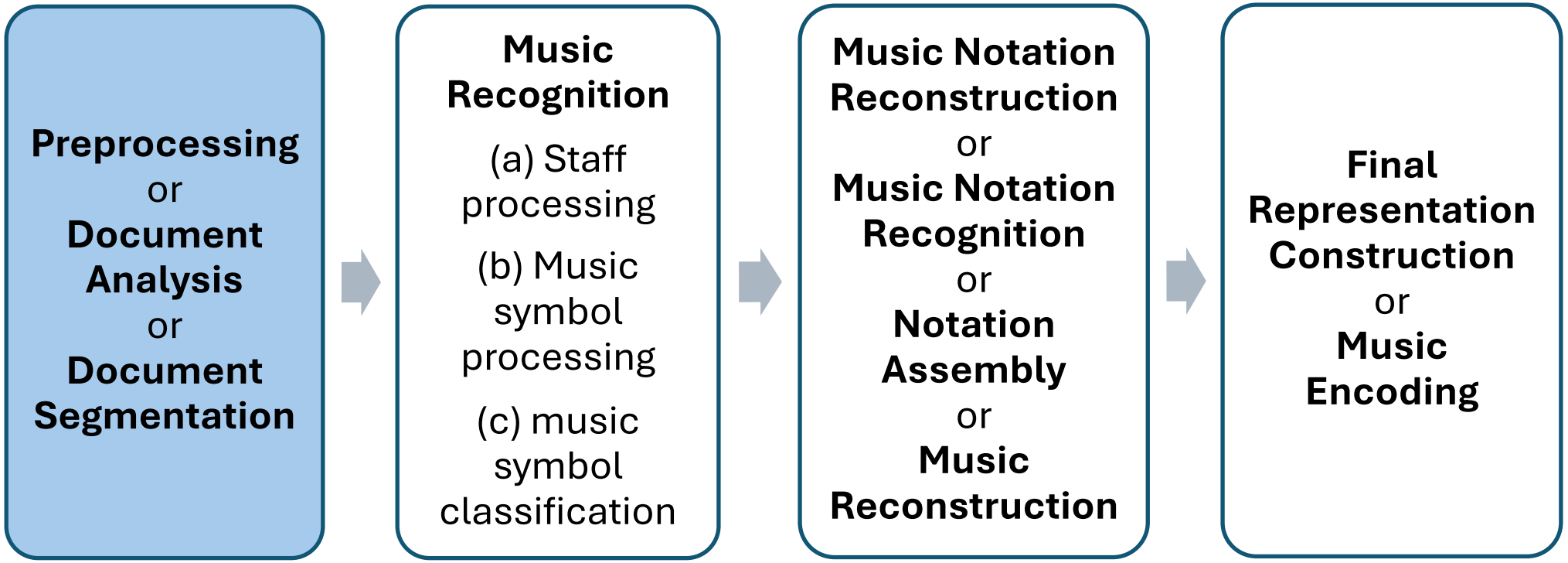
(c) music symbol classification

Music Notation Reconstruction
or
Music Notation Recognition
or
Notation Assembly
or
Music Reconstruction

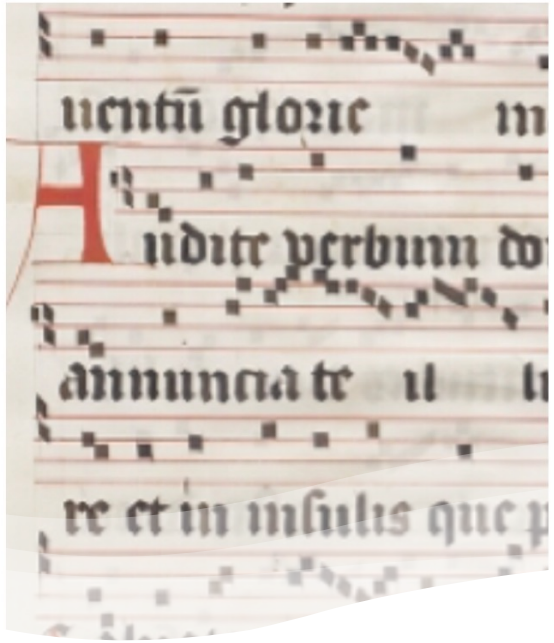
Final Representation Construction
or
Music Encoding



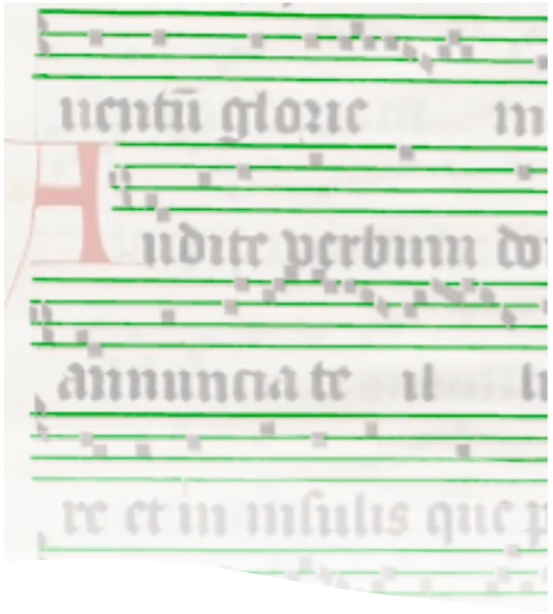
Step 1: Preprocessing or Document Analysis



Actual

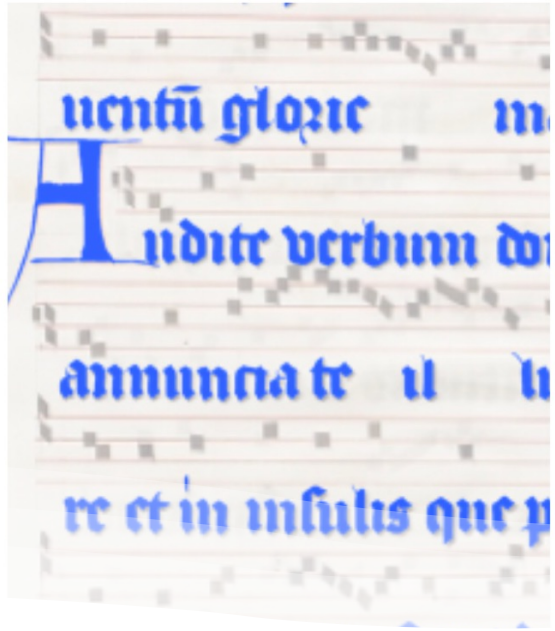


Staff



Note

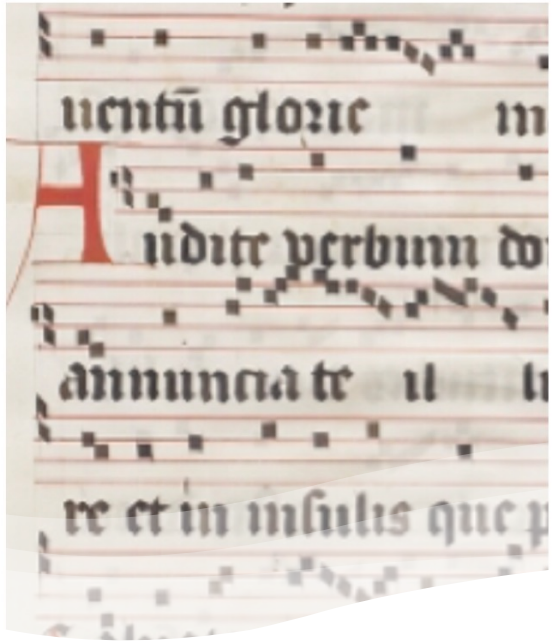
Text



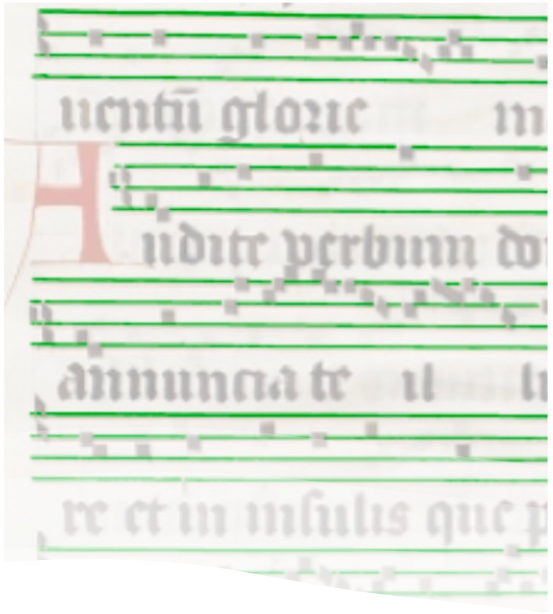
Step 1: Preprocessing or Document Analysis or Document Segmentation

- There is always a **pre-processing step** that **prepares the data** needed for the **following steps** in the pipeline
- Normally, this step segments the document into layers: music symbols & staves & text layers

Actual

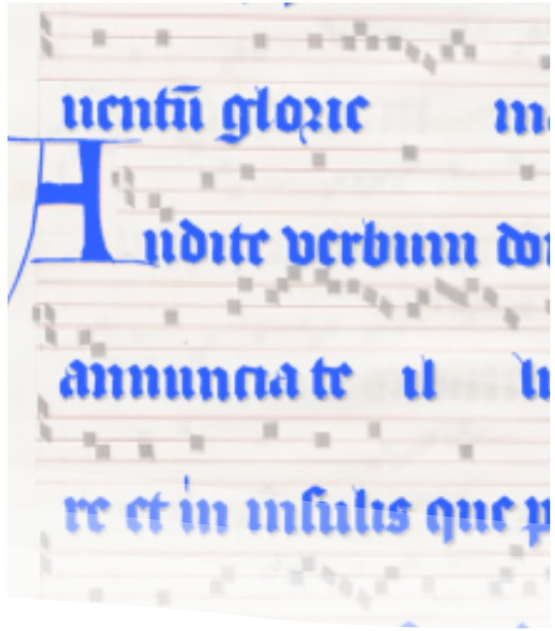


Staff



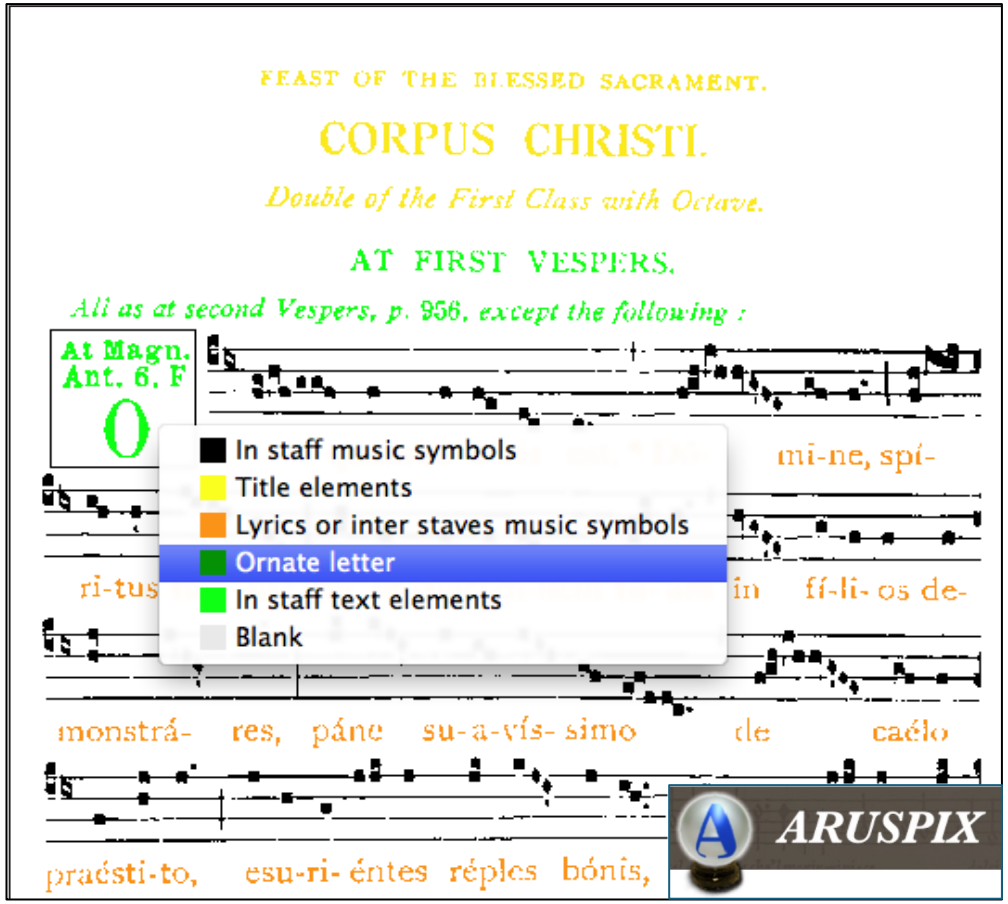
Note

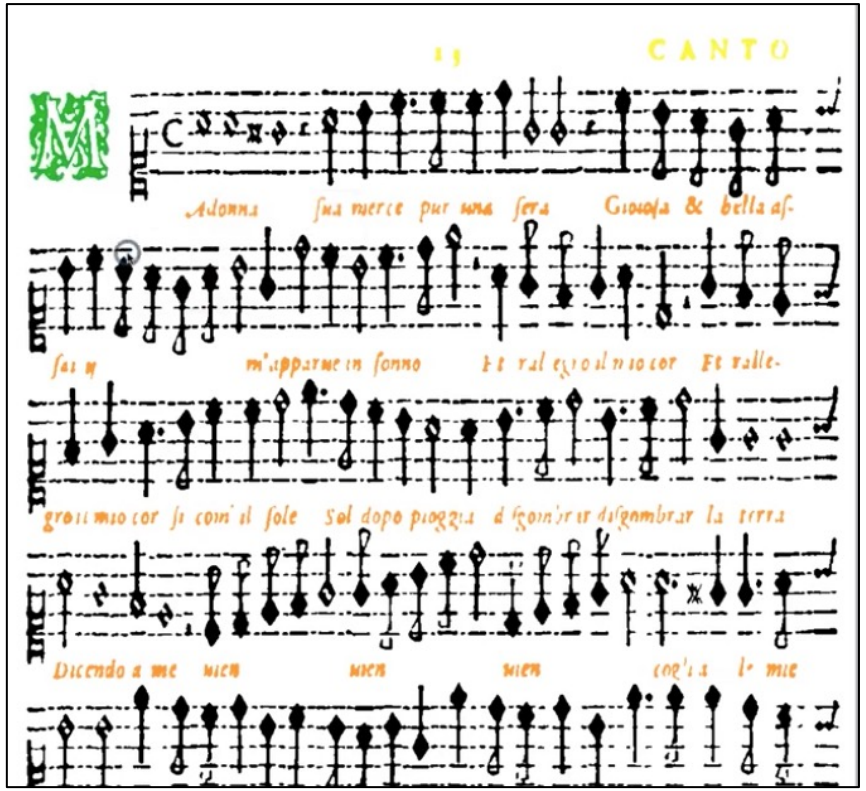
Text



Step 1: Preprocessing or Document Analysis or Document Segmentation

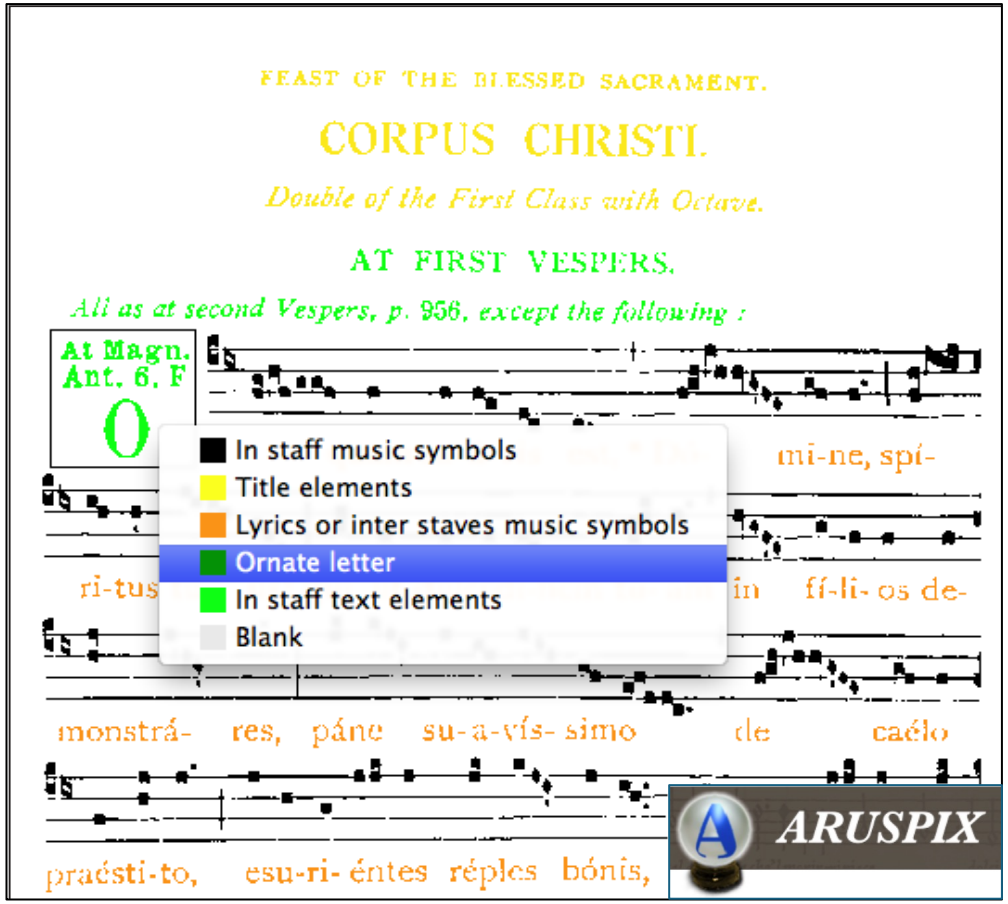
- There is always a **pre-processing step** that **prepares the data** needed for the **following steps** in the pipeline
- Normally, this step segments the document into layers: music symbols & staves & text layers
- Sometimes the music symbols are not separated from their staff and the staff regions (with their music) are used for the next step



