Symbolic Musical Data
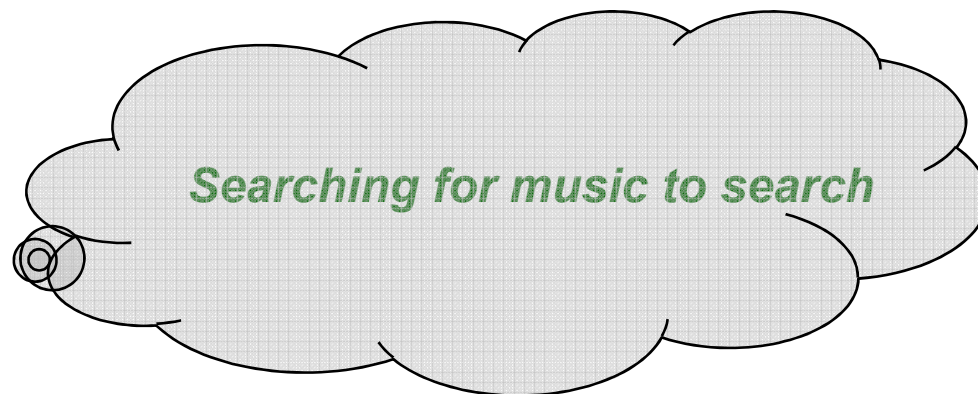
ISMIR Tutorial 2008

Eleanor Selfridge-Field and Craig Stuart Sapp

Overview

▸ 1. Issues

▸ 2. Repositories*

▸ 3. Allied Topics

▸ 4. Summary

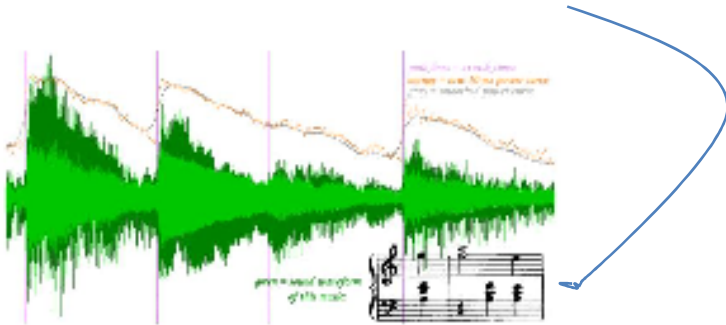▸ 5. Uses

Searching for music to search

1. Issues

Advantages and Challenges

Major advantages of symbolic data

▸ (1) Data is discreet (events isolated)

▸ (2) Data is humanly readable

▸ (3) Data representation is explicit

▸ (4) Data is parsable

Major advantages of symbolic data

▸ (1) Data is **discreet** (events isolated)

▸ Data is humanly readable

▸ Data representation is explicit

▸ Data is parsable

2008 Eleanor Selfridge-Field

Major advantages of symbolic data

▸ Data is discreet (events isolated)

▸ (2) Data is **humanly readable**

Stephen Foster: "Old Folks at Home"

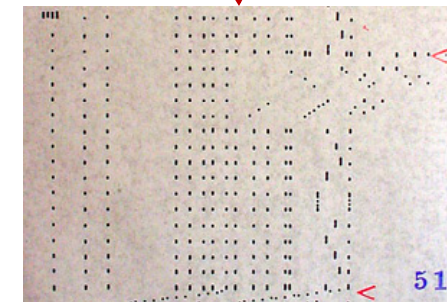3__2132|1_8_68_.|5__3_1_|2____|

▸ Data representation is explicit

▸ Data is parsable

Major advantages of symbolic data

▸ Data is discreet (events isolated)

▸ Data is humanly readable

▸ (3) Data representation is **explicit**

▸ Data is parsable

Major advantages of symbolic data

▸ Data is discreet (events isolated)

▸ Data is humanly readable

▸ Data representation is explicit

▸ (4) Data is **parsable**

Tonic sol-fa "notation"

Major issues in symbolic data

▶ (1) How complete is the data representation?

▶ (2) How accurate is the data?

▶ (3) Will the data format be maintained over the long-term?

▶ (4) How much data is available?

▶ (5) What are the terms of use?

▶ (6) Is the format widely used?

▶ (7) Is the format openly (and well) documented?

2008 Eleanor Selfridge-Field

Major issues in symbolic data

▸ (1) How **complete** is the data representation?
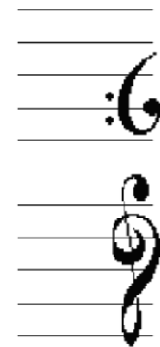
Stephen Foster: "Old Folks at Home"

3___2132|1_8_68_.|5___3_1_|2____|

E   DCED| C C AC |G   E C |D |

▸ How accurate is the data?

▸ Will the data format be maintained?

▸ How much data is available?

▸ What are the terms of use?

▸ Is the format widely used?

▸ Is the format openly (and well) documented?

Major issues in symbolic data

▸ How complete is the data representation?

▸ (2) How **accurate** is the data?

- ▸ Is the underlying source identified?
- ▸ Is metadata adequate?
- ▸ Is the file dated?
- ▸ Are data format and OS clearly identified?
- ▸ Has symbol data been verified by eye? By ear?

▸ Will the data format be maintained over the long-term?

▸ How much data is available?

▸ What are the terms of use?

▸ Is the format widely used?

▸ Is the format openly (and well) documented?

2008 Eleanor Selfridge-Field

Major issues in symbolic data

▸ How complete is the data representation?

▸ How accurate is the data?

▸ (3) Will the data format **be maintained** over the long-term?

    ▸ Archival formats

        ▸ Openly documented

        ▸ Altered only incrementally and infrequently

    ▸ Commercial formats

        ▸ Not openly documented

        ▸ Altered (or re-wrapped) frequently

▸ How much data is available?

▸ What are the terms of use?

▸ Is the format widely used?

▸ Is the format openly (and well) documented?

Major issues in symbolic data

▸ How complete is the data representation?

▸ How accurate is the data?

▸ Will the data format be maintained over the long-term?

▸ (4) **How much data** is available?

  ▸ Verbose vs compact codes (KB, MB, TB, no. of files, etc.)

  ▸ Numbers of titles, of movements, of events, of notes

▸ What are the terms of use?

▸ Is the format widely used?

▸ Is the format openly (and well) documented?

Major issues in symbolic data

▸ How complete is the data representation?

▸ How accurate is the data?

▸ Will the data format be maintained over the long-term?

▸ How much data is available?

▸ (5) What are the **terms of use**?

  ▸ Can end-user modify the data?

  ▸ Can end-user redistribute the data?

  ▸ Can end-user share the data with collaborators?

▸ Is the format widely used?

▸ Is the format openly (and well) documented?

2008 Eleanor Selfridge-Field

Copyright and other legal protections

Print copyright

- U.S.
  - Uniform (sort of)
- Europe
  - Varies by country

Recording copyright

"Big Content"

- U.S.
  - Varies by state
- Europe
  - Varies by country

Popular vs. classical

Copyright and other legal protections

Print copyright

▶ U.S.
  ▶ Uniform (sort of)
▶ Europe
  ▶ Varies by country

Other protections:
Derivative works issues

Recording copyright

▶ U.S.
  ▶ Varies by state
▶ Europe
  ▶ Varies by country

Other protections:
UK: graphical image copyright
Europe: cultural property protections

Major issues in symbolic data

▸ How complete is the data representation?

▸ How accurate is the data?

▸ Will the data format be maintained over the long-term?

▸ How much data is available?

▸ What are the terms of use?

▸ (6) Is the format **widely used**?

   ▸ Double-edged question

▸ Is the format openly (and well) documented?

Major issues in symbolic data

▸ How complete is the data representation?

▸ How accurate is the data?

▸ Will the data format be maintained over the long-term?

▸ How much data is available?

▸ What are the terms of use?

▸ Is the format widely used?

▸ (7) Is the format openly (and well) **documented**?

Generally not an issue with non-commercial data

Summary of Issues

▸ How complete is the data representation?

▸ How accurate is the data?

▸ Will the data format be maintained?

▸ How much data is available?

▸ What are the terms of use?

▸ Is the format widely used?

▸ Is the format openly (and well) documented?

