

Search Experiments

Mark Notess, Steve Harris and Julie Hardesty
Digital Library Program Brown Bag Series

23 March 2011



LIBRARIES

INDIANA UNIVERSITY

Bloomington



Outline

- V/FRBR project
- Why FRBR is interesting for music
- FRBRization Overview
- Scherzo demo
- Scherzo analysis
- Planned evaluation
- Other search experiments



INDIANA UNIVERSITY

Variations/FRBR Project @ IU

- Funded by IMLS, 10/2008-9/2011, Jenn Riley - PI
- Concrete testbed for FRBR, using music (scores/recordings) as an example
- Model for next-generation catalogs & cataloging
- Develop data model that embodies FRBR principles
- *Design and implement a new, openly-accessible search interface for discovery*



Why FRBR for Music?

- In music, especially classical, **the work is primary**.
- There are many **more instances** of a given work than for most monographs (e.g., Stardust has ~1,500 recordings).
- Item titles can matter far less than for monographs, (e.g., “Songs”).
- Music doesn’t have (just) an author. It has the composer, performers, conductor, arranger, librettist—maybe even a lithographer. Any *may* matter more than composer.
- Music also has arrangements, instrumentation, key(s), and a slew of interesting dates (composition, first performance, performance, publication)
- Albums and songbooks often have multiple works by different composers