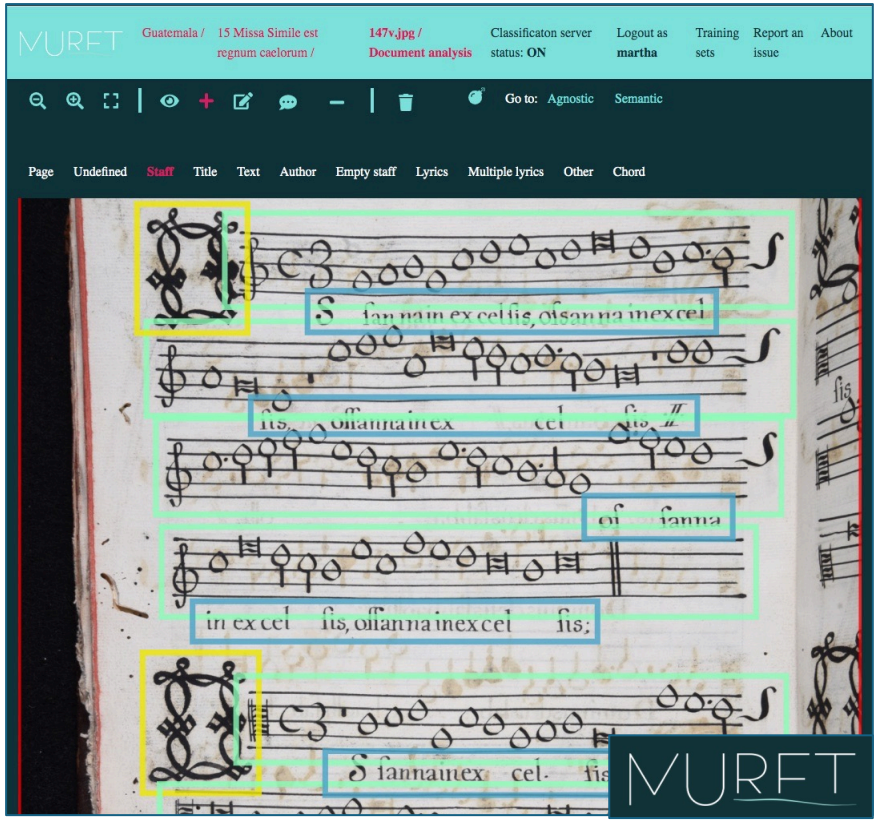


Step 1: Preprocessing or Document Analysis or Document Segmentation

- There is always a **pre-processing step** that **prepares the data** needed for the **following steps** in the pipeline
- Normally, this step segments the document into layers: music symbols & staves & text layers
- Sometimes the music symbols are not separated from their staff and the staff regions (with their music) are used for the next step

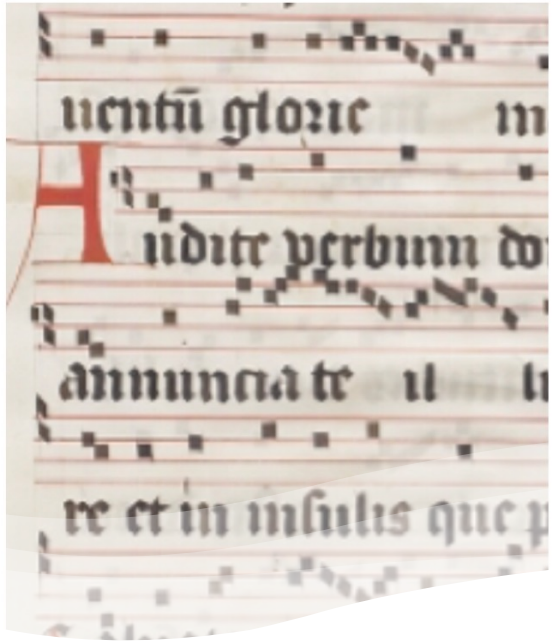


Step 1: Preprocessing or Document Analysis or Document Segmentation

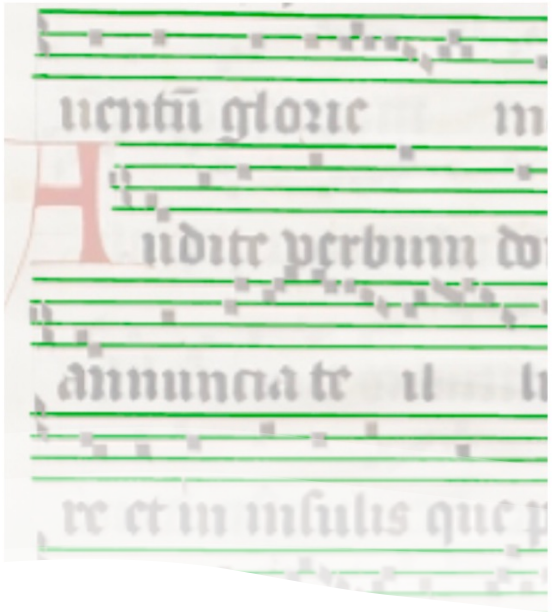


- There is always a **pre-processing step** that **prepares the data** needed for the **following steps** in the pipeline
- Normally, this step segments the document into layers: music symbols & staves & text layers
- Sometimes the music symbols are not separated from their staff and the staff regions (with their music) are used for the next step

Actual

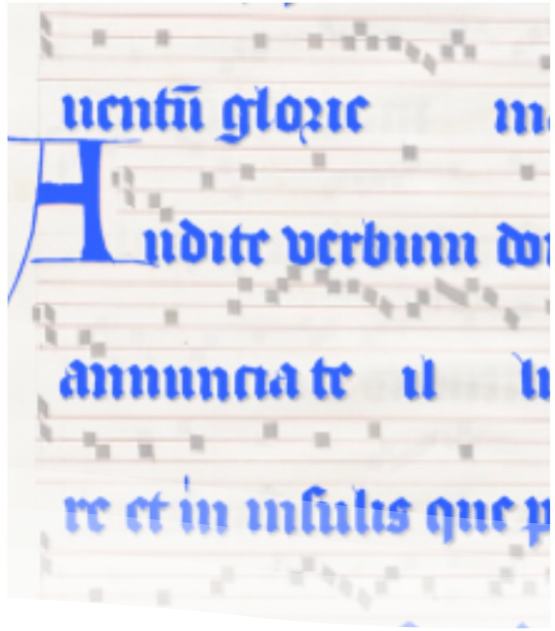


Staff



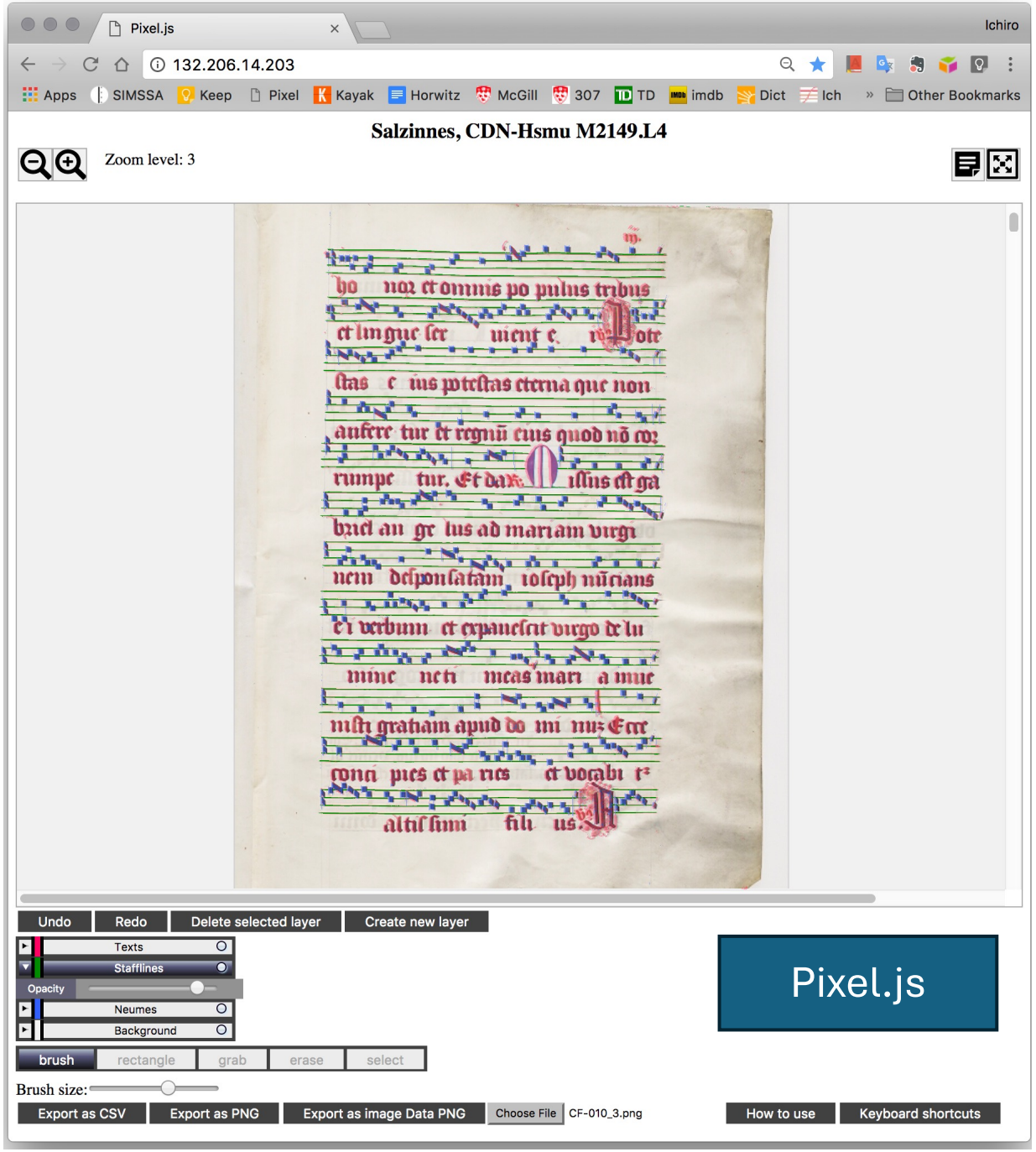
Note

Text



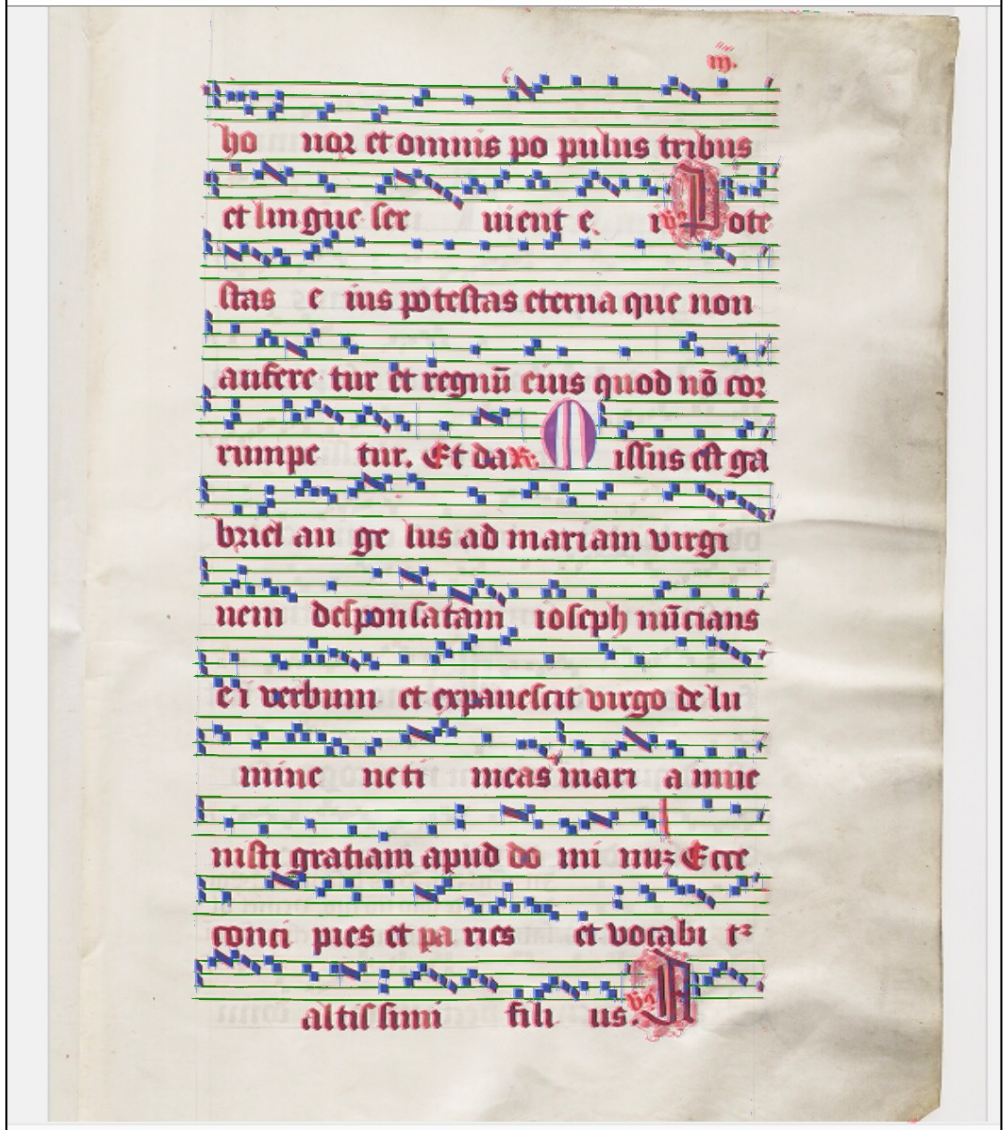
Step 1: Preprocessing or Document Analysis or Document Segmentation

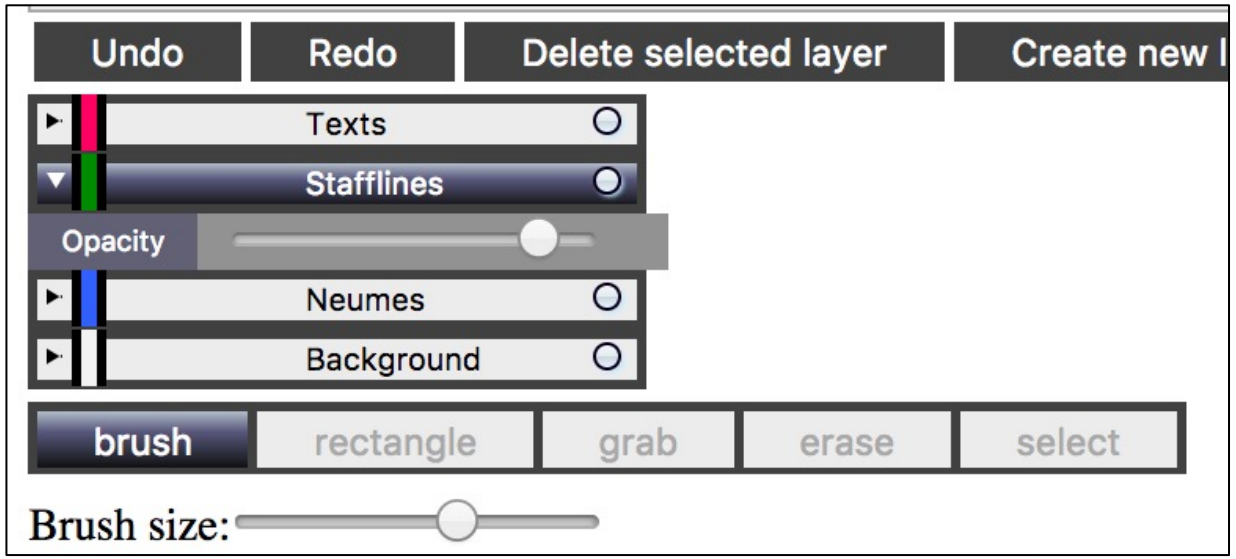
- There is always a **pre-processing step** that **prepares the data** needed for the **following steps** in the pipeline
- Normally, this step segments the document into layers: music symbols & staves & text layers
- Train the computer to do this layer separation



- There is always a **pre-processing step** that **prepares the data** needed for the **following steps** in the pipeline
- Normally, this step segments the document into layers: music symbols & staves & text layers
- Train the computer to do this layer separation