2. Repositories

Purposes and Means

Symbolic Data =

- Data for notation
- Data for musical analysis
- Data used in layered applications
  - Pedagogy
  - Analytical markup
  - Music perception
  - Music cognition
  - Electro-acoustic music

Applications here representative, not exhaustive

2008 Eleanor Selfridge-Field   9/14/2008
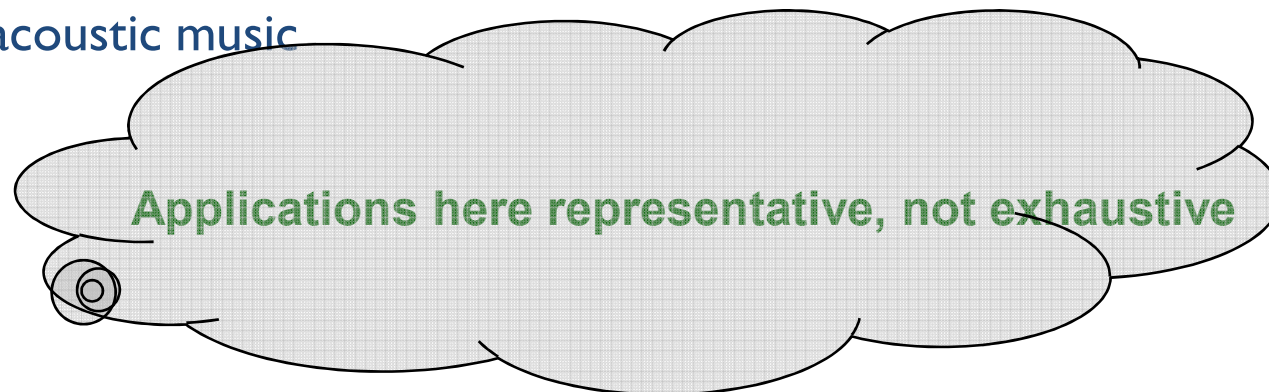
Preface: Common Music Notation (CMN) (or music from Common Practice Period)

Assumptions

- Large body of classical and popular music

- No market for special notation

- Can't accommodate
  - non-Western music
  - Early music
  - Recent music

…and (but)….

- _____confined to limited symbolic sets

- _____only affects a few small repertories

- _____not important because there is **so much** in the core repertories

Core of CMN elusive
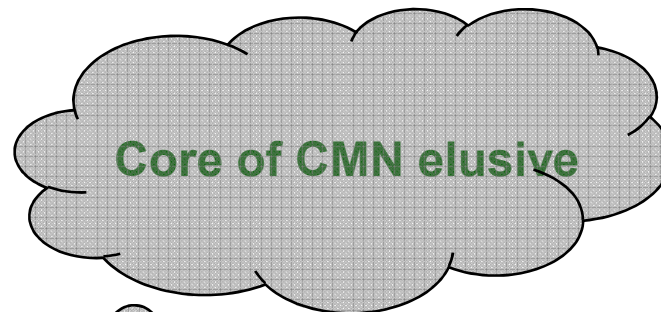
Writing systems outside CMN

Writing systems inside CMN?

17th c.   18th c.   19th c.   20th c.