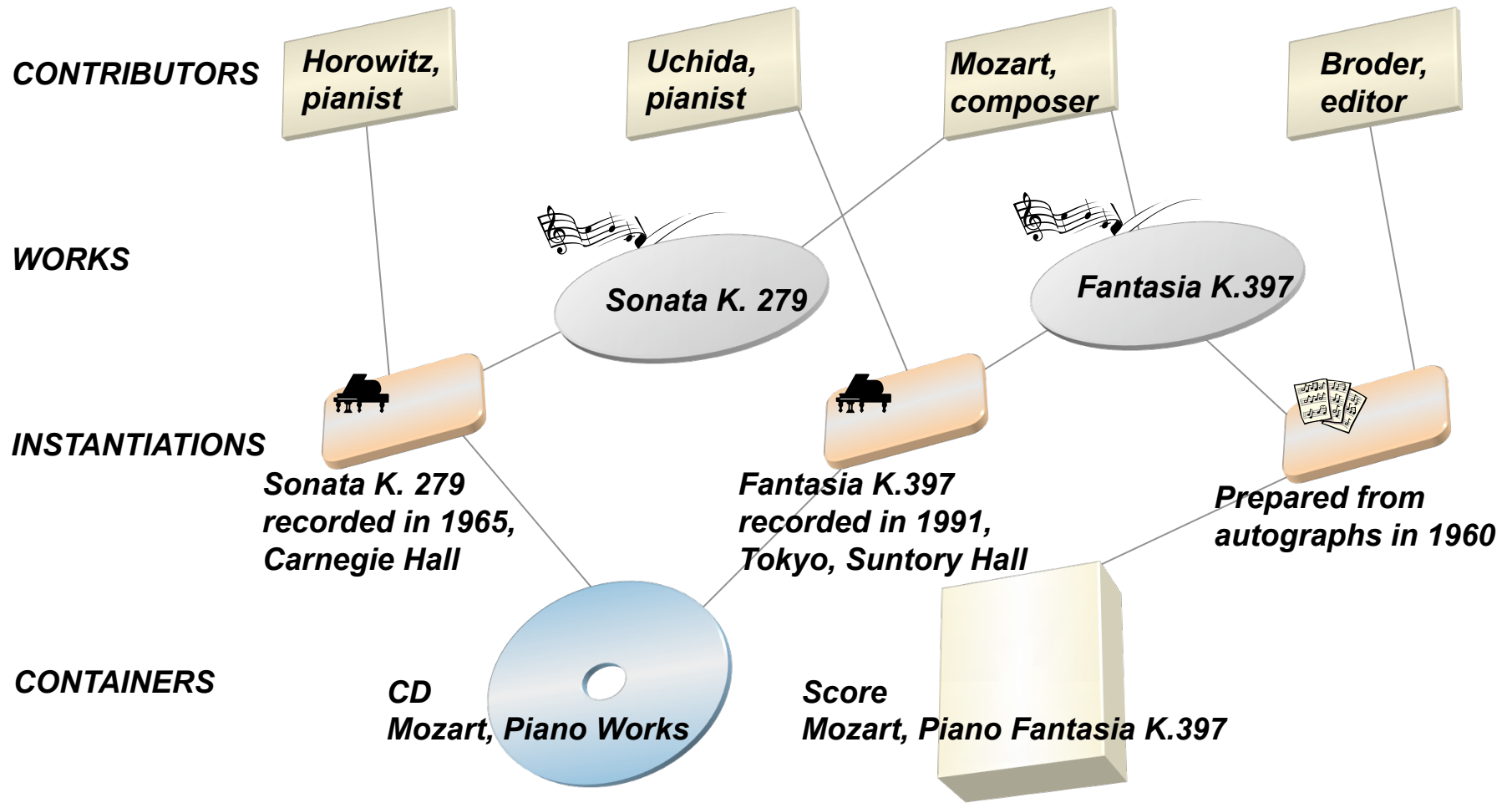
INDIANA UNIVERSITY

Variations2 project & FRBR

- Variations2 project (2000-2005) developed a FRBR-like data model and search
- Required additional hand-cataloging beyond the MARC record importing to function
- Cataloging was done by grant-funded workers
- Unsustainable model—we never got much above 20% cataloged



V2 Data Model Example



Variations2 Search Window 1

File Edit Bookmarks Window Help

Basic | Advanced | Keyword | Browse

Creator / Composer Examples: bach, "Johann Sebastian", moz, faure

Performer / Conductor Examples: bernstein, nbc orch, beatle








Work Title Examples: sonat, sym 94, "symphonies, H.I, 94", hymn

Key **Media Format**

Sort By: Show:

13 Works matching sonata were found where:
- the Creator matched **Bach, Johann Sebastian**

[Select All](#)

1. Work Title: [Sonatas, organ, BWV 529, C major](#) 
Composer: [Bach, Johann Sebastian](#) 1685-1750 
2. Work Title: [Sonatas, violin, continuo, BWV 1021, G major](#) 
Composer: [Bach, Johann Sebastian](#) 1685-1750 
3. Work Title: [Sonatas, violin, continuo, BWV 1023, E minor](#) 
Composer: [Bach, Johann Sebastian](#) 1685-1750 
4. Work Title: [Sonatas, violin, harpsichord, BWV 1014, B minor](#) 



The V in V/FRBR

- Originally, VFRBR search envisioned as a replacement for the search in Variations
- But
 - We wanted to include all recordings and scores, not just the digitized ones
 - Very few Variations adopters were interested in adopting music-specific discovery or cataloging
 - Variations has moved away from providing discovery—defaults to no search window
 - So now, Variations is decoupled from discovery; discovery is rebranded as **Scherzo**



V/FRBR Schema Development

- Locally developed a suite of FRBR Schemas
- To provide a model for others encoding and sharing FRBRized data
- 3-level approach:
 - frbr – strict interpretation of FRBR report(s)
 - efrbr (extended FRBR) – make FRBR useful
 - vfrbr (Variations/FRBR) – add/remove data elements to optimize model for music
- Covers Group 1, 2, and 3 Entities, plus Relationships
- Created record packaging structure



From Variations2 to FRBR

V2 Data Model →	V/FRBR Schema	Examples
Work	Work (abstract creative entity)	Symphony
Instantiation	Expression (realization of work via performance or scoring)	Concert or Critical edition
Container	Manifestation (embodiment via publication)	CD or Book



People and Dates

V/FRBR Schema	Examples	People	Dates
Work (abstract creative entity)	Symphony	Composer Librettist	Composition 1 st Performance
Expression (realization of work via performance or scoring)	Concert or Critical edition	Performer Conductor Editor Arranger	Performance
Manifestation (embodiment via publication)	CD or Book	Producer	Publication