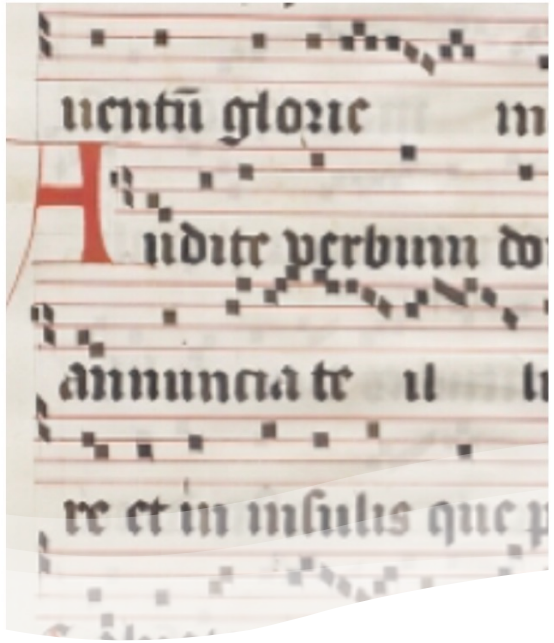
Salzinnes, CDN-Hsmu M2149.L4



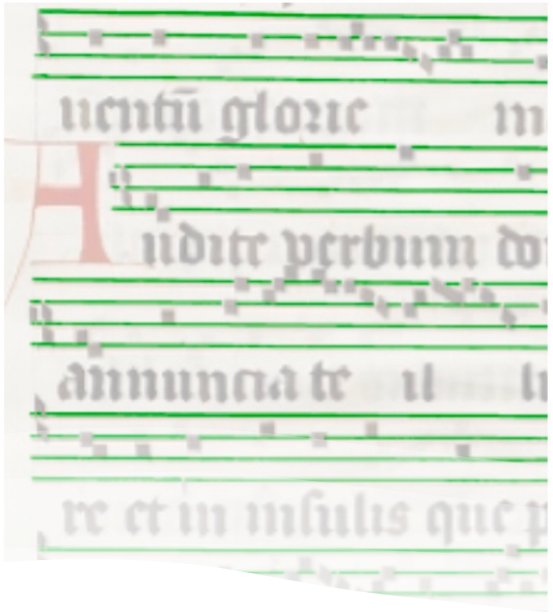


- There is always a **pre-processing step** that **prepares the data** needed for the **following steps** in the pipeline
- Normally, this step segments the document into layers: music symbols & staves & text layers
- Train the computer to do this layer separation

Actual

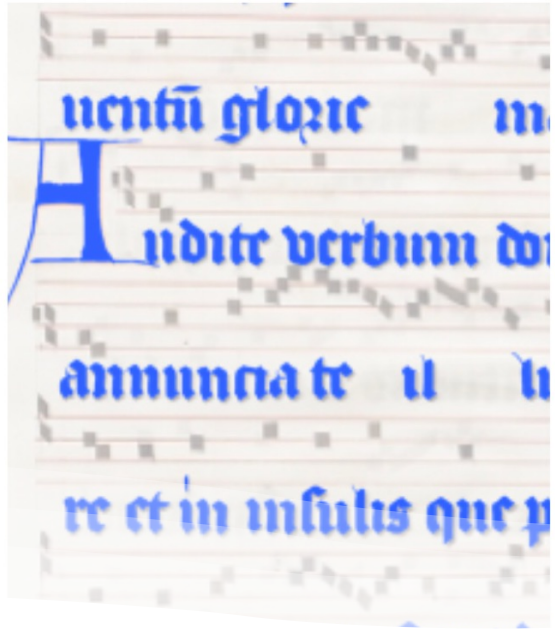


Staff



Note

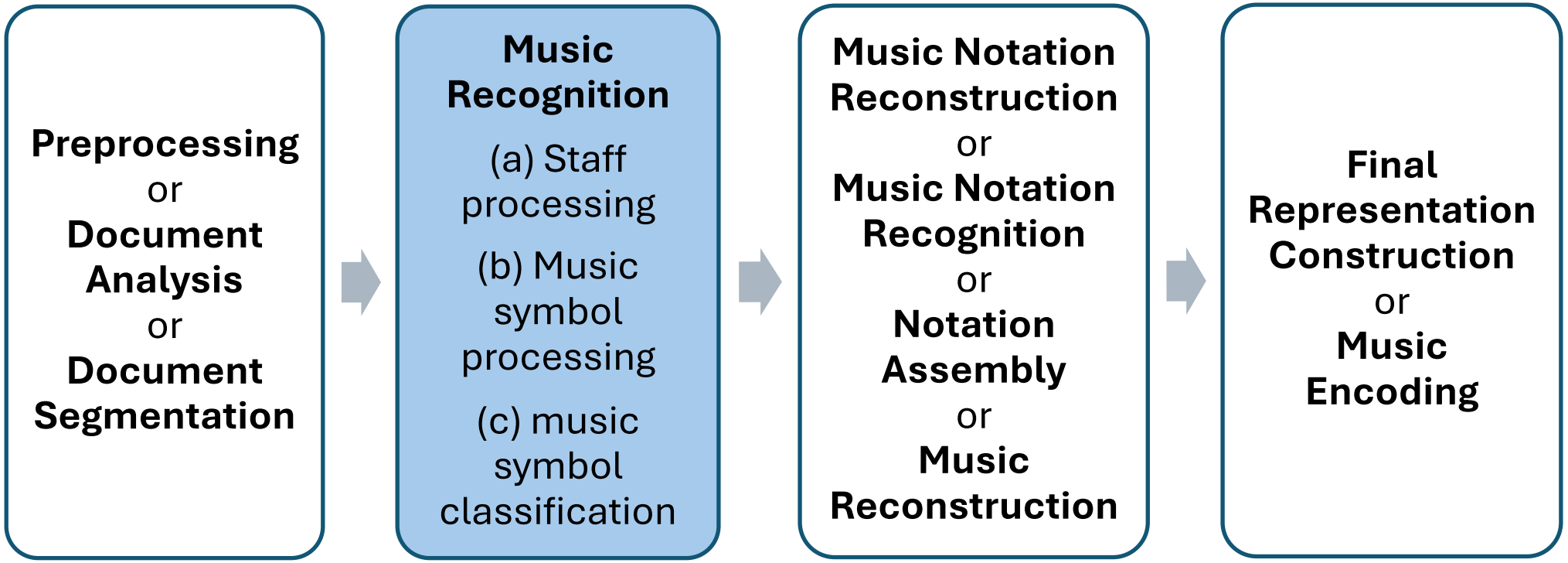
Text



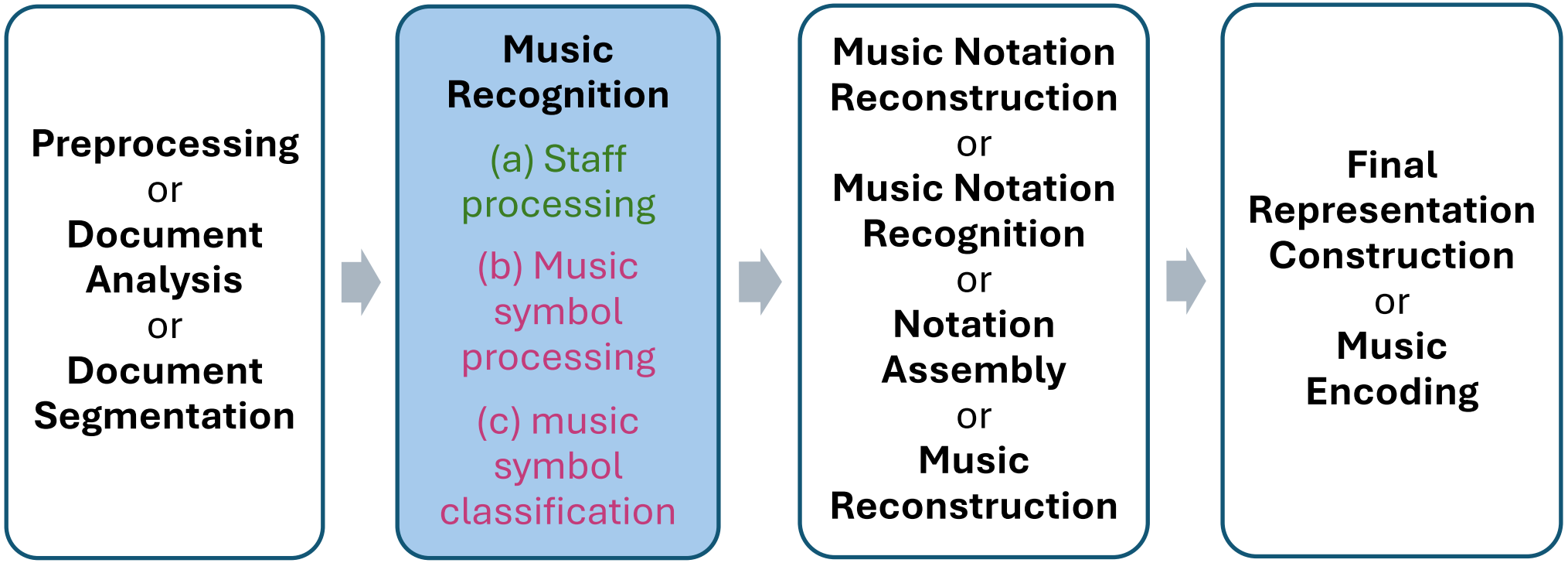
Step 1: Preprocessing or Document Analysis or Document Segmentation

- There is always a **pre-processing step** that **prepares the data** needed for the **following steps** in the pipeline
- Normally, this step segments the document into layers: music symbols & staves & text layers
- Train the computer to do this layer separation