2008 Eleanor Selfridge-Field
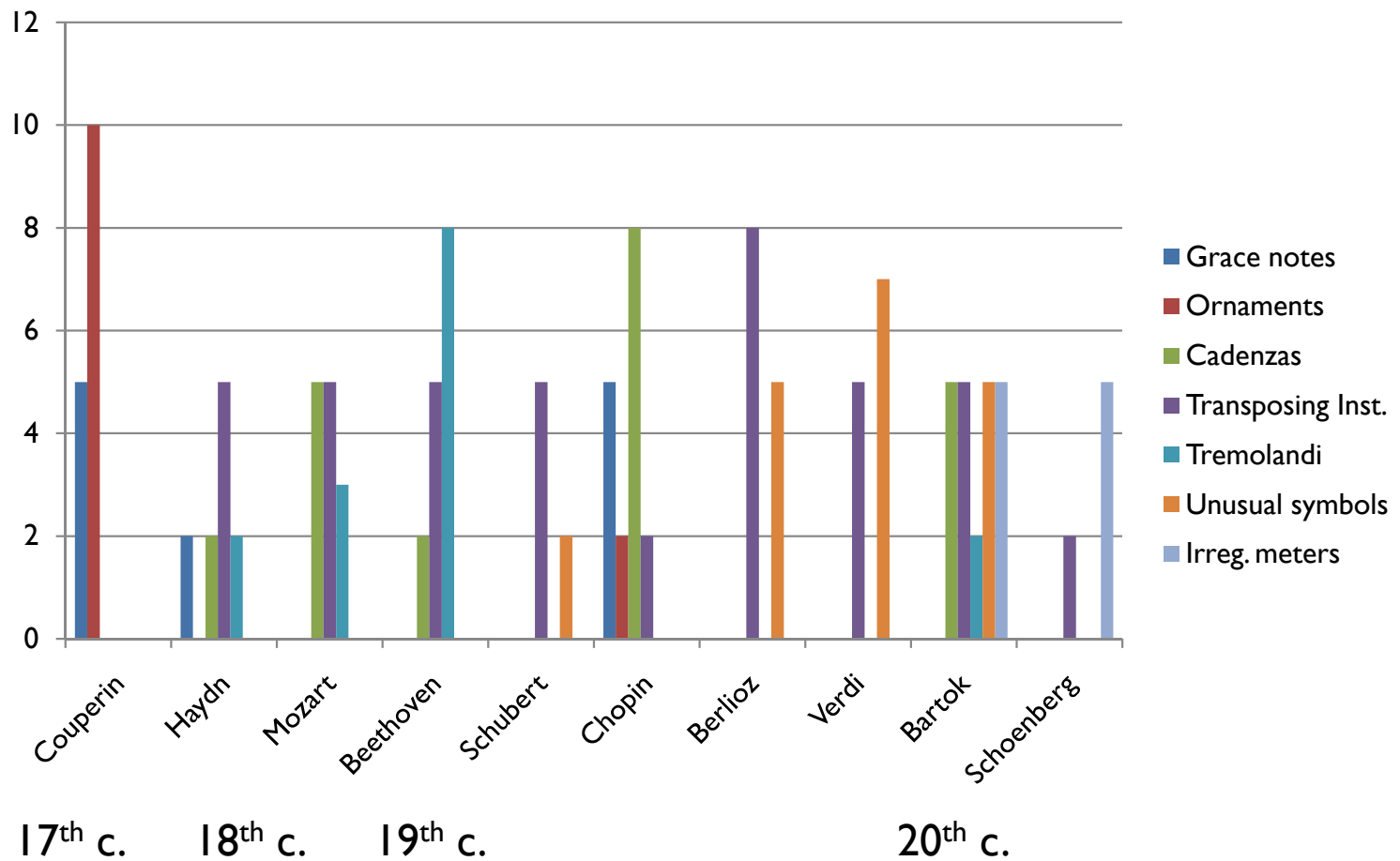
Inside Common Western Notation?

2008 Eleanor Selfridge-Field 9/14/2008
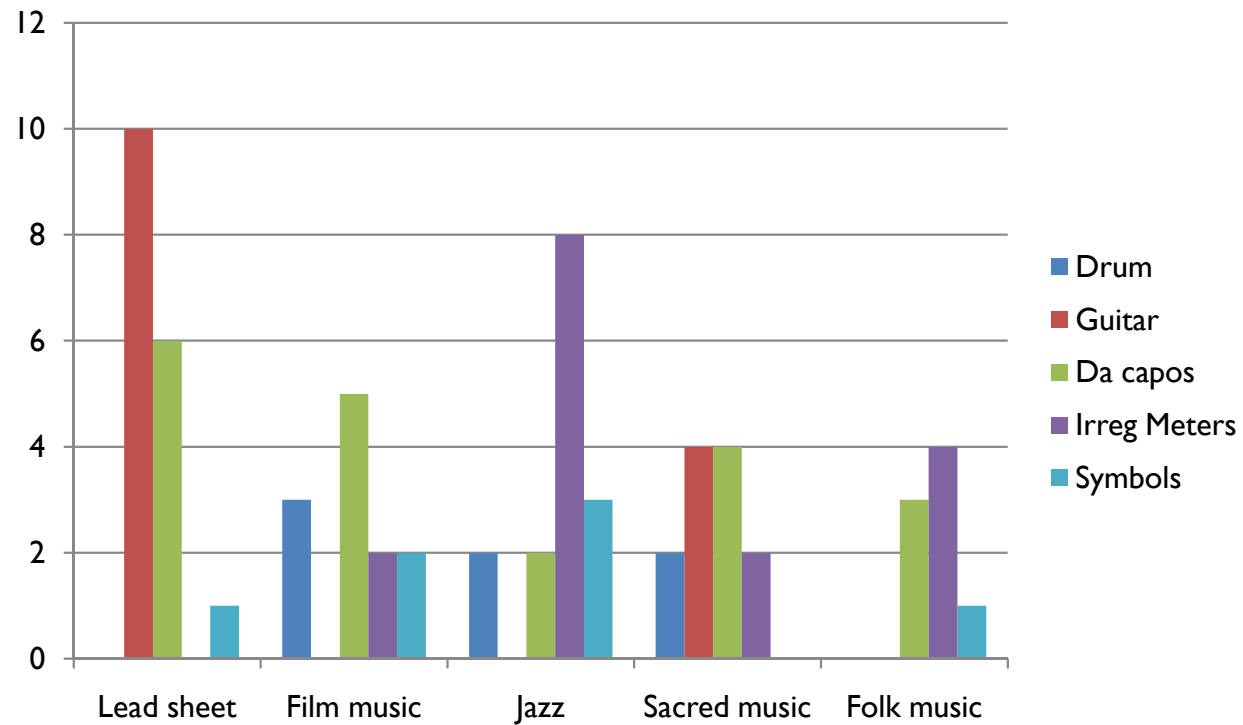
Repository Types

- (1) Image archives

- (2) Edition-based archives

- (3) Musical-data archives

- (4) Music-bibliographical archives

- (5) Collection-based archives

- (6) Cultural-heritage archives

- (7) Collaborative projects

Repository Types

- (1) **Image archives**
  - Images only
  - Enhanced image
  - Images with extensive metadata
  - Images with stored symbolic data
- Edition-based archives
- Musical data archives
- Music-bibliography archives
- Collection-based archives
- Cultural-heritage archives
- Collaborative projects

Images only

(a) U. of Chicago Library, Chopin Op. 61
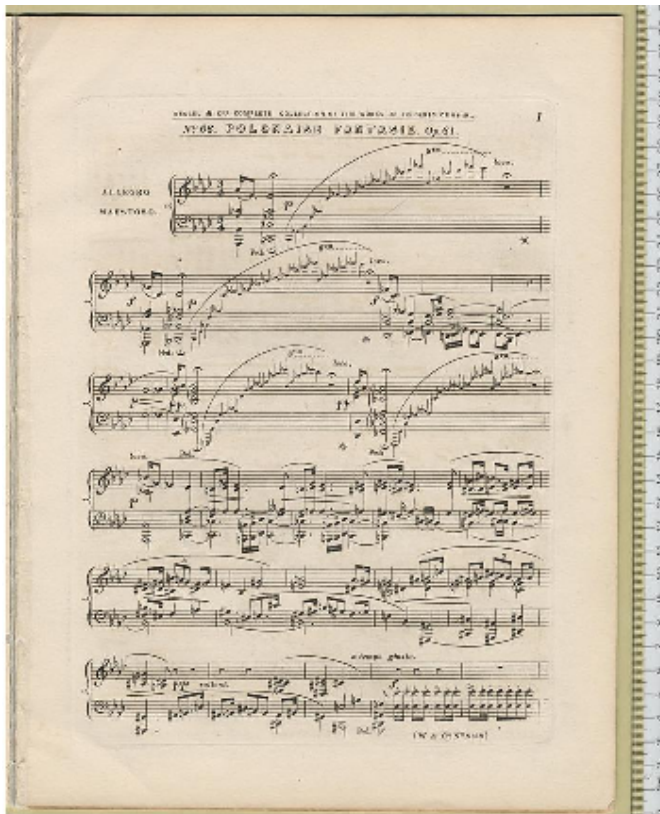
(b) Beethoven-Haus, Bonn Op. 132, sketch