FRBRization Process

- Started w/MARC Bib and Authority Files
 - ~ 80,000 recordings
 - ~ 100,000 scores
 - Authority files fetched via z39.50
- Identify works and people
 - If we've already seen this one, just link to it
 - If we haven't, see if we have an authority file
 - If not, create a new record
- Map fields
- Geared specifically for music



Work Identification Algorithm

Uses clues in MARC bib records to pull out works

- Presence of fields, subfields, and indicators
- Values of subfields compared to Collective Title and Forms lists

If the value in 240 |a equals the phrase "Chamber Music" do not identify 240 as a work



Example mapping rules

Work from Authority record

- Uniform Title 100,110,111 |t |m |n |
r
- Instrumentation 100,110,111,130 |m
-- make separate entries from each
string delimited by comma; do not
include (x); map value inside () to
number



Some Issues with Work Identification

- 31,340 total Manifestations with no Works
 - 19,017 recordings (22%)
 - 12,323 scores (12%)
- Reasons for work identification failure
 - **Works represented in inaccessible formats**
 - IU recordings – sheer volume precludes full cataloging
 - Soundtracks – considered works (work may be present in 245, but algorithm doesn't detect)
- Works may not match when they should
 - Differences or typos in names could cause a new work to be created when it shouldn't be



Inaccessible Work Information

- Many recordings just have contents notes:
 - 505: 0 : So what -- Freddie Freeloader -- Blue in green -- All blues --Flamenco sketches.
 - 505: 0 : So what (9:02) -- Freddie freeloader (9:33) -- Blue in green (5:26) -- All blues (11:31) -- Flamenco sketches (9:25).
 - 505: 00 : |gCD side.|tSo what|g(9:22) --|tFreddie Freeloader|g(9:46) --|tBlue in green|g(5:37) --|tAll blues|g(11:33) --|tFlamenco sketches|g(9:26) --|tFlamenco sketches|g(alternate take)|g(9:32).



Relating Performers to Works

Three examples from three bib records:

- 511: 0 : Miles Davis, trumpet ; Julian "Cannonball" Adderley, alto saxophone (except #3) ; John Coltrane, tenor saxophone ; Wynton Kelly, piano (#2) ; Bill Evans, piano (all others) ; Paul Chambers, bass ; Jimmy Cobb, drums.
- 511: 0 : Miles Davis, trumpet ; Julian Adderl[e]y, alto saxophone (in 1st- 2nd, 4th-5th works) ; John Coltrane, tenor saxophone ; Wynton Kelly (2nd work) or Bill Evans (remainder), piano ; Paul Chambers, bass ; James Cobb, drums.
- 511: 0 : Miles Davis, trumpet ; Julian "Cannonball" Adderley, alto sax ; John Coltrane, tenor sax ; Wynton Kelly or Bill Evans, piano ; Paul Chambers, bass ; James Cobb, drums.



Scherzo Design Process

- Conducted observations and interviews with 8 participants (students and faculty) using Variations search; made recommendations
- Designed new search based on recommendations, other search experience, and new capabilities (e.g., desire to take advantage of FRBR work-centricity)



INDIANA UNIVERSITY

Scherzo Demo

- Scherzo: <http://vfrbr.info/search>
-



INDIANA UNIVERSITY

Scherzo Analysis



INDIANA UNIVERSITY

Scherzo Evaluation Plan



FRBR Implementation Flavors

Three general approaches to “FRBR”:

1. FRBRize data and store in that form
– V/FRBR
2. Just use FRBR concepts during indexing - Blacklight
3. Apply FRBR concepts w/in MARC –
RDA as being tested now



INDIANA UNIVERSITY

Other Search Experiments

- Virgo: <http://search.lib.virginia.edu>
- Blacklight:
<http://walnut.dlib.indiana.edu:8500/>
(temporary link)



INDIANA UNIVERSITY

For more information

- Try out Scherzo: vfrbr.info/search
- vfrbr.info – project's public site
 - Schemas & sample instance files
 - FRBRization algorithm documentation
 - Papers & presentations