Step 2: Music Recognition



Step 2: Music Recognition



Staff

Note

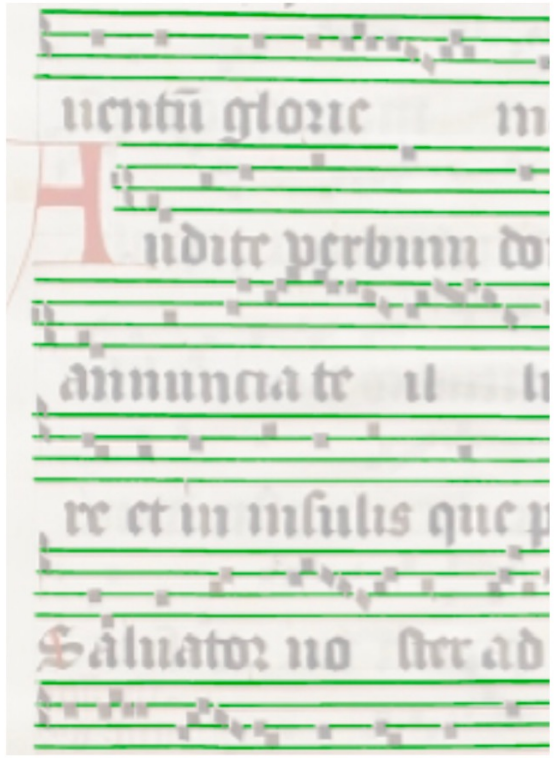
Step 2: Music Recognition

(a) Staff processing

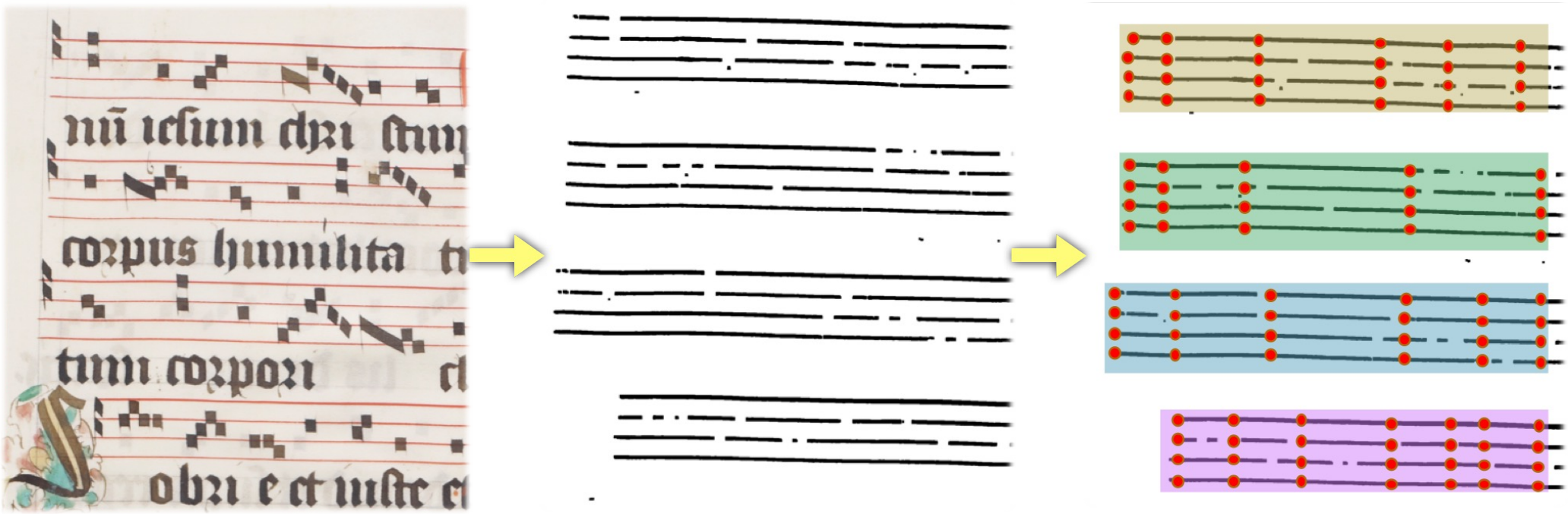
(b) Music symbol processing &

(c) Music symbol classification

Staff



Step 2: Music Recognition **(a)** Staff processing

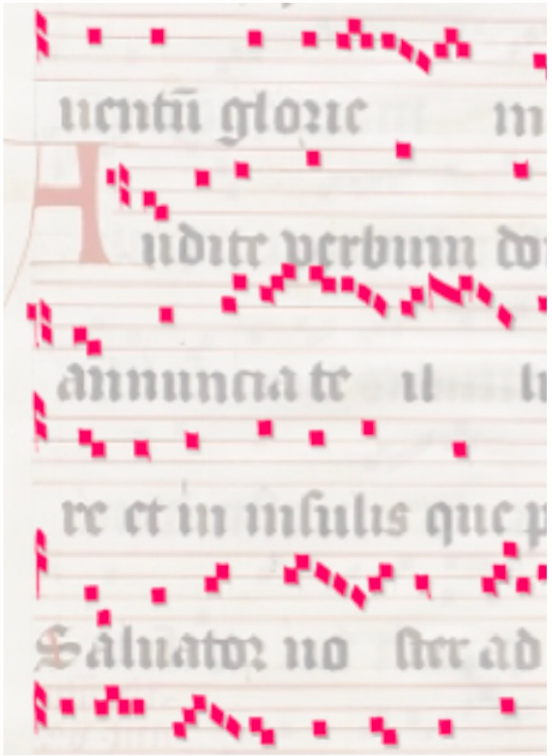


Step 2: Music Recognition

(a) Staff processing

- Track the staff line path

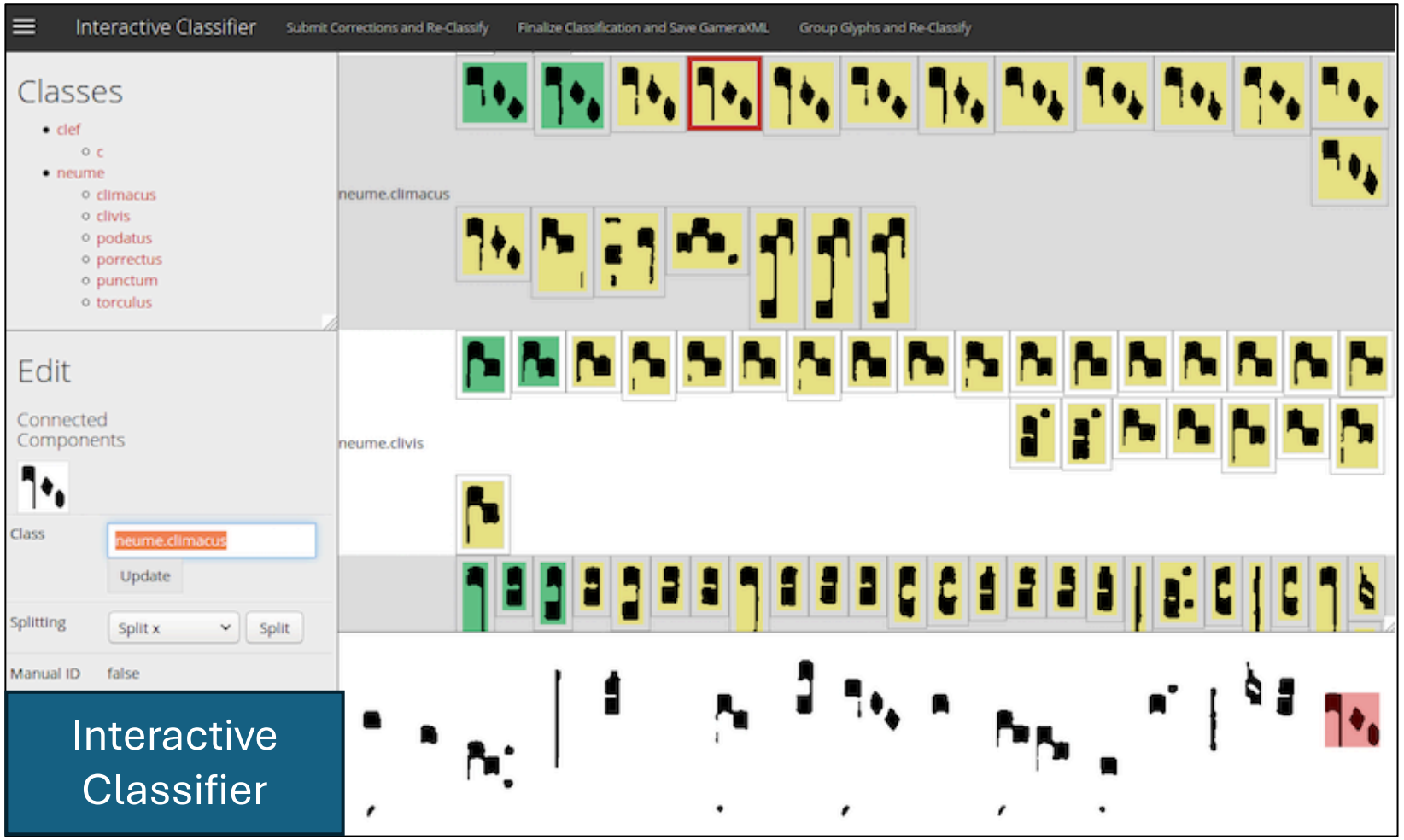
Note



Step 2: Music Recognition

(b) Music symbol processing

- “Detection” of music symbols



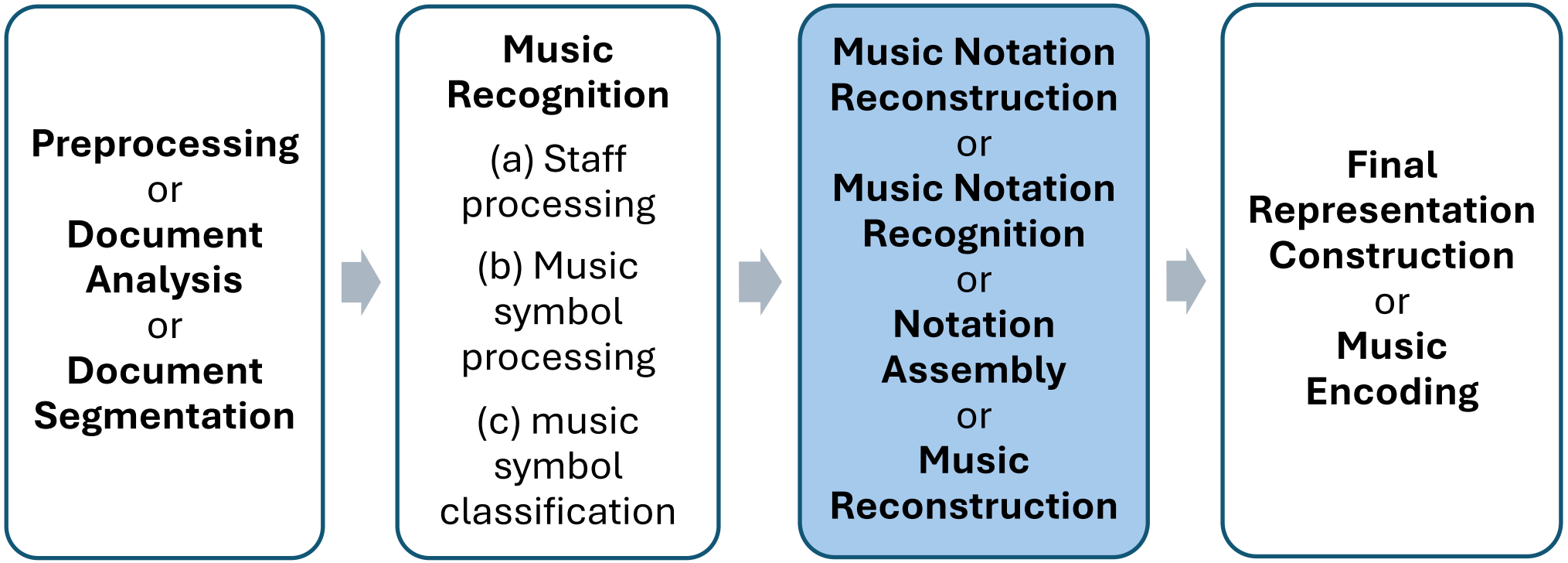
Step 2: Music Recognition

(b) Music symbol processing

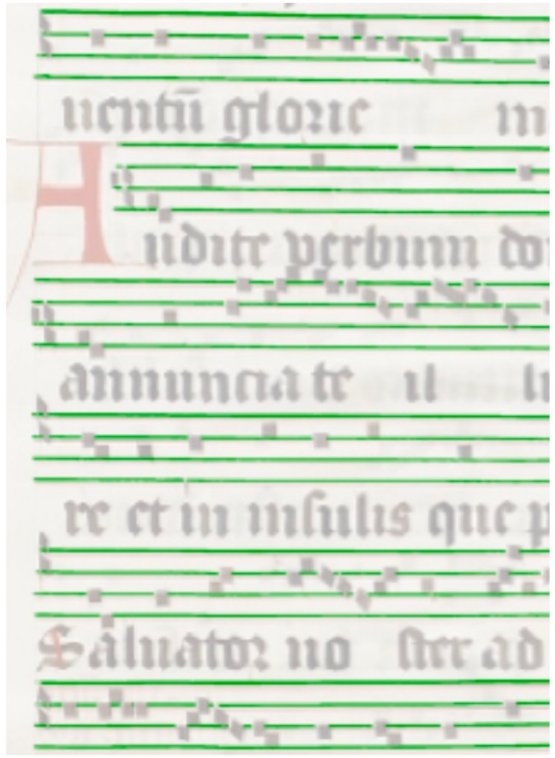
- "Detection" of music symbols

(c) Music symbol **classification**

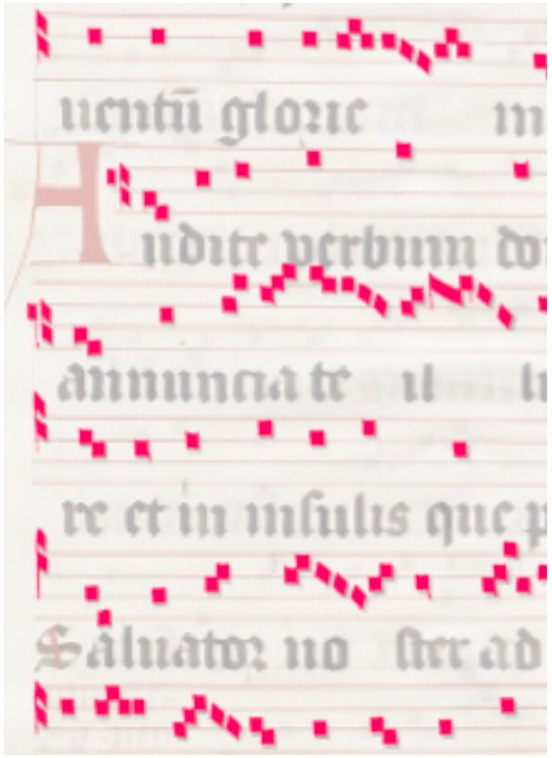
Step 3: Music Notation Reconstruction



Staff



Note

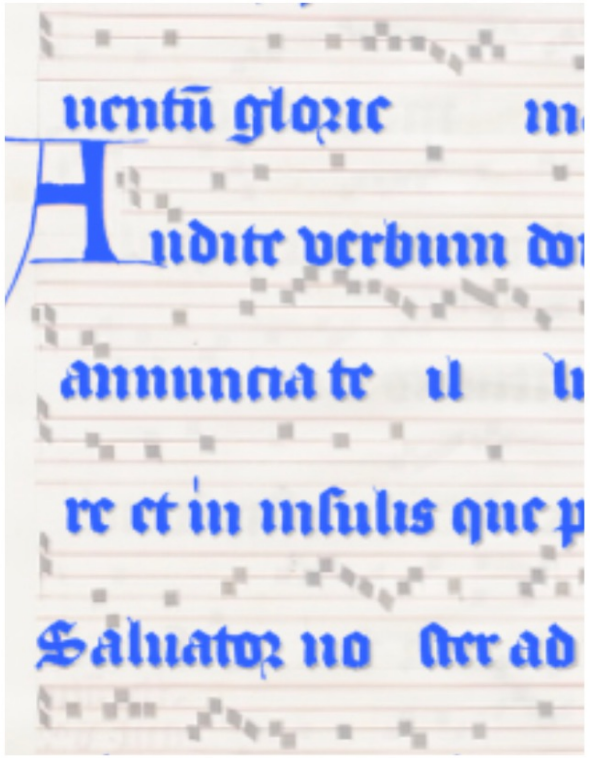


Step 3: Music Notation Reconstruction
also known as "notation assembly" or "music reconstruction"

Take the various individual sources of information recognized on previous steps (e.g., music symbols and staff lines) and bring them together to reconstruct the "music notation" (or "music semantics")

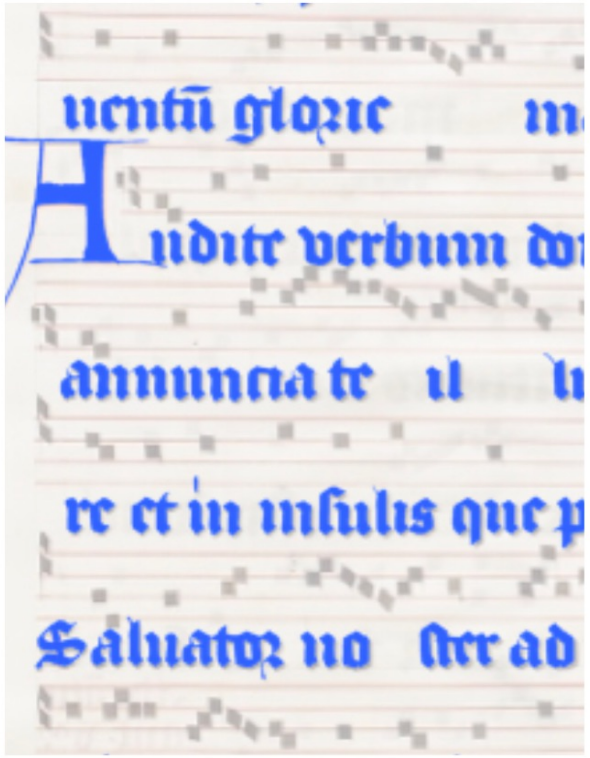
→ **Pitch detection**

Overlay the symbols on top of the staff lines and use the tracked staff lines path to interpret the pitch of each symbol



What about text?

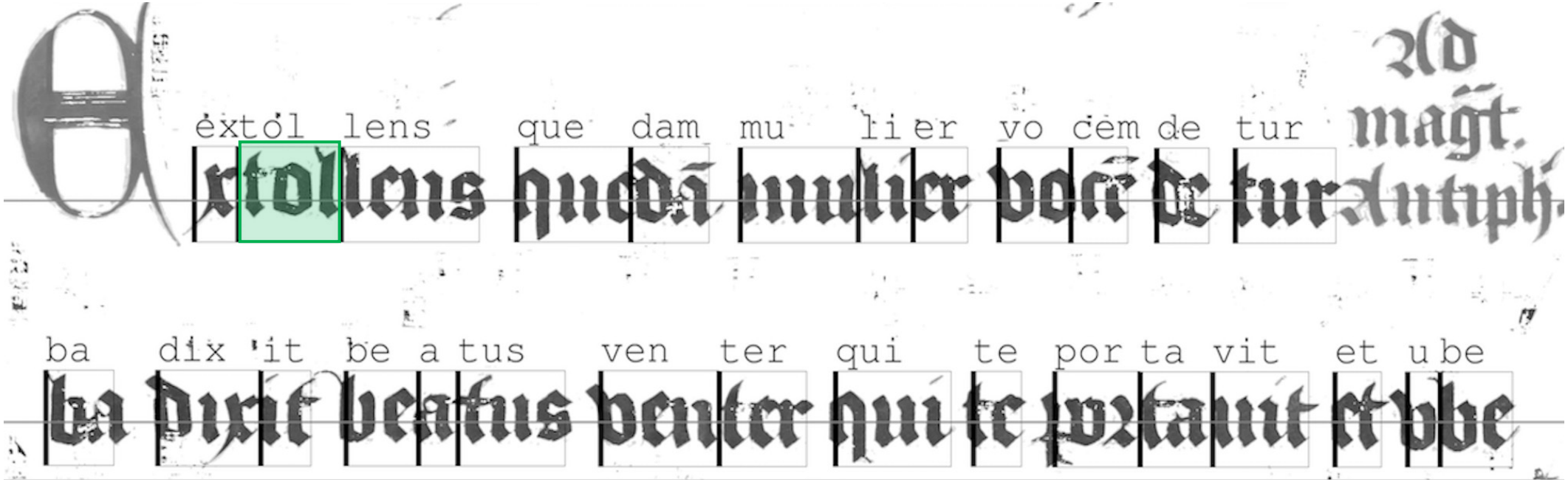
Step 3: Music Notation Reconstruction
also known as "notation assembly" or "music reconstruction"



What about text?

Step 3: Music Notation Reconstruction
also known as "notation assembly" or "music reconstruction"

- OCR has issues with recognizing handwritten text



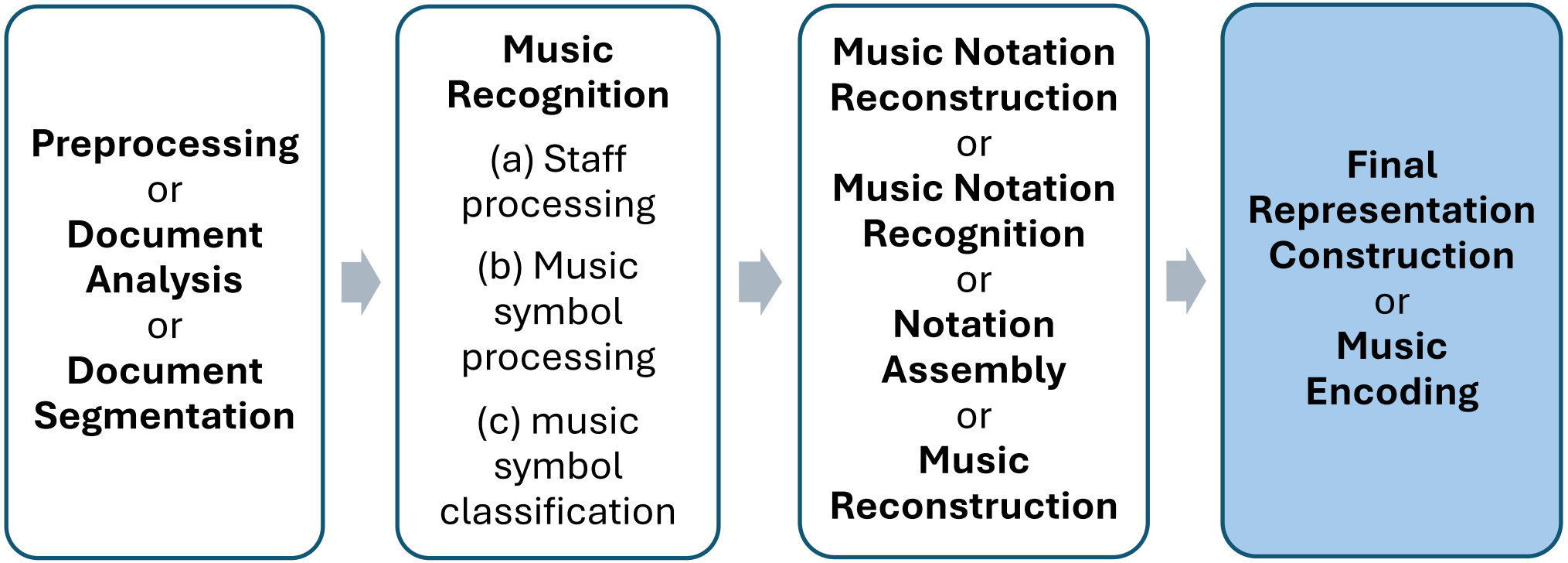
From Cantus Database

Extollens quaedam mulier vocem de tur-
ba dixit beatus venter qui te portavit et ube…

Step 3: Music Notation Reconstruction
also known as "notation assembly" or "music reconstruction"

- OCR has issues with recognizing handwritten text

Step 4: Final Representation or Music Encoding



```
<section xml:id="section-0000001229415468">
    <measure xml:id="measure-L6" n="1">
        <staff xml:id="staff-L6F2N1" n="1">
            <layer xml:id="layer-L6F2N1" n="1">
                <rest xml:id="rest-L7F2" dur="2" fermata="above" />
                <note xml:id="note-L11F2" dur="2" oct="4" pname="g" accid.ges="n" />
            </layer>
        </staff>
        <staff xml:id="staff-L6F1N1" n="2">
            <layer xml:id="layer-L6F1N1" n="1">
                <note xml:id="note-L7F1" dots="1" dur="4" oct="3" pname="c" accid.ges="n" />
                <tuplet xml:id="tuplet-L8F1-L10F1" num="3" numbase="2" num.format="count">
                    <beam xml:id="beam-L8F1-L10F1">
                        <note xml:id="note-L8F1" dur="16" oct="3" pname="d" accid="s" />
                        <note xml:id="note-L9F1" dur="16" oct="3" pname="e" accid="ff" />
                        <note xml:id="note-L10F1" dur="16" oct="3" pname="f" accid="x" />
                    </beam>
                </tuplet>
                <note xml:id="note-L11F1" dur="4" oct="3" pname="a" accid.ges="n" />
                <note xml:id="note-L12F1" dur="4" oct="3" pname="a" accid.ges="n" />
            </layer>
        </staff>
        <fermata xml:id="fermata-L10F1" staff="2" startid="#note-L10F1" place="above" />
        <tie xml:id="tie-L11F1-L12F1" startid="#note-L11F1" endid="#note-L12F1" />
        <slur xml:id="slur-L11F2-L16F3N1" staff="1" startid="#note-L11F2" endid="#note-L16F3" />
        <tie xml:id="tie-L12F1-L15F1" startid="#note-L12F1" endid="#note-L15F1" />
    </measure>
    <measure xml:id="measure-L13" n="2">
        <staff xml:id="staff-L13F2N1" n="1">
            <layer xml:id="layer-L13F2N1" n="1">
                <note xml:id="note-L15F3" dur="2" oct="4" pname="b" accid.ges="n" />
                <note xml:id="note-L16F3" dur="2" oct="5" pname="d" accid.ges="n" />
            </layer>
        </staff>
        <staff xml:id="staff-L13F1N1" n="2">
            <layer xml:id="layer-L13F1N1" n="1">
                <note xml:id="note-L15F1" dur="1" oct="3" pname="a" accid.ges="n" />
            </layer>
            <layer xml:id="layer-L15F2N2" n="2">
                <note xml:id="note-L15F2" dur="2" oct="2" pname="b" accid.ges="n" />
                <note xml:id="note-L16F2" dur="2" oct="2" pname="a" accid.ges="n" />
            </layer>
        </staff>
        <fermata xml:id="fermata-L16F3" staff="1" startid="#note-L16F3" place="above" />
    </measure>
    <measure xml:id="measure-L18" right="end" n="3">
        <staff xml:id="staff-L18F2N1" n="1">
            <layer xml:id="layer-L18F2N1" n="1">
                <note xml:id="note-L19F2" dur="4" oct="5" pname="c" accid.ges="n">
                    <artic xml:id="artic-L19F2" artic="marc" />
                </note>
                <note xml:id="note-L20F2" dur="4" oct="5" pname="d" accid.ges="n">
```