2008 Eleanor Selfridge-Field
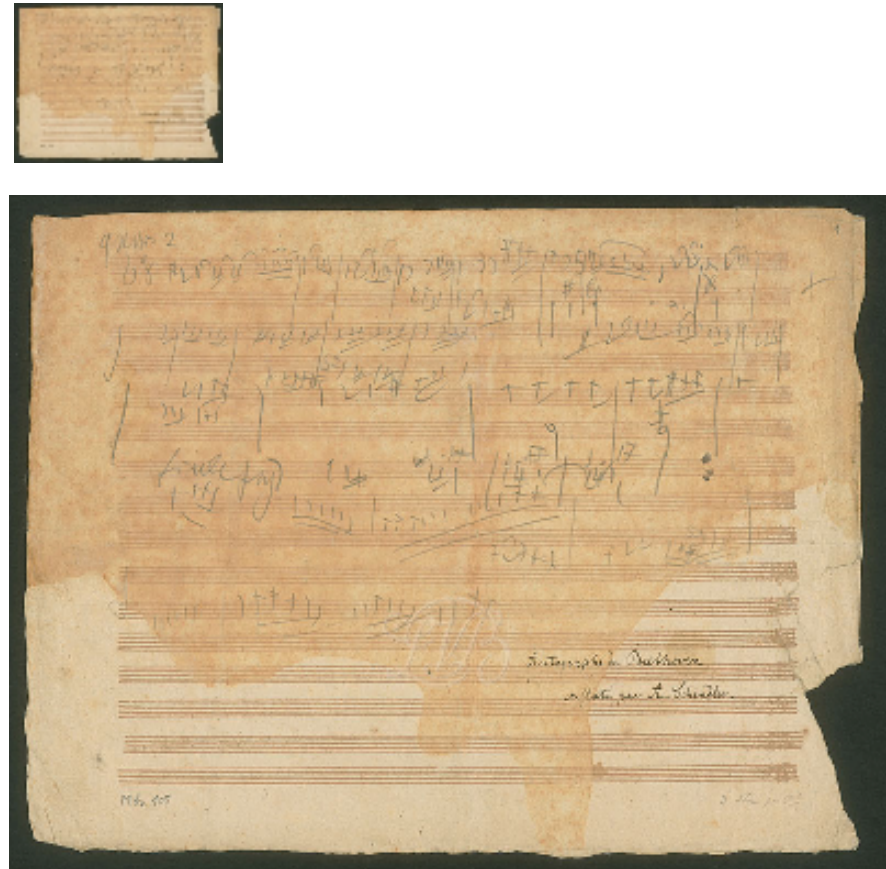
Turin National Library

Vivaldi Cantata (Lo-res)   **Vivaldi cantata,(Hi-res)**
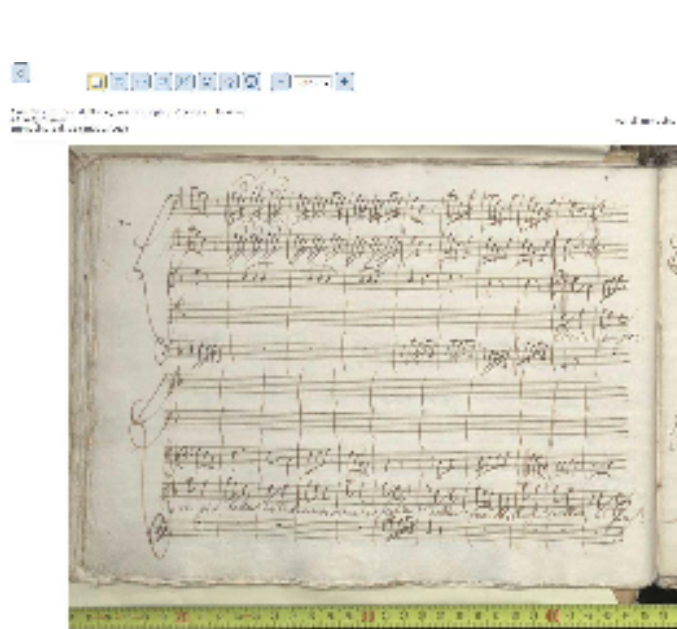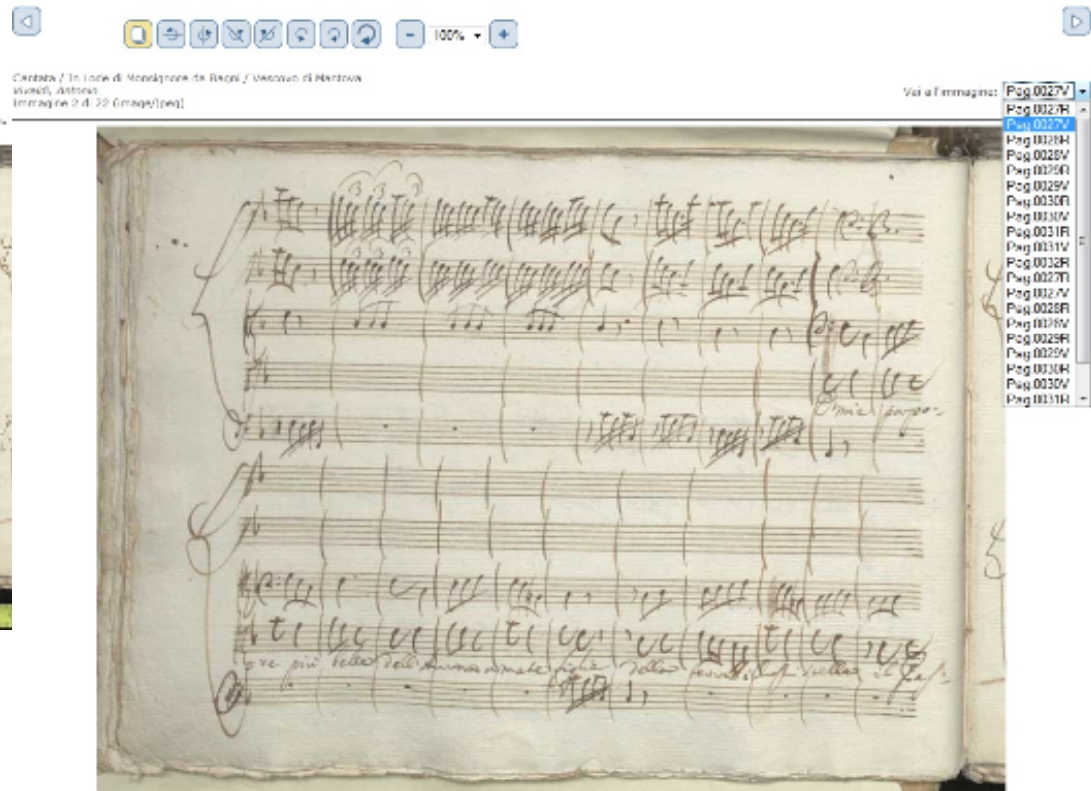
University of North Texas: Virtual Rare-Book Music Room

- About 100 scan collections for long works, mainly operas, 17-19th centuries
- High-quality images suitable for research

British Library, Music Division

Purcell |

←Handel

W. F. Bach →

Enhanced images

Digital Image Archive of Medieval Music (DIAMM)

DIAMM

Magnification*

2008 Eleanor Selfridge-Field 9/14/2008
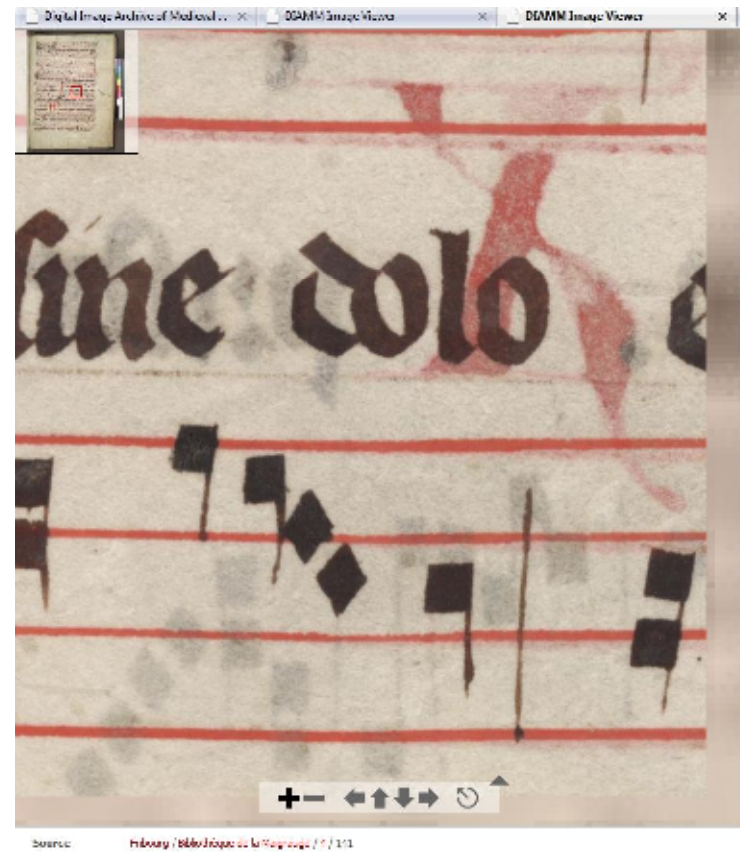
Images with extensive metadata

Variations 3 (Indiana U.):

Brahms's Academic Fest. Ov.

Recording server

Score server   **Metadata***

Tools for music theory

"container" concept

FRBR

Images with (some) stored symbolic data

Levy (Civil War) Sheet Music Coll., Johns Hopkins U.*

Bavarian State Library, Munich (Vifamusik Project)*

Guido Music Notation

Several parallel formats

Repository Types

- Image archives

- **Edition-based archives**

    - Replication of printed edition with critical notes

    - Electronic edition

    - Electronic apparatus for created new printed/virtual edition

    - Viewable edition with sound playback

- Musical data archives

- Music-bibliography archives

- Collection-based archives

- Cultural-heritage archives

- Collaborative projects

2008 Eleanor Selfridge-Field

Digital replication of printed editions

Neue Mozart Ausgabe

Salzburg Mozarteum

Electronic apparatus for new editions

EDIROM Project, U. Detmold: CM v. Weber
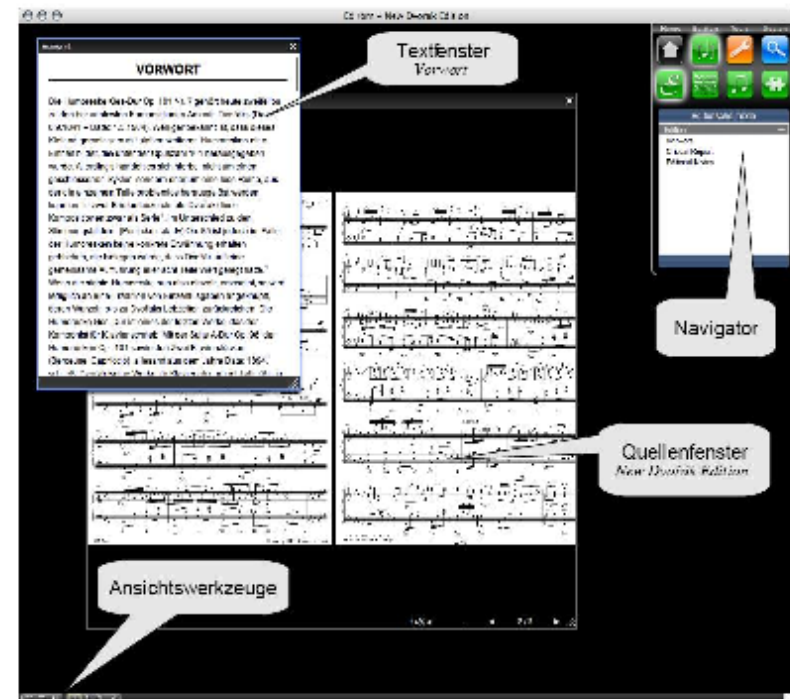
EDIROM Project, Prague: Dvorak Edition

"Born digital" editions

Web Library of Seventeenth-Century Music (Duke U.)

Werner Icking Music Library

Introduction, critical notes, editorial indications

Marini: Sonata variata from Op. 8 (1629)

Repository Types

▸ Image archives

▸ Edition-based archives

▸ **Musical data archives**

    ▸ For teaching

    ▸ For performance

    ▸ For research

▸ Music-bibliography archives

▸ Collection-based archives

▸ Cultural-heritage archives

▸ Collaborative projects

Viewable edition with sound playback

Variations 3 (teaching)

Also indirectly …

▸ Choral Public Domain Library

▸ Werner Icking Music Library

Data Archives for Performers (Collaborative Projects)

Choral Public Domain Library (CPDL)

- Formats
  - Finale
  - Sibelius
  - MIDI
  - PDF (Finale, Sibelius), *.ps
  - MUP
- Features
  - Wiki format
  - Text translations (> 3,000)
  - >9700 pp.
  - Score and composer catalogs

Werner Icking Music Archive (WIMA)

- Formats
  - MusiXTeX (*.ps)
  - Finale
  - PDF (Finale, capella)
  - mp3
- Features:
  - Almost 500 composers
  - Apparatus for revisions
  - Identification of sw versions
  - Performing parts

Data archives for performers

Werner Icking Archive (WIMA)

Choral Public Domain Library (CPDL)

Data archives for research

MuseData

▸ c. 1,200 encoded full scores

▸ Reportories:

  ▸ 17-19th centuries

  ▸ Mainly orchestra, chamber, opera, oratorio

▸ File types

  ▸ Raw MuseData (edited and verified)

  ▸ Kern

  ▸ MIDI (two versions)

  ▸ SCORE

  ▸ PDF

KernScores

▸ c. 1,000 encoded full scores

▸ Reportories:

  ▸ 15th-early 20th centuries

  ▸ Mainly keyboard

▸ File types

  ▸ Kern

  ▸ MIDI

  ▸ PDF

  ▸ On-the-fly

    ▸ Guido Music Notation

    ▸ Piano-roll files

    ▸ Keyscapes (graphical harmonic analysis)

Data archives for research: MuseData

- Main strength
  - Multi-format consistency:
    - data content matches **across formats**
    - data content matches **between score and parts**
- Repertories with PDFs
  - Bach, J.S.: WTC I, II; Chorales
  - Beethoven: 6 symphonies
  - Haydn: 12 symphonies
  - Vivaldi: 12 concertos
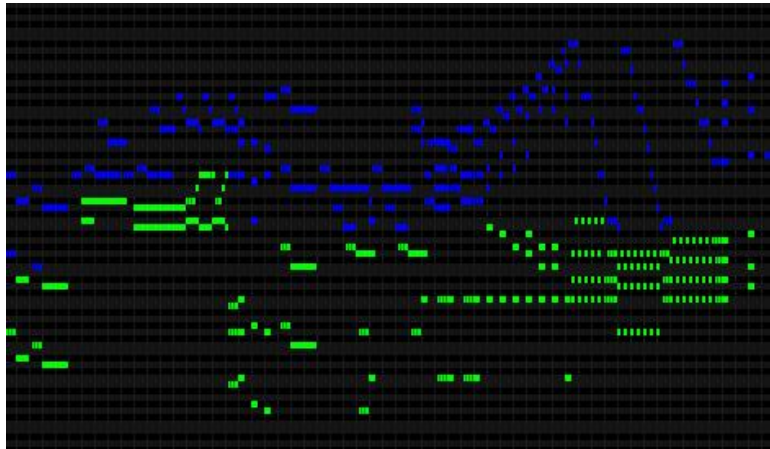
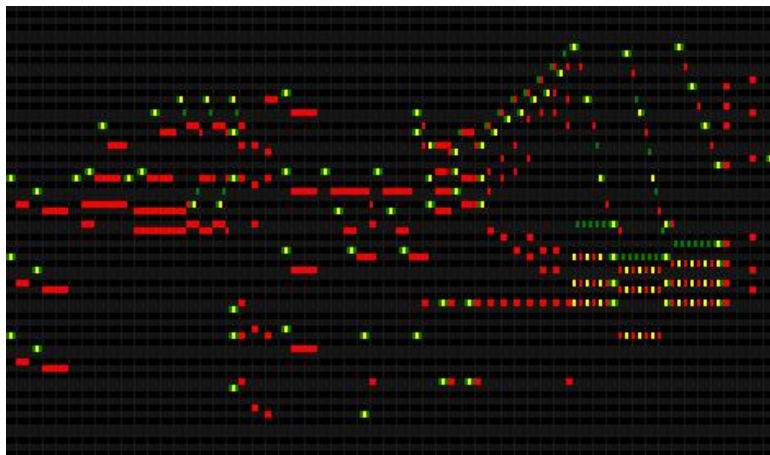When consistency matters

Data archives for research: KernScores*

Beethoven's Fifth Symphony (beginning)

Kern: piano rolls

Kern: visual harmonic analyses

Green = left hand

Red = downbeat

Seeing

Hearing

(a)   By pitch and duration
(b)   By metrical accent

(a) Without repeats; (b) with repeats

Repository Types

▸ Image archives

▸ Edition-based archives

▸ Musical data archives

▸ **Music-bibliography archives**

  ▸ For metadata

  ▸ For cross-reference

▸ Collection-based archives

▸ Cultural-heritage archives

▸ Collaborative projects

Music-bibliographical repositories

RISM (Rép. Int. des Sources Musicales)

- Audiences
  - librarians
  - researchers
- Purposes
  - Manuscript inventory
  - Musical concordances
- Contents
  - Metadata
  - Musical code

Themefinder (CCARH)

- Audiences
  - Search researchers
  - Music librarians
  - A few analysts, psychologists
- Purposess
  - Study of search methods
  - Study of query constructions

2008 Eleanor Selfridge-Field

RISM Interface Dialects

RISM-UK

RISM-CH

2008 Eleanor Selfridge-Field 9/14/2008

More RISM, Themefinder

RISM-Italy

RISM-US

2008 Eleanor Selfridge-Field   9/14/2008

Repository Types

▸ Image archives

▸ Edition-based archives

▸ Musical data archives

▸ Music-bibliography archives

▸ **Collection-based archives**

▸ Cultural-heritage archives

▸ Collaborative projects

Bayerische Staatsbibliothek (Vifamusik)*

Edition Reprints: G. F. Händel (Chrysander)

Early prints and MSS

Ricordi Archive (Milan)

Respighi: Pines of Rome
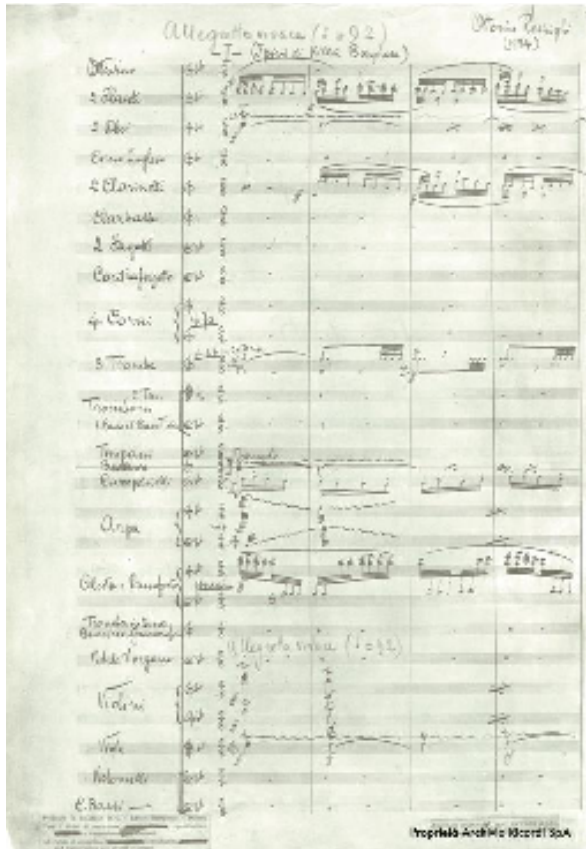
Paganini Capriccio

Signed musical score, Paganini, Capricci, 1816.
Click to view full-size image
(2 / 5)

Repository Types

▸ Image archives

▸ Edition-based archives

▸ Musical data archives

▸ Music-bibliography archives

▸ Collection-based archives

▸ **Cultural-heritage archives**

▸ Collaborative projects

Cultural Heritage
Library of Congress—American Memory

Aaron Copland, 1908-1918

Aaron Copland, 1965

Repository Types

▸ Image archives

▸ Edition-based archives

▸ Musical data archives

▸ Music-bibliography archives

▸ Collection-based archives

▸ Cultural-heritage archives

▸ **Collaborative projects**

Collaborative Projects: MIDI-based

(Die) Kunst der Fuge

- 16,500 classical MIDI files
- PDF sheet music (<8000 scanned pp from 171 books)
- Mainly piano, Romantic era, Russian, Bach, anything out of copyright
- Some quality-of-performance IDs

Classical MIDI Archive

- >40,000 files by >2,000 composers
- Multiple performances of some items

Collaborative, non-MIDI projects

Aruspix scanning*

- Scan (Aruspix) to *Wolfgang*
  - *Notation software optimized for 16th-17th centuries*
  - *Projects*
    - *Geneva University*
    - *McGill University*
    - *Marenzio edition*