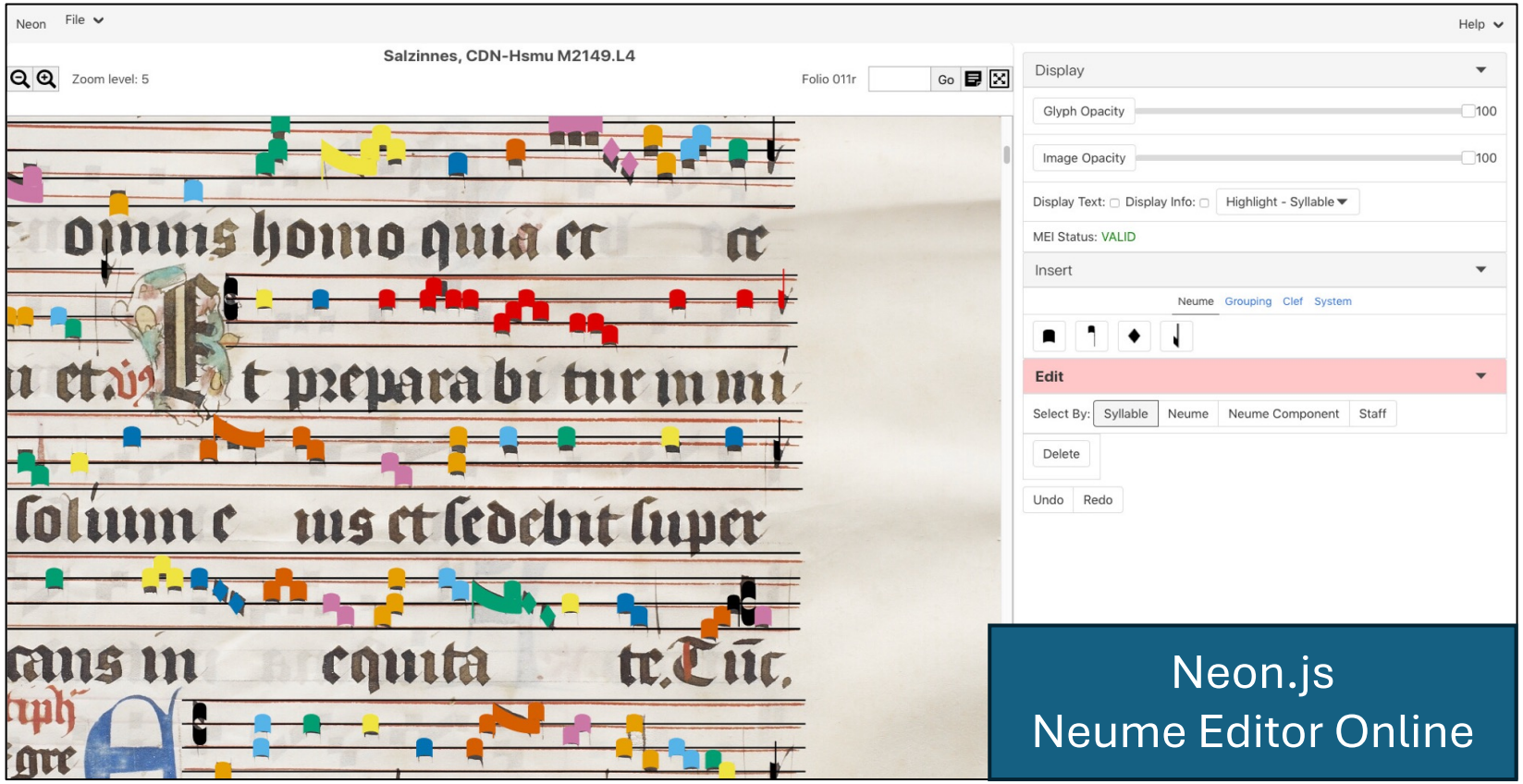
Step 4: Final Representation Construction (or Music Encoding)

- The music is encoded in a standard music-encoding format (e.g., MusicXML and MEI)

```
<section xml:id="section-0000001229415468">
    <measure xml:id="measure-L6" n="1">
        <staff xml:id="staff-L6F2N1" n="1">
            <layer xml:id="layer-L6F2N1" n="1">
                <rest xml:id="rest-L7F2" dur="2" fermata="above" />
                <note xml:id="note-L11F2" dur="2" oct="4" pname="g" accid.ges="n" />
            </layer>
        </staff>
        <staff xml:id="staff-L6F1N1" n="2">
            <layer xml:id="layer-L6F1N1" n="1">
                <note xml:id="note-L7F1" dots="1" dur="4" oct="3" pname="c" accid.ges="n" />
                <tuplet xml:id="tuplet-L8F1-L10F1" num="3" numbase="2" num.format="count">
                    <beam xml:id="beam-L8F1-L10F1">
                        <note xml:id="note-L8F1" dur="16" oct="3" pname="d" accid="s" />
                        <note xml:id="note-L9F1" dur="16" oct="3" pname="e" accid="ff" />
                        <note xml:id="note-L10F1" dur="16" oct="3" pname="f" accid="x" />
                    </beam>
                </tuplet>
                <note xml:id="note-L11F1" dur="4" oct="3" pname="a" accid.ges="n" />
                <note xml:id="note-L12F1" dur="4" oct="3" pname="a" accid.ges="n" />
            </layer>
        </staff>
        <fermata xml:id="fermata-L10F1" staff="2" startid="#note-L10F1" place="above" />
        <tie xml:id="tie-L11F1-L12F1" startid="#note-L11F1" endid="#note-L12F1" />
        <slur xml:id="slur-L11F2-L16F3N1" staff="1" startid="#note-L11F2" endid="#note-L16F3" />
        <tie xml:id="tie-L12F1-L15F1" startid="#note-L12F1" endid="#note-L15F1" />
    </measure>
    <measure xml:id="measure-L13" n="2">
        <staff xml:id="staff-L13F2N1" n="1">
            <layer xml:id="layer-L13F2N1" n="1">
                <note xml:id="note-L15F3" dur="2" oct="4" pname="b" accid.ges="n" />
                <note xml:id="note-L16F3" dur="2" oct="5" pname="d" accid.ges="n" />
            </layer>
        </staff>
        <staff xml:id="staff-L13F1N1" n="2">
            <layer xml:id="layer-L13F1N1" n="1">
                <note xml:id="note-L15F1" dur="1" oct="3" pname="a" accid.ges="n" />
            </layer>
            <layer xml:id="layer-L15F2N2" n="2">
                <note xml:id="note-L15F2" dur="2" oct="2" pname="b" accid.ges="n" />
                <note xml:id="note-L16F2" dur="2" oct="2" pname="a" accid.ges="n" />
            </layer>
        </staff>
        <fermata xml:id="fermata-L16F3" staff="1" startid="#note-L16F3" place="above" />
    </measure>
    <measure xml:id="measure-L18" right="end" n="3">
        <staff xml:id="staff-L18F2N1" n="1">
            <layer xml:id="layer-L18F2N1" n="1">
                <note xml:id="note-L19F2" dur="4" oct="5" pname="c" accid.ges="n">
                    <artic xml:id="artic-L19F2" artic="marc" />
                </note>
                <note xml:id="note-L20F2" dur="4" oct="5" pname="d" accid.ges="n">
```

Step 4: Final Representation Construction (or Music Encoding)

- The music is encoded in a standard music-encoding format (e.g., MusicXML and MEI)
- For early music notation → MEI



Step 4: Final Representation Construction (or Music Encoding)

- The music is encoded in a standard music-encoding format
- For early music notation → MEI
- Interfaces for correction of the resulting MEI file (e.g., Neon for neume notation)

Some OMR Frameworks for **Neume** Scripts

OMMR4all

Rodan and the **Single Interface for Music Score Searching & Analysis** (**SIMSSA**) Project

*Workflow manager created by **Andrew Hankinson**, used for OMR in the **SIMSSA** project, led by **Ichiro Fujinaga** (McGill)*

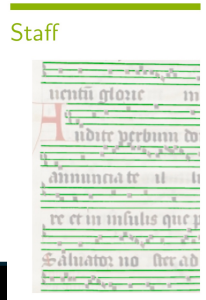
Rodan and the **Single Interface for Music Score Searching & Analysis** (**SIMSSA**) Project

*Workflow manager created by **Andrew Hankinson**, used for OMR in the **SIMSSA** project, led by **Ichiro Fujinaga** (McGill)*

Document Segmentation

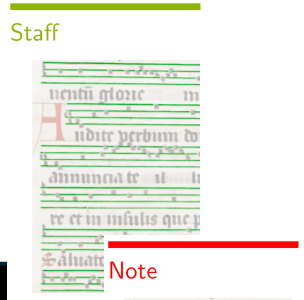
Rodan and the **Single Interface for Music Score Searching & Analysis** (**SIMSSA**) Project

*Workflow manager created by **Andrew Hankinson**, used for OMR in the **SIMSSA** project, led by **Ichiro Fujinaga** (McGill)*



Document Segmentation

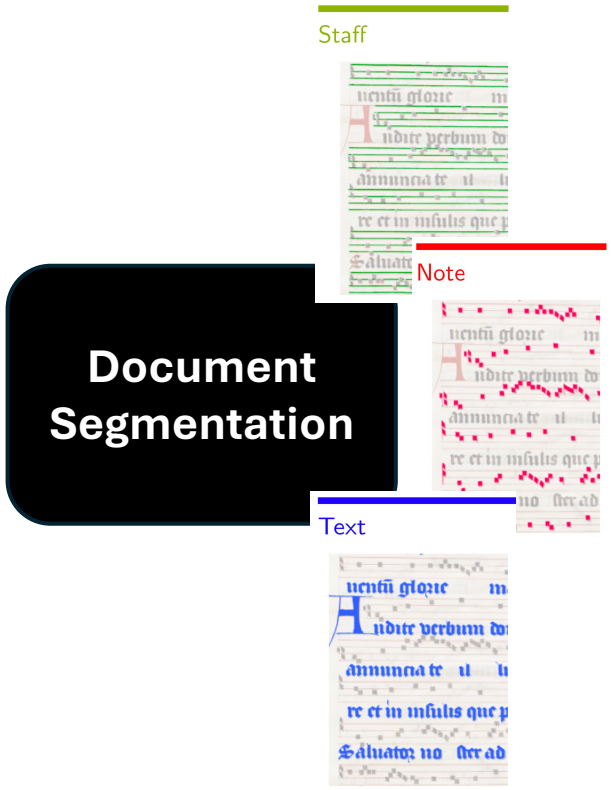
Rodan and the **Single Interface for Music Score Searching & Analysis (SIMSSA)** Project



Document Segmentation

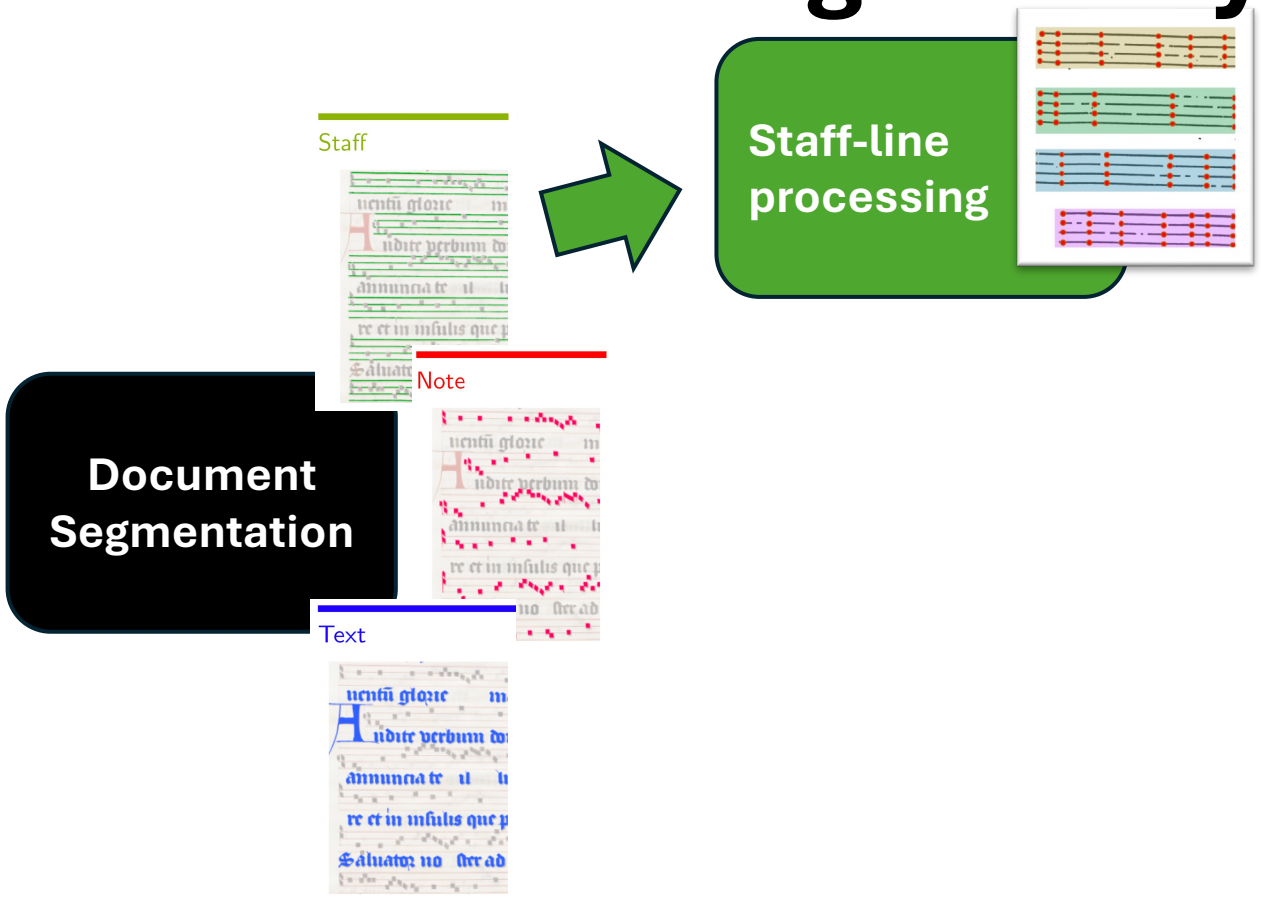
*Workflow manager created by **Andrew Hankinson**, used for OMR in the **SIMSSA** project, led by **Ichiro Fujinaga** (McGill)*

Rodan and the **Single Interface for Music Score Searching & Analysis (SIMSSA)** Project



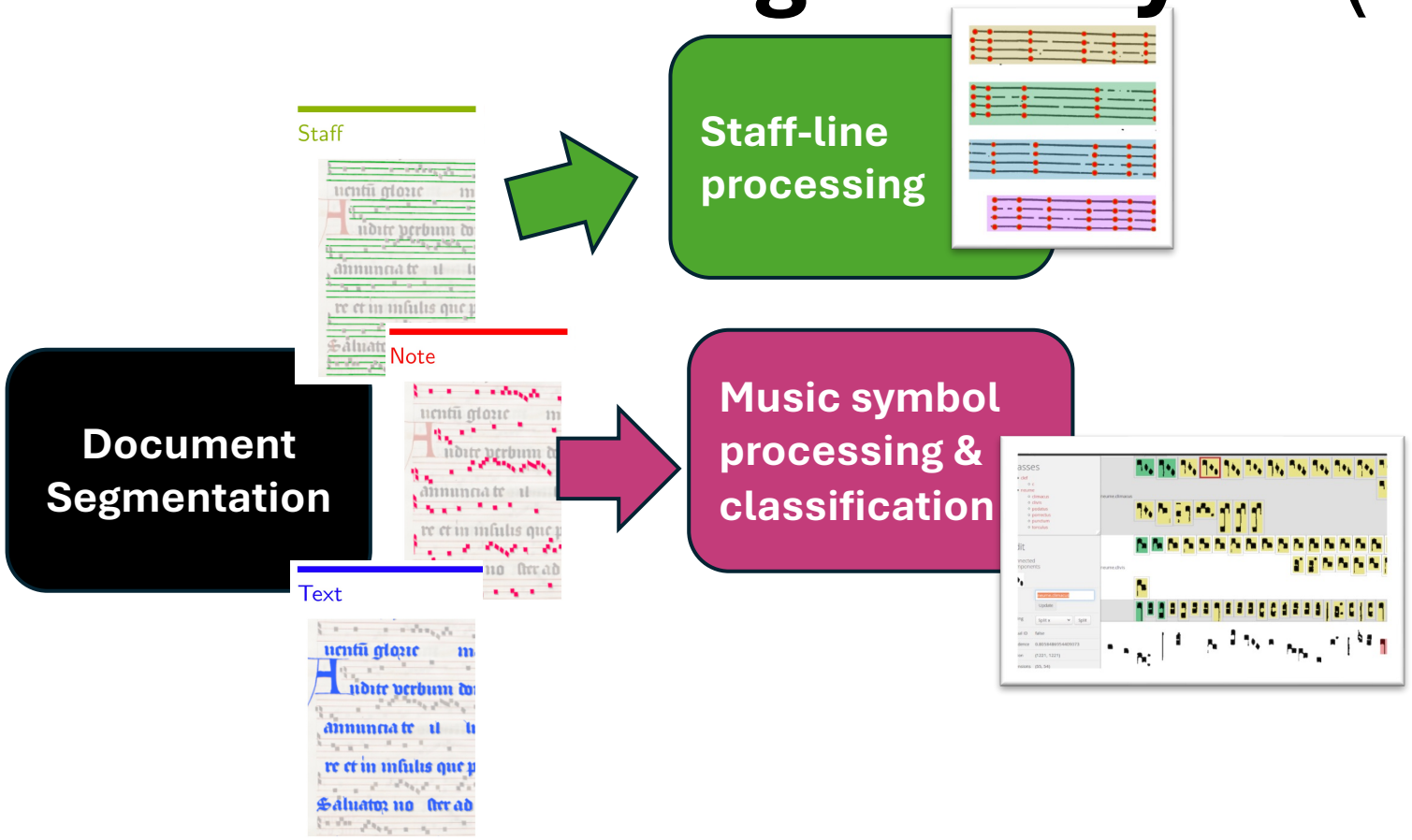
*Workflow manager created by **Andrew Hankinson**, used for OMR in the **SIMSSA** project, led by **Ichiro Fujinaga** (McGill)*