IMSLP

- PDF (bit-mapped) archive
- User-contributed
- Claim 20,000 PDF scores (many restrictions due to copyright*)
- c. 1,200 composers
- Repertory: mainly 19th-century orchestral, Russian and Polish chamber music and songs
- Operates under Canadian copyright provisions
- Linked (off-and-on) to Project Gutenberg

Collaborative, non-MIDI

- Notation-software sharing sites (aided by viewers)
    - Finale
    - Sibelius
- Free-agent sites
    - Gordon Callon's web sites (Acadia U.)
        - Baroque
        - Teaching pieces
        - Canadian traditional music
    - Alain Naigeon (France)
        - Mainly Renaissance

Mutopia

- OS: Linux

- Typesetting software: Lilypond

- Other formats: MIDI, PDF

- Features

  - Lulu link

  - Shows incipits at website

BWV 1006a - Gigue

by J. S. Bach (1685-1750)

3. Allied Topics

Data interchange and linking

Allied topics

▸ Manually encoded repertories

▸ Interchange codes and schemes

▸ PDF viewer/players

▸ Linking symbolic data with audio

Manually encoded repertories

- Principally monophonic

- Component data

  - Standard metadata

  - Lyrics

  - Musical transcription

- Some examples

  - Austrian Volkslied Archiv

  - Deutsche Volkslied Archiv

- Best-known example

  - Essen Databases (EsAC format)

- Extent: undetermined

Übersicht über die thematische Gruppierung der Dokumentationsmappen

Gr I	Ballade, erzählendes Lied
Gr II	Historisch-politisches Lied
Gr III	Liebeslied
Gr IV	Abschieds- und Wanderlied
Gr V	Tagelied, Kiltlied
Gr VI	Ehestandslied
Gr VII a	Tanzlied
Gr VII b	Vierzeiler, Schnaderhüpfl
Gr VII c	Anhängsel und Schnörkel
Gr VIII	Rätsel-, Wunsch- u. Wettlied
Gr IX	Trink- und Tabaklied
Gr X	Ansingelied zu Kalenderfesten
Gr X a	Ansingelied ohne festes Datum
Gr X b	Ansingelied (Lebenslauf)
Gr XI	Soldatenlied
Gr XI b	Jägerlied
Gr XI c	Hirten- und Bauernlied
Gr XI d	Ständelied allg. (Handwerk und andere Berufe)
Gr XI e	Studentenlied
Gr XI f	Arbeitslied und -reime
Gr XII	Scherz- und Spottlied
Gr XIII	Heimat-, Auswandererlied
Gr XIV	Schlager 20. Jahrhundert
Gr XV a	Weihnachtslied
Gr XV b	Geistliches Lied (allgemein)
Gr XV c	Marien-, Wallfahrts-, Legendenlied
Gr XV d	Buß-, Sterbe- und Totenlied
Gr XVI	Fremdsprachiges Lied
KiV	"Kunstlied im Volksmund"
DVldr	Balladen (DVA-Publikation)

Genre

EsAC Record Structure

Interchange codes

- XML schemes for music

  - MusicXML (Recordare)

  - MusicXML Library (Grame)

  - Music-Encoding Initiative (MEI)

  - IEEE 1599 (Laboratorio d'Informatica Musicale, Milan U.)

  - Corpus Mensurabilis Musicarum Electronicum (CMME)

- Electronic repertories: ChucK et al. (Stanford)

MusicXML*

- Provider
  - Recordare, LLC
- (Mainly) commercial code
  - Finale
  - Sibelius
  - PriMus
- Aims
  - Archival interchange format
- Versions
  - 1=OS specific
  - 2=Java-based

2008 Eleanor Selfridge-Field   9/14/2008

MusicXML Library (GRAME)

Sourceforge.net

- OS's supported:
  - Mac
  - Linux
  - Win
- Versions
  - 2
- Main developer
  - Dominique Fober

Available data
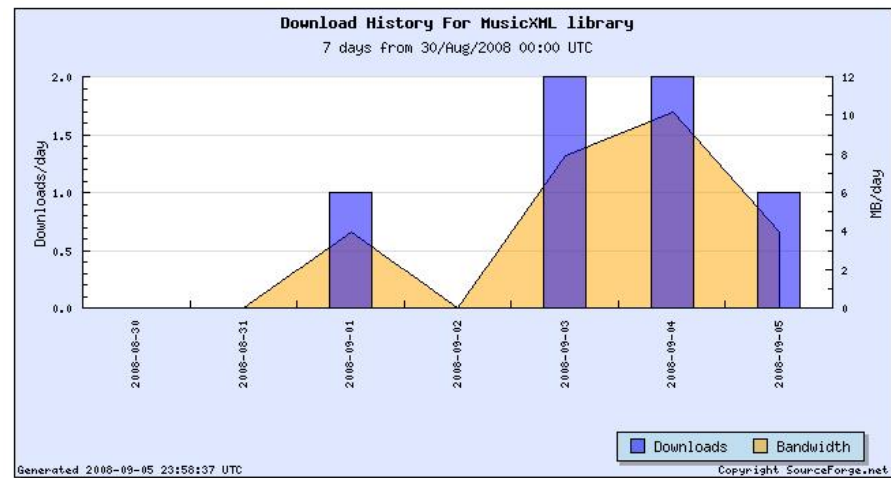
MusicXML import to notation software (PriMus)