Rodan and the **Single Interface for Music Score Searching & Analysis (SIMSSA)** Project



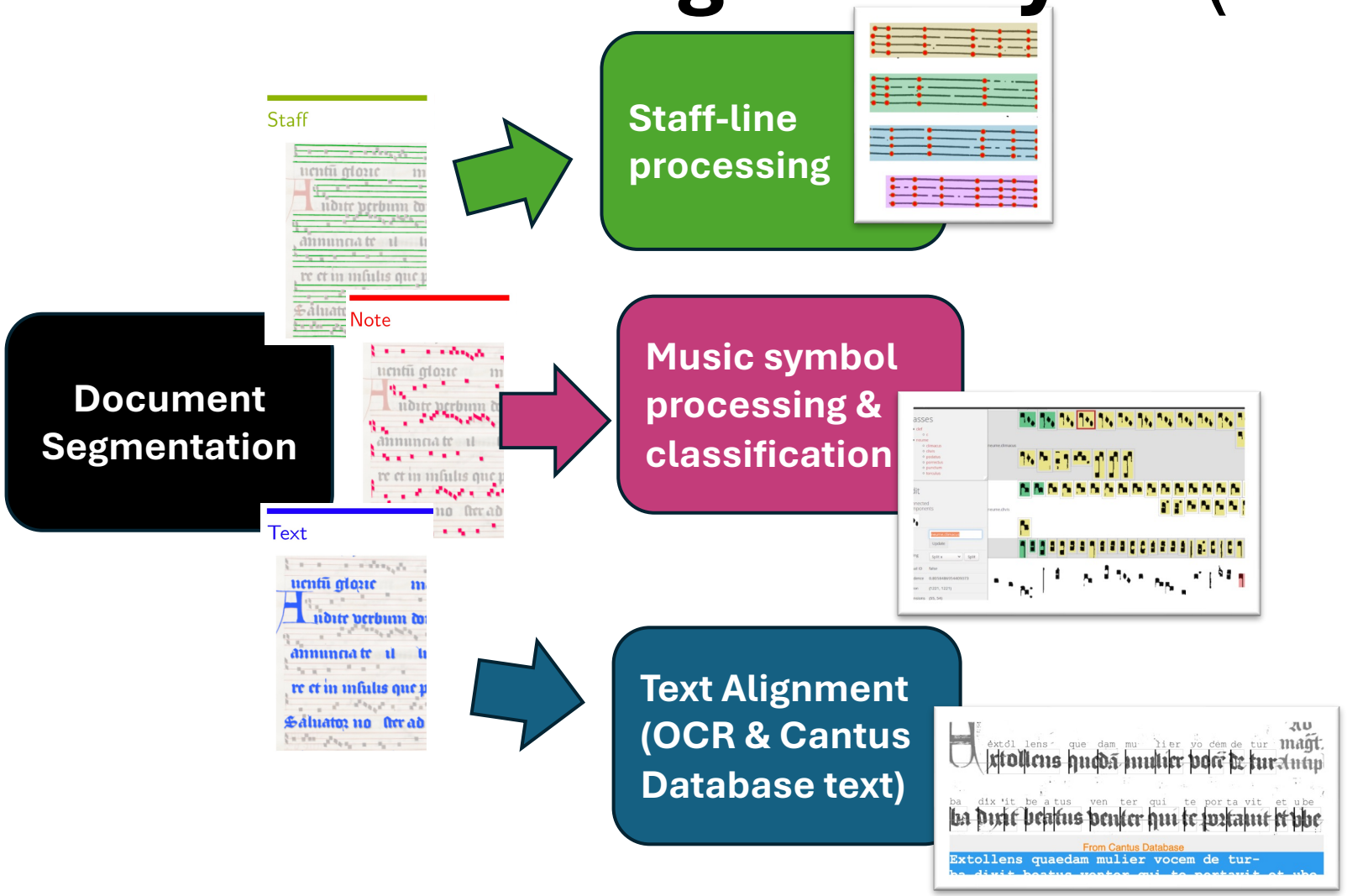
*Workflow manager created by **Andrew Hankinson**, used for OMR in the **SIMSSA** project, led by **Ichiro Fujinaga** (McGill)*

Rodan and the **Single Interface for Music Score Searching & Analysis** (**SIMSSA**) Project



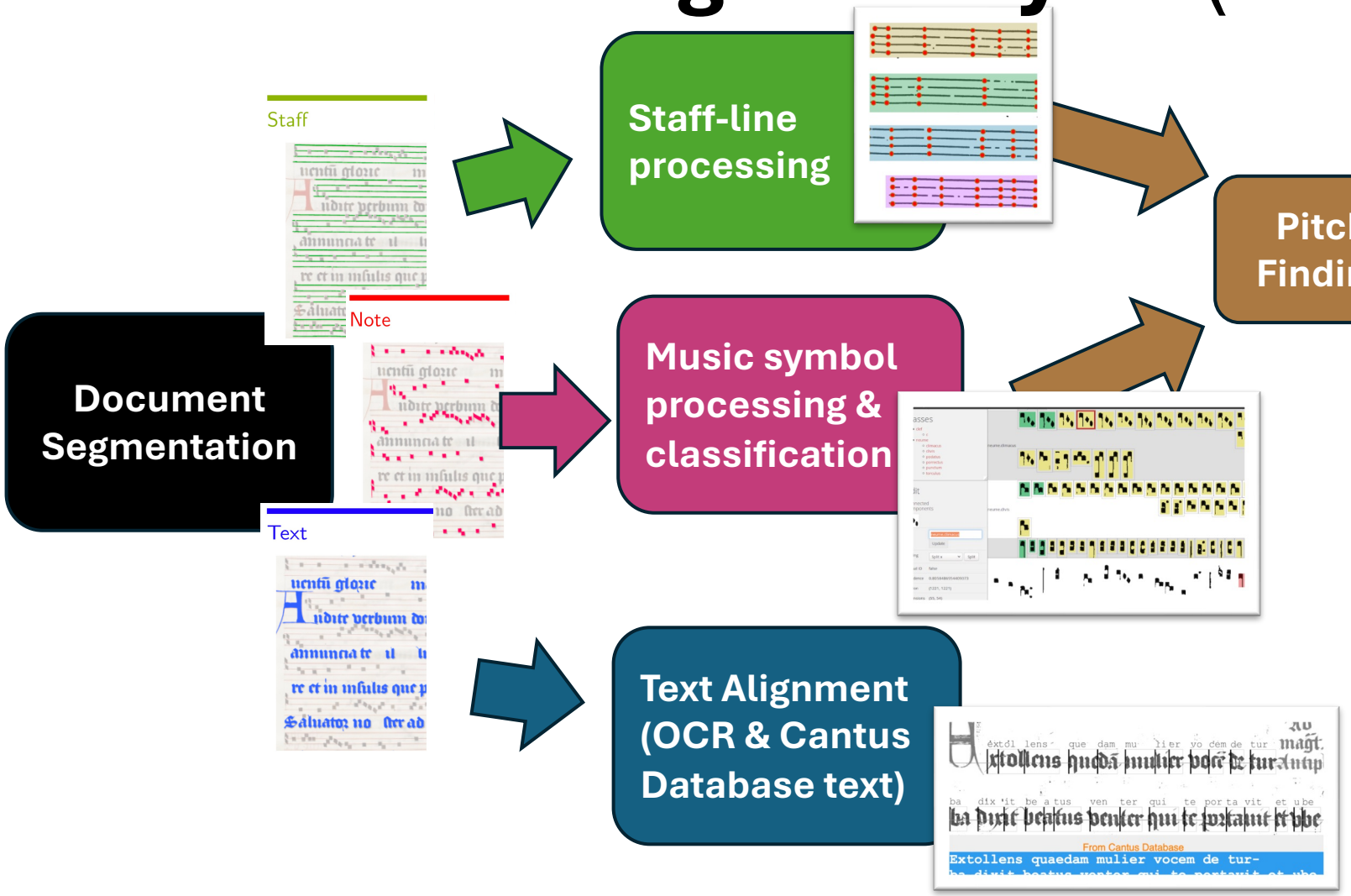
*Workflow manager created by **Andrew Hankinson**, used for OMR in the **SIMSSA** project, led by **Ichiro Fujinaga** (McGill)*

Rodan and the **Single Interface for Music Score Searching & Analysis** (**SIMSSA**) Project



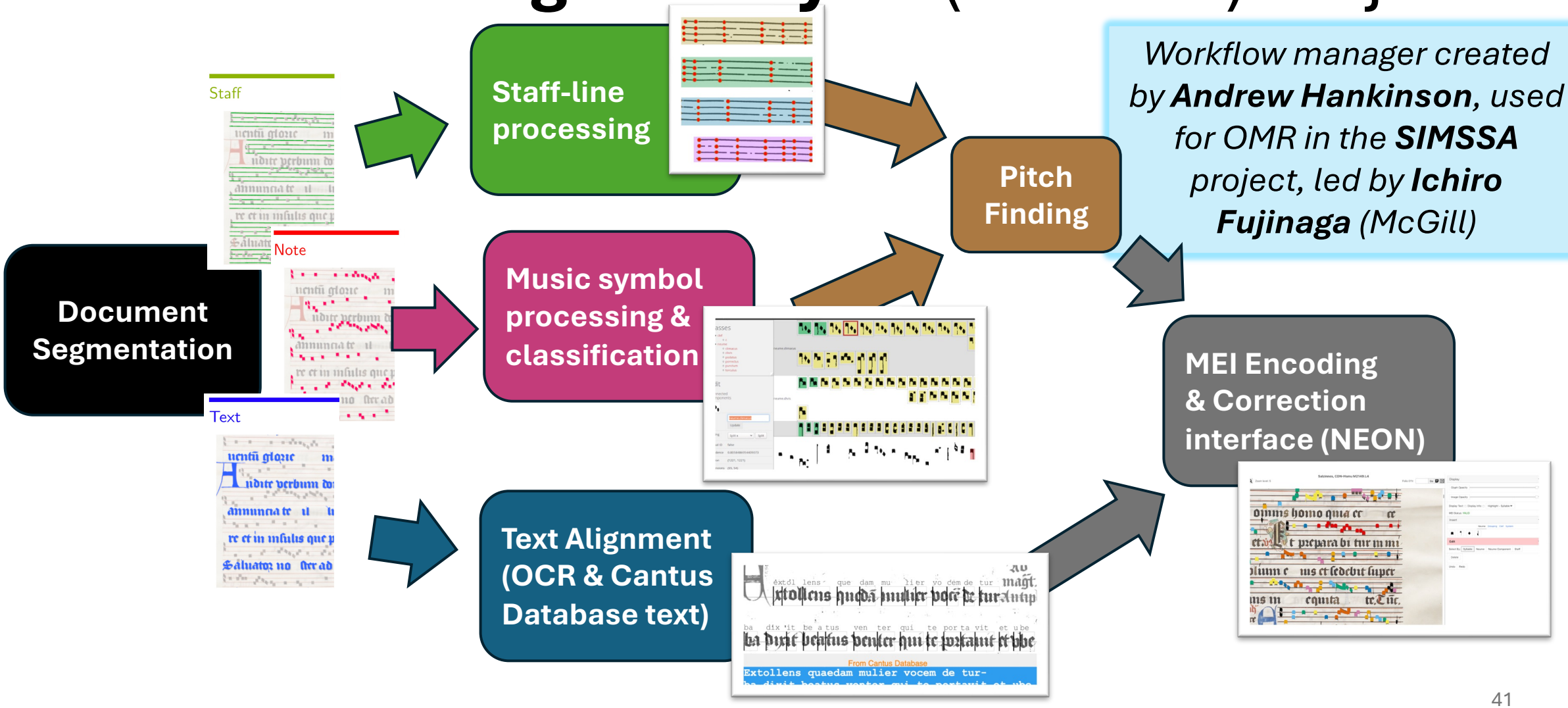
*Workflow manager created by **Andrew Hankinson**, used for OMR in the **SIMSSA** project, led by **Ichiro Fujinaga** (McGill)*

Rodan and the **Single Interface for Music Score Searching & Analysis (SIMSSA)** Project

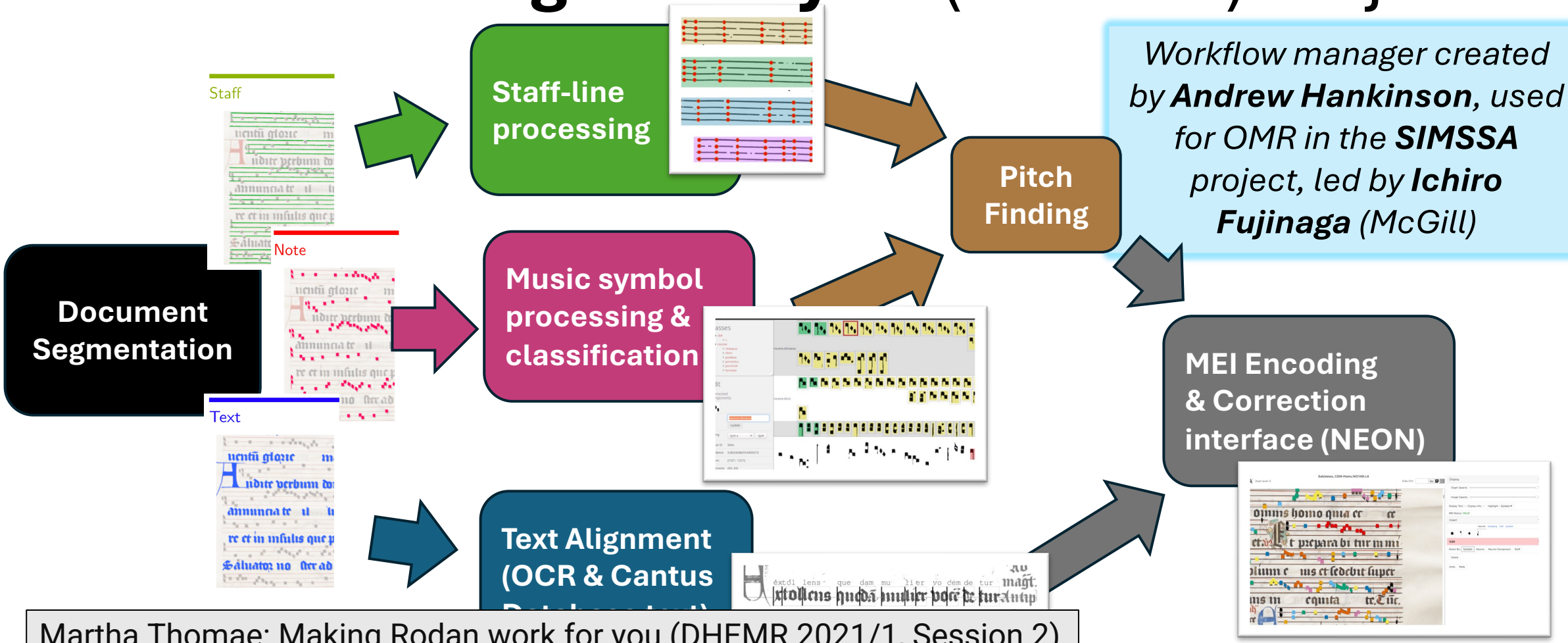


*Workflow manager created by **Andrew Hankinson**, used for OMR in the **SIMSSA** project, led by **Ichiro Fujinaga** (McGill)*

Rodan and the **Single Interface for Music Score Searching & Analysis (SIMSSA)** Project



Rodan and the **Single Interface for Music Score Searching & Analysis (SIMSSA)** Project

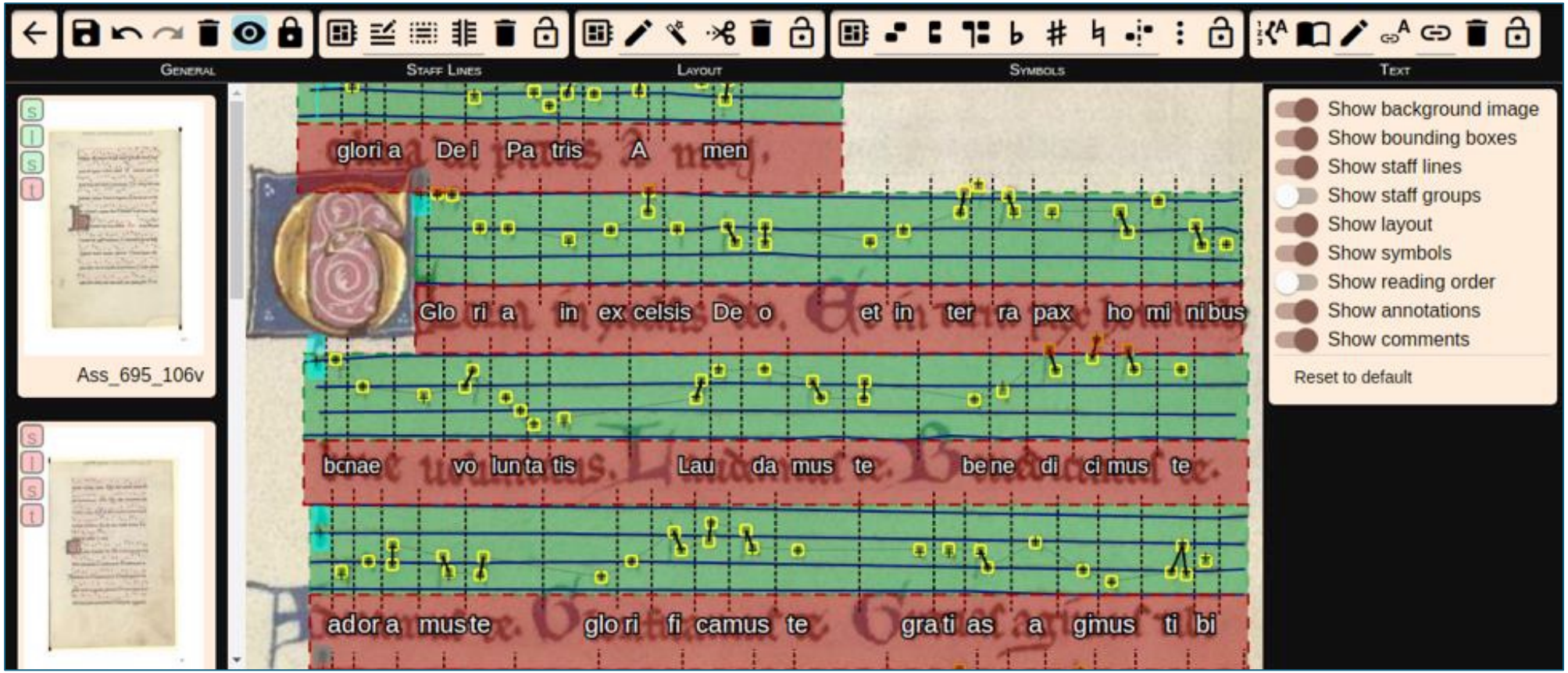


Martha Thomae: Making Rodan work for you (DHEMR 2021/1, Session 2)
https://youtu.be/_TeGXG9Fh2M?si=v_IQOBysIGi7GtLL

OMMR4All

*Created by **Cristopher Wick** (University of Würzburg), and it is being used in the **Corpus Monodicum** project*

general staff lines layout symbols text



Alexander Hartelt & Jan Hajič: Collaborative digital editions with OMMR4All (DHEMR 2021/1 sess. 3): https://youtu.be/hX9pGOdfbZ8?feature=shared

OMR Frameworks for Neume Notation

- **Old Myths, New Facts (OMNF).** Recordings available from first Digital Humanities in Early Music Research workshop of 2021. https://www.smnf.cz/en/news/Recordings-available-from-first-Digital-Humanities-in-Early-Music-Research-workshop-of-2021/

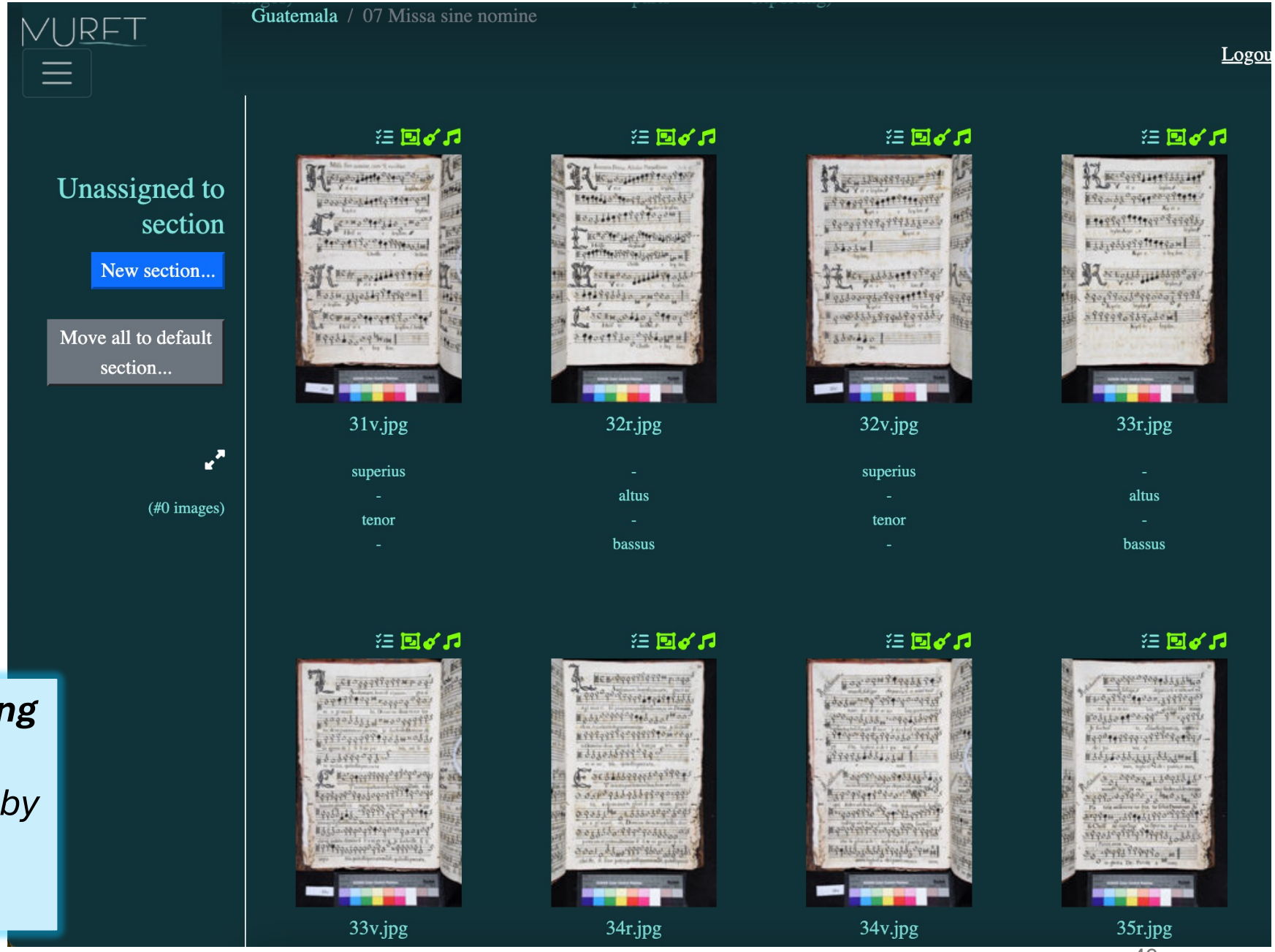
  The project "Old Myths, New Facts: Czech Lands in Center of 15-century Music Developments" (Czech Science Foundation EXPRO 19-28306X) focuses on the study of music repertory from 15th century Bohemia and its reflection in modern Czech music culture.

  - **Ichiro Fujinaga:** Large-scale OMR for neumes with Rodan (DHEMR 2021/1, Session 1). https://www.youtube.com/watch?v=FXkKrwatIwM
  - **Martha Thomae:** Making Rodan work for you (DHEMR 2021/1, Session 2). https://www.youtube.com/watch?v=_TeGXG9Fh2M
  - **Alexander Hartelt & Jan Hajič:** Collaborative digital editions with OMMR4All (DHEMR 2021/1 session 3). https://www.youtube.com/watch?v=hX9pGOdfbZ8

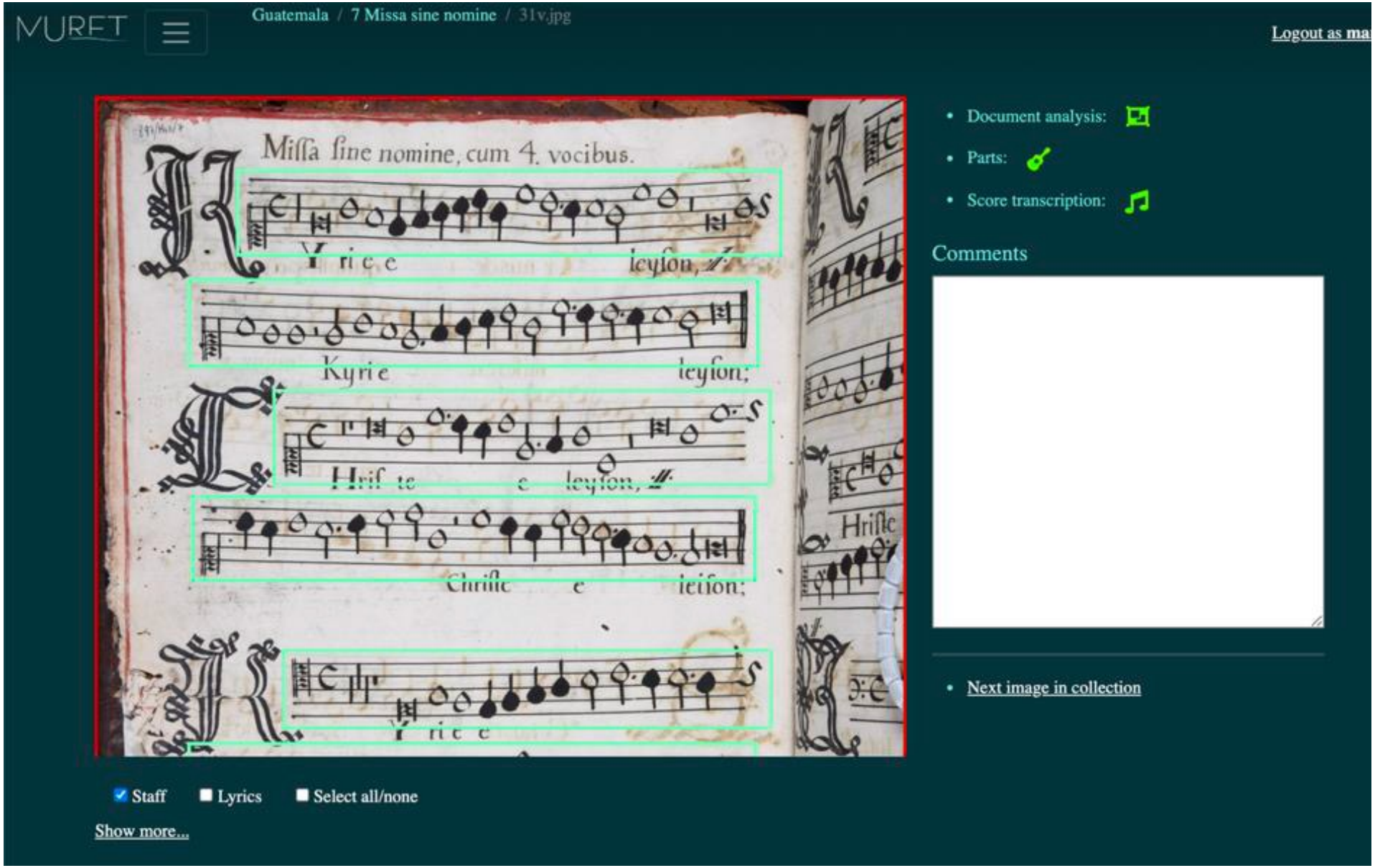
Some OMR Frameworks for **Mensural** Scripts

MuRET
The Piece Interface

Music Recognition Encoding and Transcription (MuRET)
OMR framework developed by ***David Rizo (University of Alicante)***