MusicXML example imported into Primus
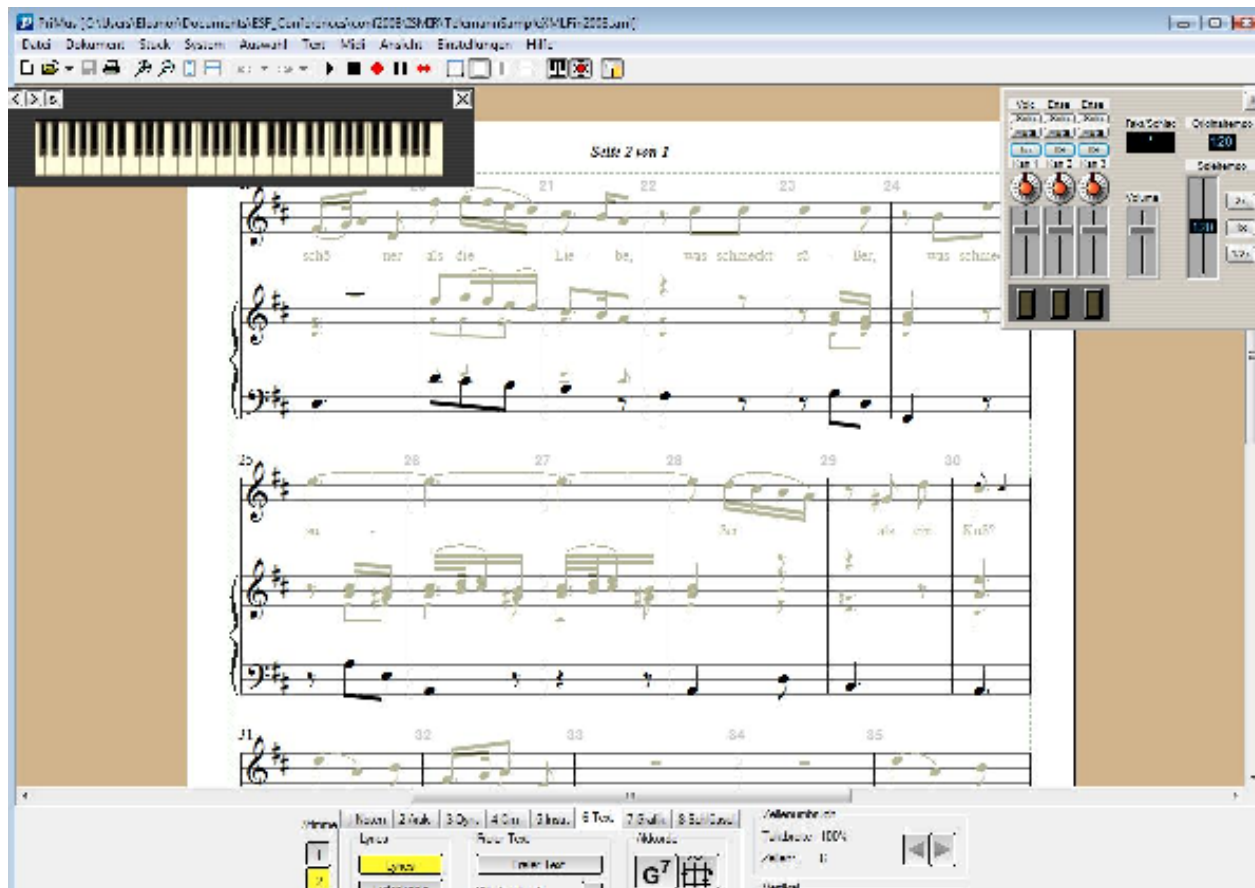
MusicXML import to notation software (PriMus)
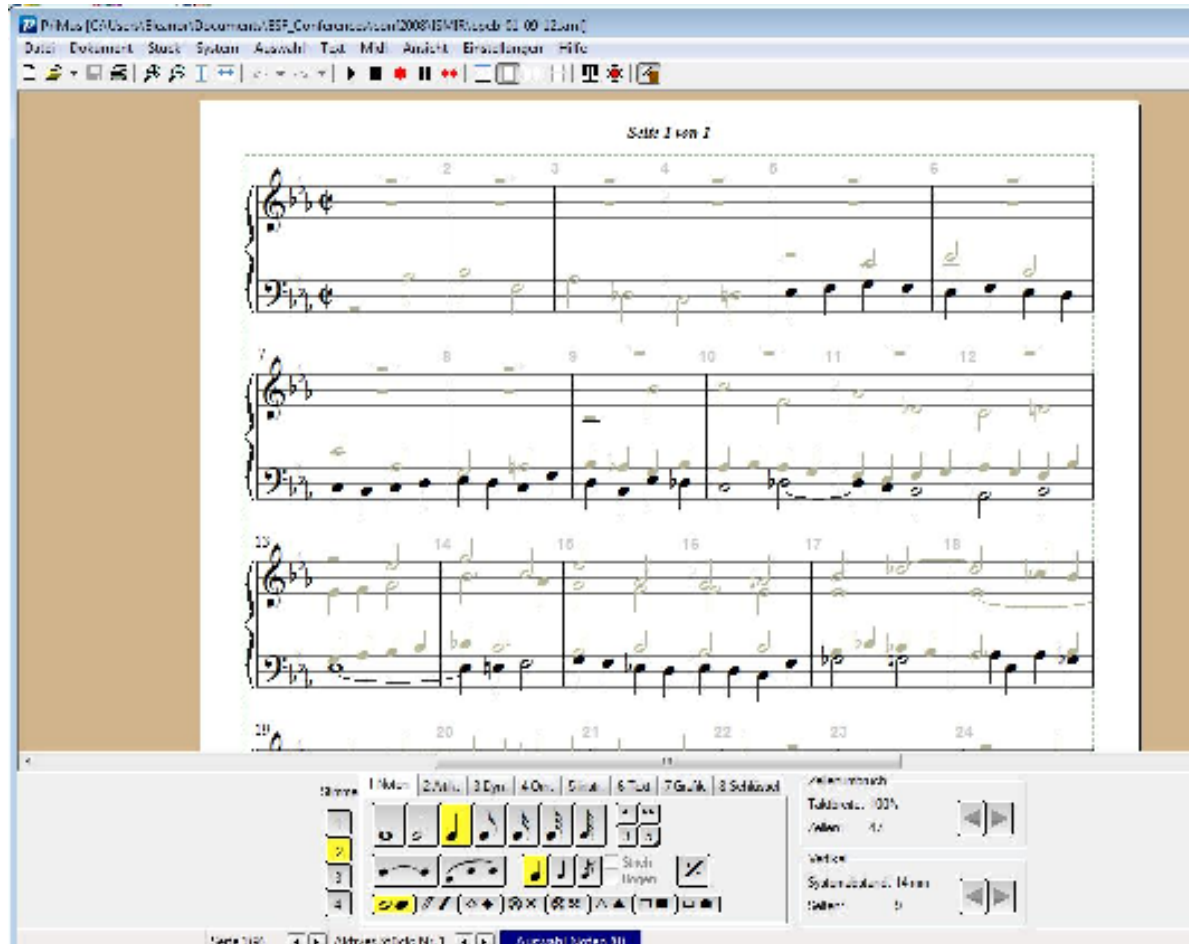
SCORE *.eps to print to scan …

MusicXML import to notation software (Primus)

KernScores automatically-generated MusicXML, imported into Primus
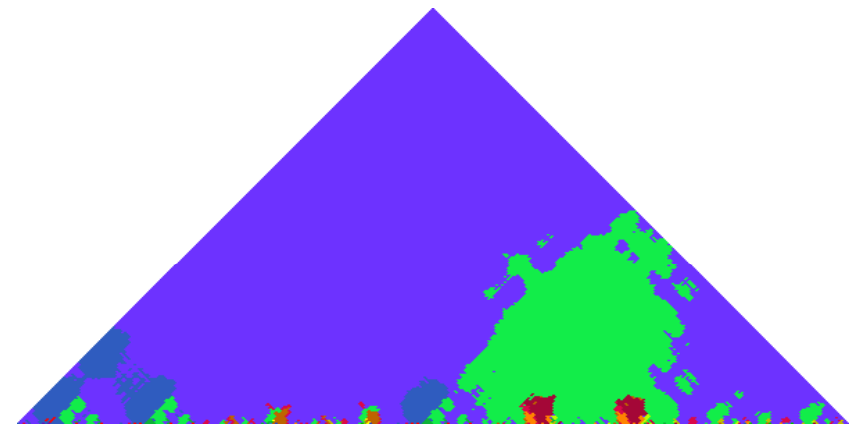
Keyscape showing harmonic structure of Joplin's Maple Leaf Rag

Music Encoding Initiative (MEI)

- Where: U. of Virginia
- Orientation
  - Bibliographical apps
    - Thematic indices
  - Scholarly apps
    - Critical apparatus supported
  - TEI analogues
    - General features in place

- I/O links with
  - MuseData
  - MusicXML
  - PS (via MUP)
  - PDF
  - TEI

IEEE 1599—U. of Milan

- Background: coord. With Ricordi Archive
- Aims:
  - synchronization of
    - Manuscripts
    - Prints
    - Pictures (still and moving)
    - Recordings
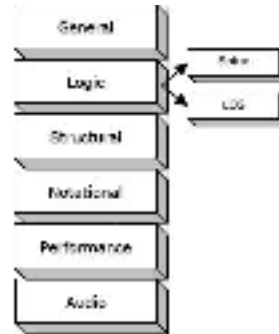- Status: IEEE listed project

Figure 1. MX multi-layer structure.

Figure 2. General structure of an MX file.
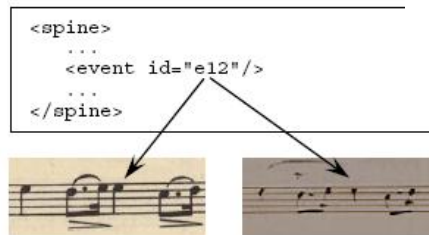
Figure 3. Event e12 graphical mappings.
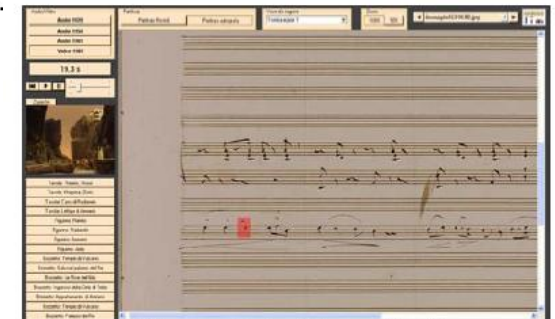
Figure 4. Screenshot of the application.

Corpus Mensurabilis Musicarum Electronicarum (CMME)

▶ Aims

  ▶ Archive for music in mensural notation (variant system[s] of meter and prolation)

  ▶ Critical apparatus

  ▶ Reference pages for sources and incipits

  ▶ Flexible graphical representations in modern notation

  ▶ Expandable apparatus for integration of multiple projects concerned with Renaissance music

CMME 2

Beate omnes

CMME Description

CMME 3

Salve radix

Structural description

Measures:	1-8	9-27	28-32	33-49	50-63
Section:	A	B	C	B'	A'

Salve radix details

PDF viewers/players (*PDF-to-MusicPro*)*

- What it does (viewing)
  - Captures vector graphics info from PDFs
  - What programs output this?
    - Finale
    - Sibelius
    - SCORE
  - Secondary production: *.ps files distilled by Acrobat; *.eps files
- What it does not do
  - Read bit-mapped images

- What it does (playing)
  - Outputs piano rendition of keyboard music if not too complicated
- What it does not do
  - Read orchestral music (tries to output individual tracks to separate playbacks)
  - Keep beat synchronized with score
  - Always start at beginning of file

PDF-to-Music-Pro Examples

MuseData PDF to MusicPro

2008 Eleanor Selfridge-Field

PDF-to-Music-Pro Examples

MuseData PDF to MusicPro

2008 Eleanor Selfridge-Field

PDF-to-Music-Pro Examples

LUCA MARENZIO, QUELL'AUGELLIN CHE CANTA,
(IL SETTIMO LIBRO DE MADRIGALI, 1595, no. 2)

Gordon Callon's website
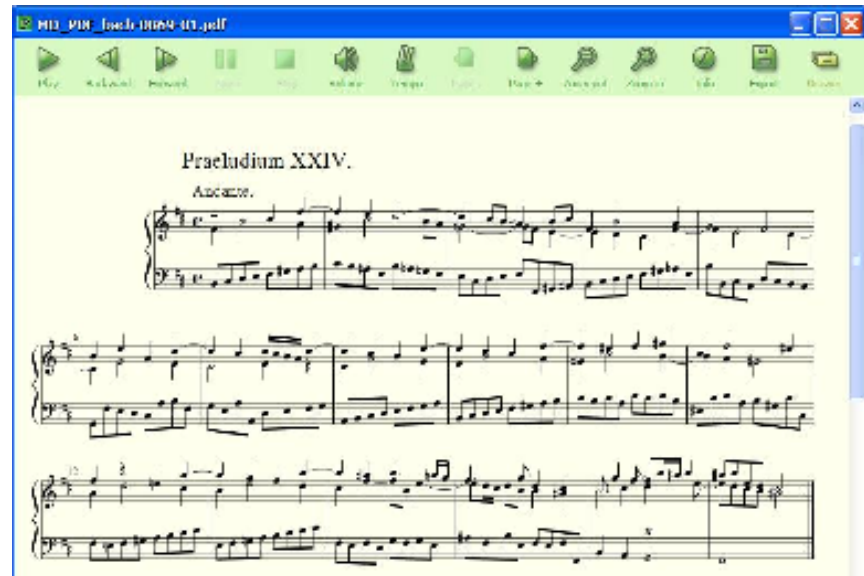
SCORE *.eps to Distiller to PDF to MusicPro

PDF-to-Music-Pro Examples

Icking Music Archive *.ps to Distiller to PDF to MusicPro

2008 Eleanor Selfridge-Field
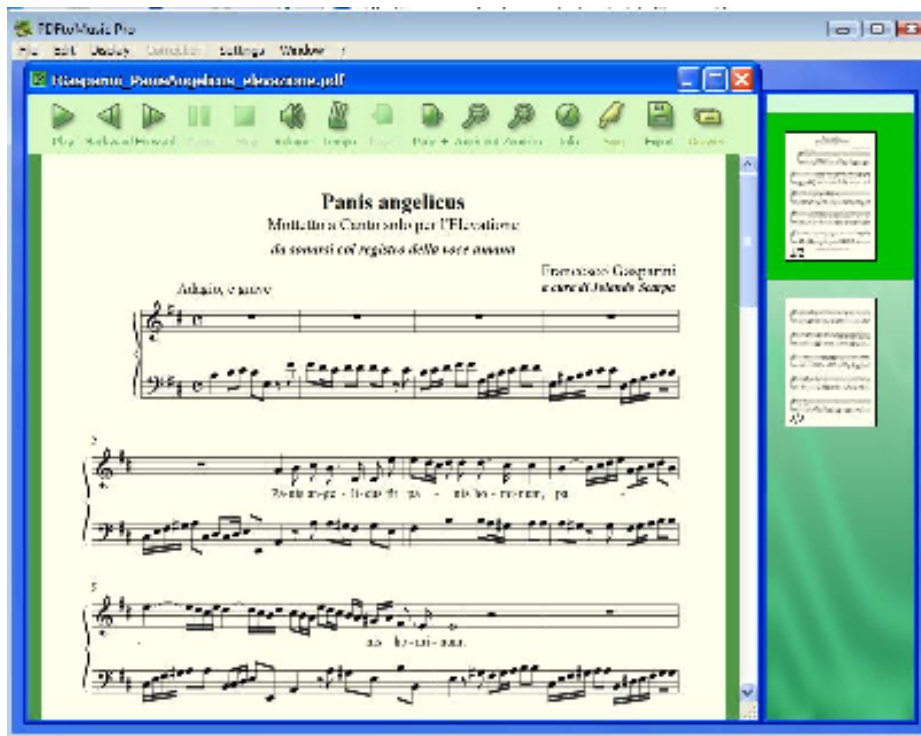
Symbolic-audio synchronization

MIDI-based

- Jeremy Pickens et al. (2003)*
  - Project: Use of MIDI as intermediate layer between audio and symbolic data (2003)
  - Motive: synchronization
- Meinard Müller et al. (2005—)*
  - Project: All of above plus video synchronization
  - Motivations: tools for use with RWC (Real World Computing database)

Kern-based

- Kyogu Lee (Yahoo), 2007*
  - Project: Use of kern as intermediate layer between MIDI "encodings" and audio (2007)
  - Motive: chordal analysis
- Craig Sapp, CHARM (2006—)*
  - Motive: alignment of temporal information in score, MIDI, and audio data

Image-symbol-audio based
VifaMusik (Munich SB)
IEEE 1599 (Milan)

Symbolic-audio synchronization: new music

Rob Hamilton

- Archival format for (rescue of) electro-acoustic music
  - Humdrum with spine for controller description

Dexter Morrill, John Chowning (1980s)

Open Music and ChucK*

- Matt Wright
  - Open Music (200—): open standard for electronic music
- Ge Wang
  - ChucK: realtime interactive synthesis among collaborating computers (time control fundamental)

Slork 2008

4. Summary

Review of Symbolic Data

(1) How complete is the data representation?
(2) How accurate is the data?
(3) Will the data format be maintained over the long-term?
(4) How much data is available?

(5) Does the code suit all the needs of the project?
(6) Is the format potentially archival?
(7) Are originating software, encoding versions, and other contextual details adequately documented?

Musical needs, stability, annotation

▶ Evaluation difficult because

  ▶ High scores in all categories unlikely

▶ Each goal important in a different way

  ▶ Musical needs fundamental

  ▶ Code stability suggests long(er) life for projects

  ▶ Annotation essential for survival

Projects and data with potential

▶ Musical needs

Cymbal (C)
Snare (S)
Bass (B)
Hi-Hat (H)
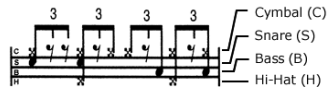
CMN??

▶ Code stability

- ▶ MusicXML
- ▶ MEI (?)
- ▶ SCORE
- ▶ Kern
- ▶ MuseData
- ▶ Lilypond (?)
- ▶ *.ps, *.eps, *.pdf (!!)
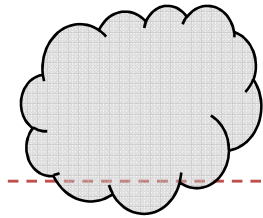
Projects and data with potential

- Code acceptance
  - Interchange
    - MusicXML
    - CMME
  - Native
    - Kern
    - MuseData
    - MIDI????
  - With version creep
    - Finale
    - Sibelius

- Contextual annotation
  - Werner Icking Music Archive
  - MuseData
  - KernScores
  - Variations
  - DIAMM
  - CMME
  - British Library
  - U Texas

2008 Eleanor Selfridge-Field 9/14/2008

Hidden code clusters

Editions

- Neue Mozart Ausgabe (SCORE and PHI code)
- C.P.E. Bach Edition (SCORE and PHI code)
- Most Schott collected editions (SCORE)
- Many Bärenreiter collected editions (SCORE et al.)
- Some Ricordi editions (e.g., Verdi) SCORE
- Some Henle editions (customized version of Finale)

Archives

- Bayerische Staatsbibliothek
- Werner Icking Archive
- MuseData Archive
- KernScore Archives

Converging projects

- CMME/DIAMM/Motet DB
- Edirom/new editions

Commercial publishing

- In-house data translation sw

5. Using Symbolic Data

Go to Craig Sapp's presentation