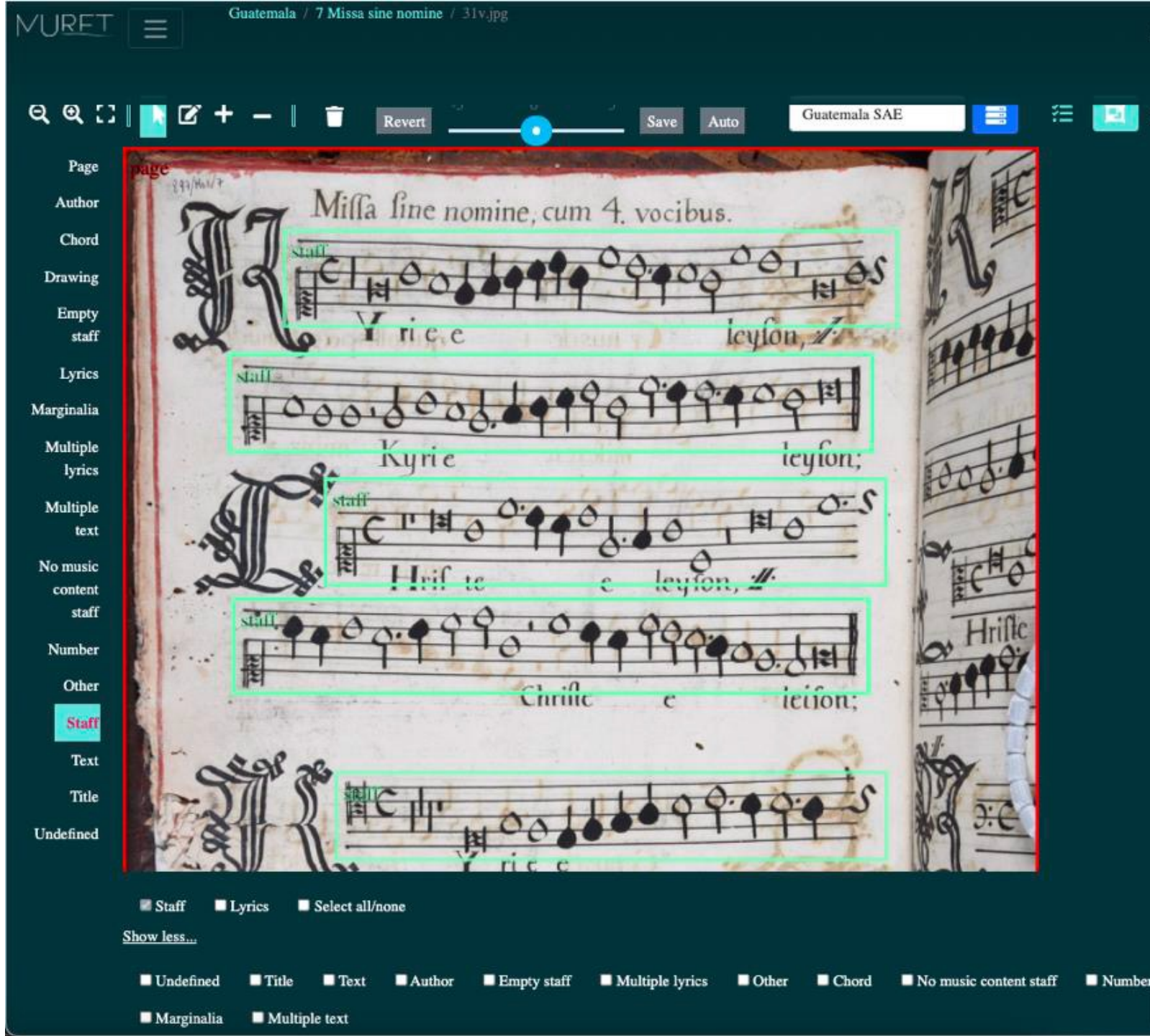


MuRET
The Page Interface

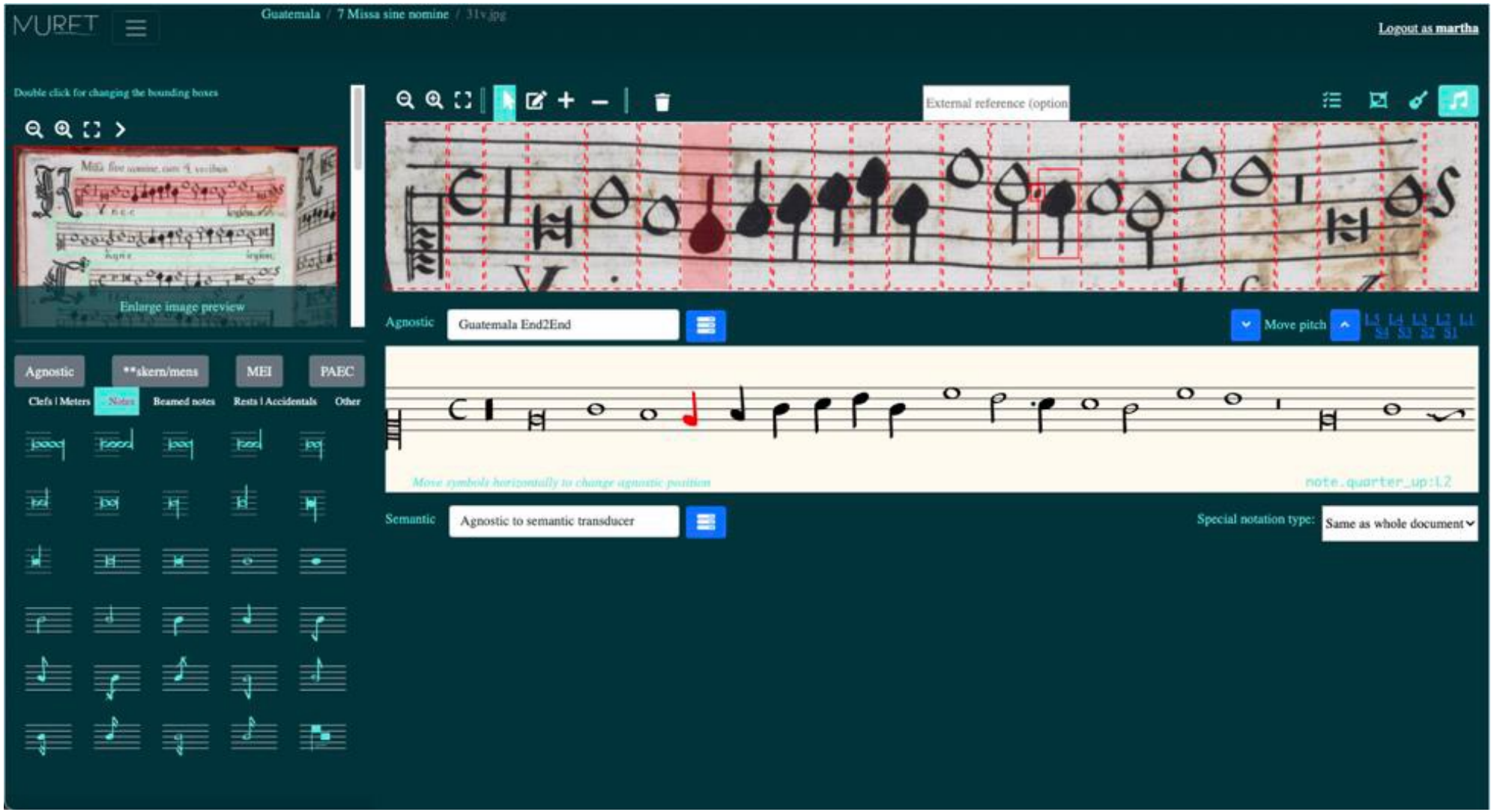


MuRET

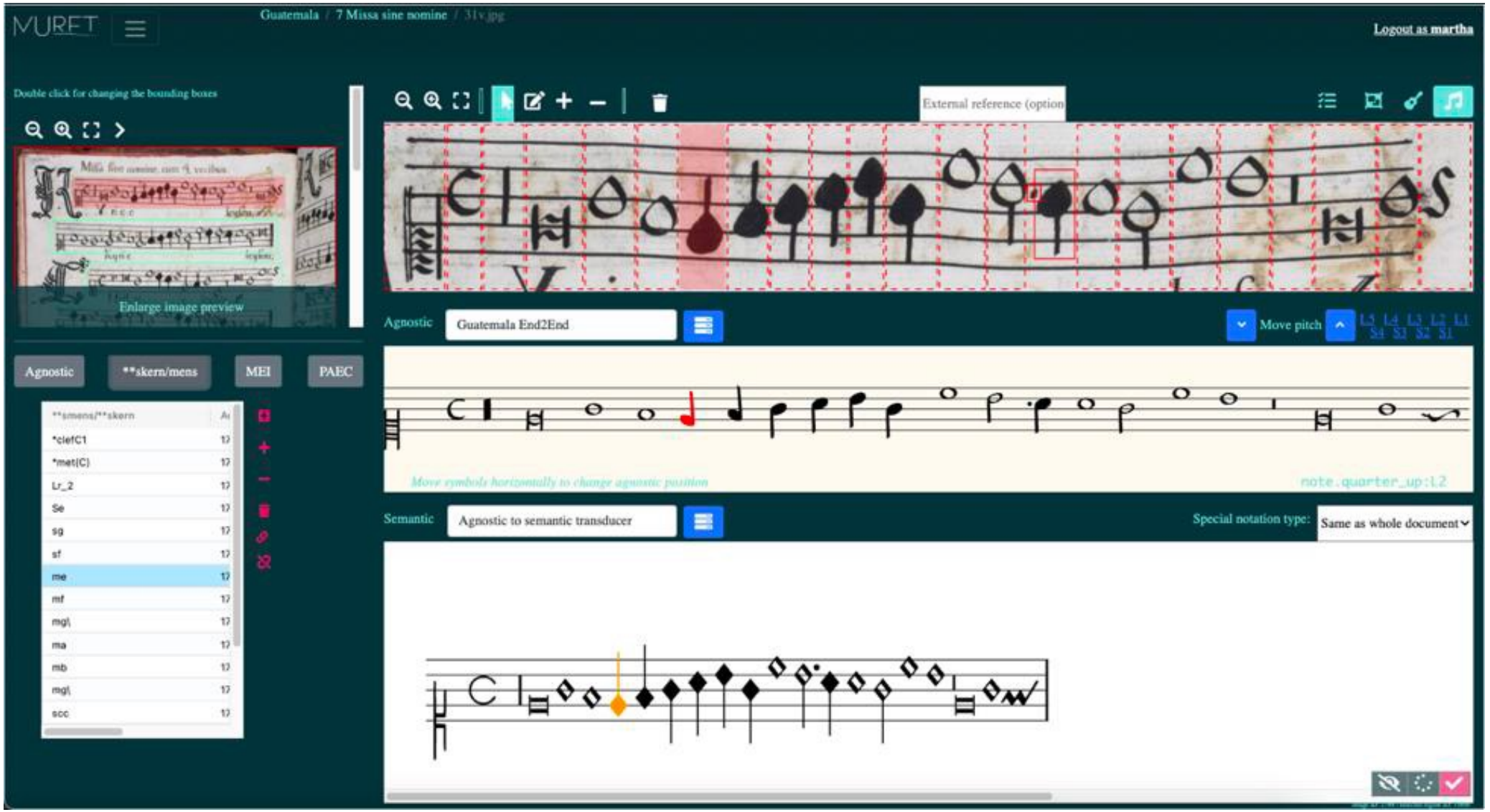
OMR Stage 1: Preprocessing or Document Analysis



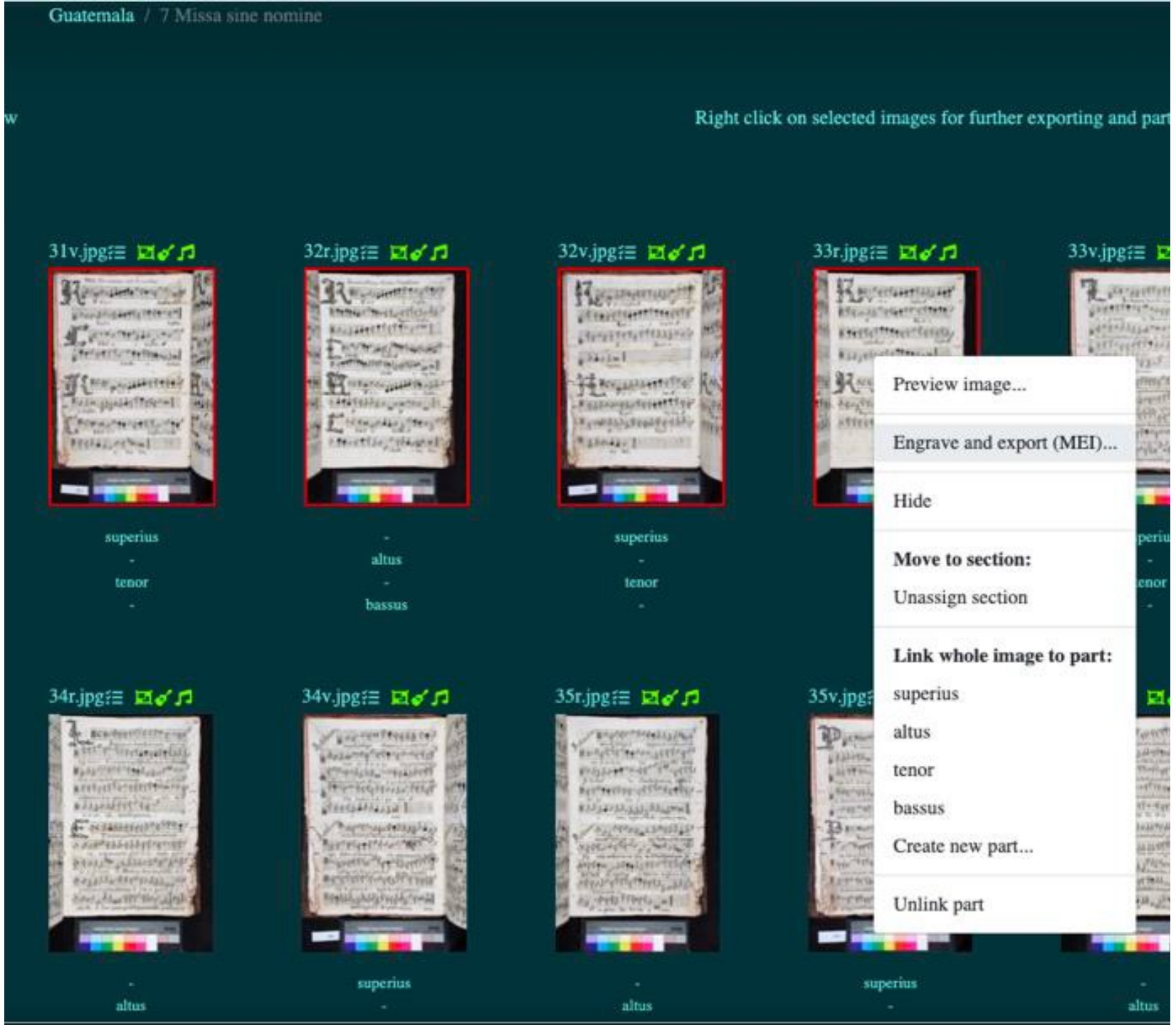
MuRET OMR Stage 2: Music Symbol Recognition



MuRET OMR Stage 3: Music Notation Reconstruction

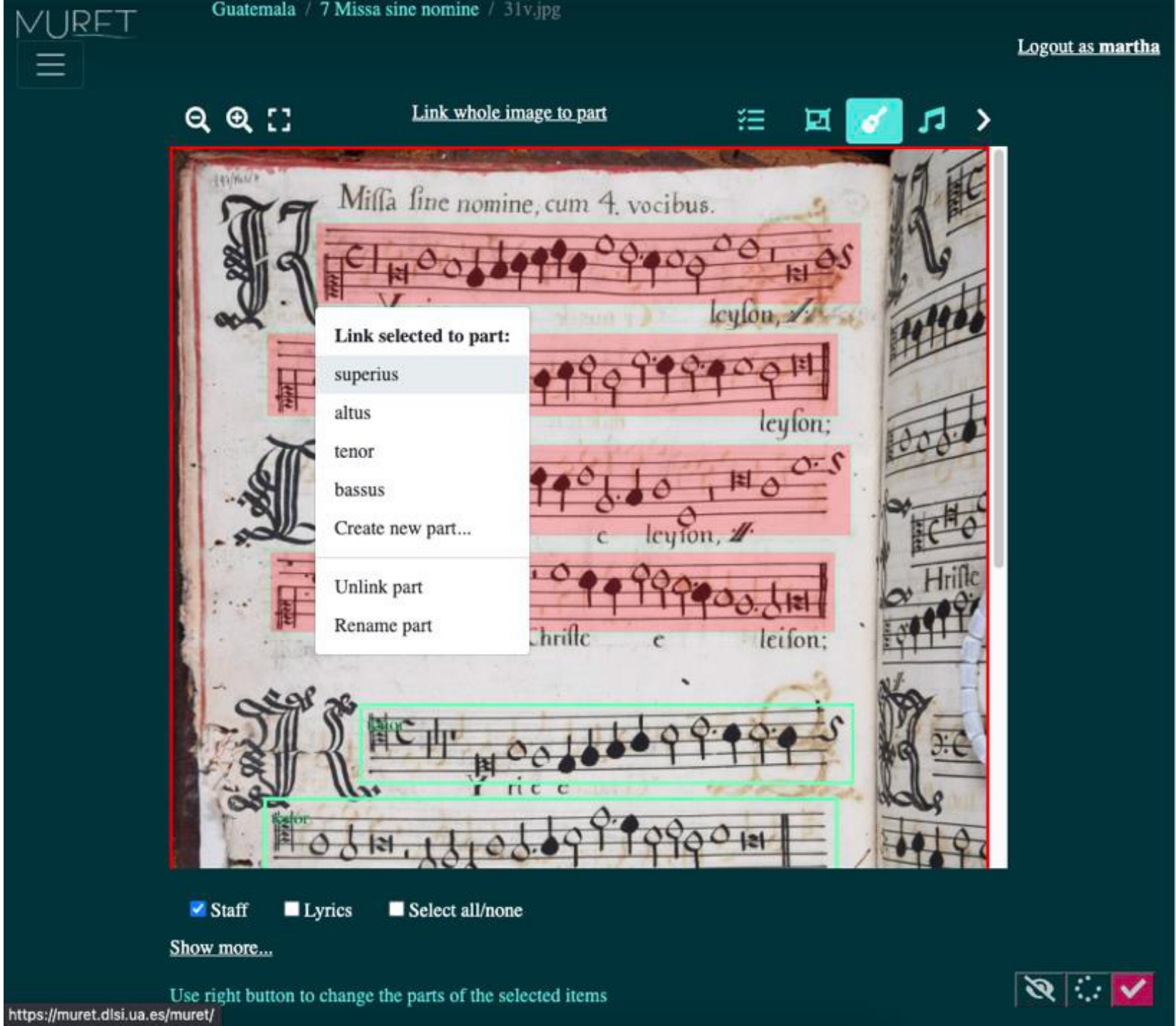


MuRET
OMR Stage 4: Music Encoding



MuRET

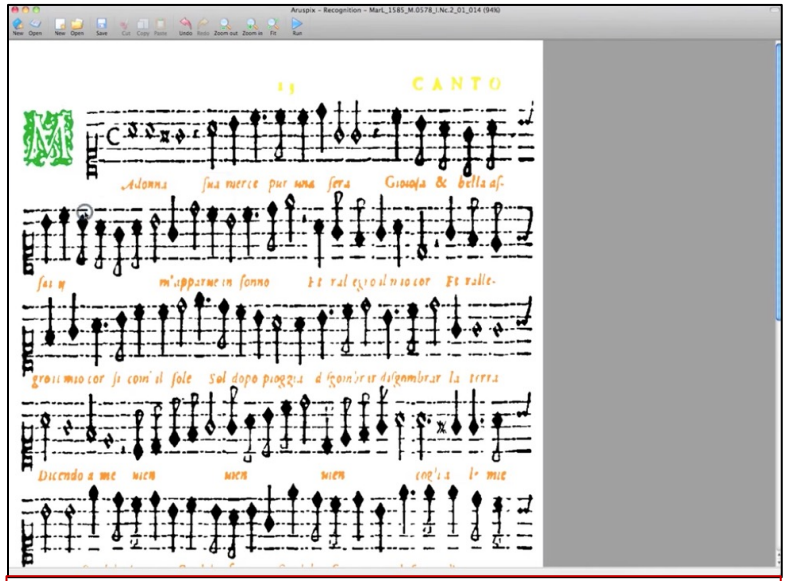
Extra Step:
Assignation
of Voices



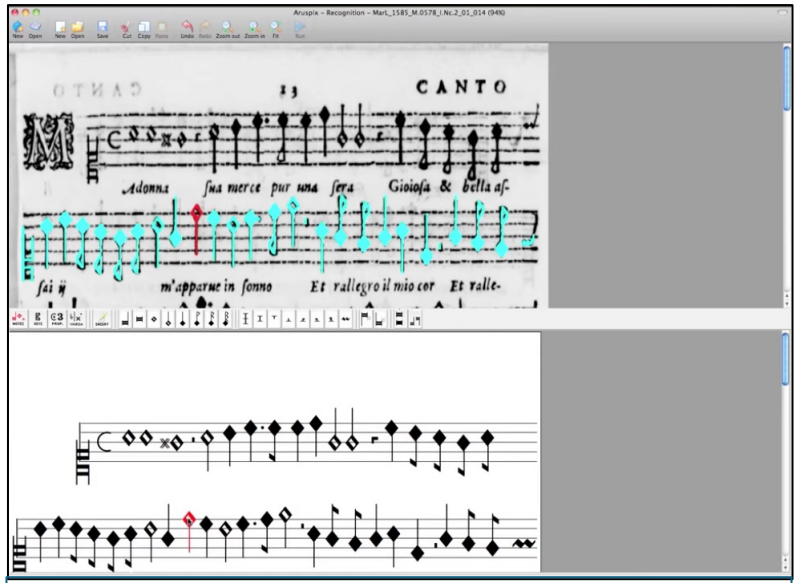
Aruspix

Developed by Laurent Pugin
http://www.aruspix.net

- Early typographic music prints
- Music printed during the sixteenth and seventeenth centuries with **movable typefaces**

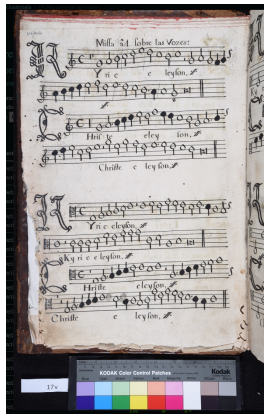


Preprocessing = Document Analysis



Symbol Recognition

Optical Music Recognition (OMR)



Optical Music Recognition

```
<mei meiversion="5.0">
   <meiHead>
      <fileDesc>
         <titleStmt>
            <title>06 Missa sobre las voces</title>
            <composer>Cristobal de Morales</composer>
         </titleStmt>
         <pubStmt/>
      </fileDesc>
   </meiHead>
   <music>
      <body>
         <mdiv>
            <score>
               <scoreDef>
                  <staffGrp>
                     <staffDef n="1" lines="5" label="superius" notationtype="mensural">
                        <clef line="2" shape="G"/>
                        <mensur sign="C" tempus="2" prolatio="2"/>
                     </staffDef>
                  </staffGrp>
               </scoreDef>
               <section>
                  <staff n="1">
                     <layer n="1">
                        <rest dur="brevis" loc="4"/>
                        <rest dur="semibrevis" loc="4"/>
                        <note dur="semibrevis" pname="g" oct="4"/>
                        <note dur="minima" pname="a" oct="4"/>
                        <note dur="minima" pname="b" oct="4"/>
                        <note dur="minima" pname="c" oct="5"/>
                        <note dur="minima" pname="d" oct="5"/>
```

Now… What can we do with the retrieved music that is encoded in these machine-readable files?

Now… What can we do with the retrieved music that is encoded in these machine-readable files?

Content-based searches
& computational music analysis