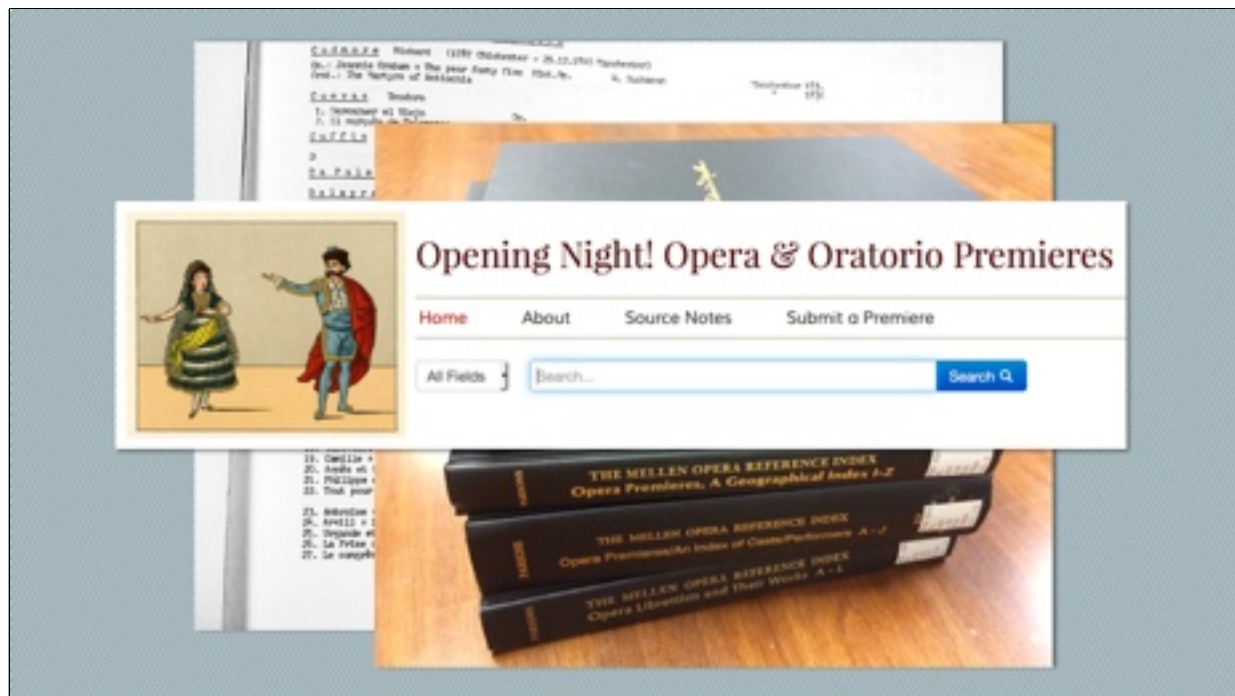


TEI & MEI MODELS FOR RESEARCHING LIBRETTOS & PERFORMANCE HISTORY

DAVID A. DAY
BRIGHAM YOUNG UNIVERSITY
IAML, ROME, 2016

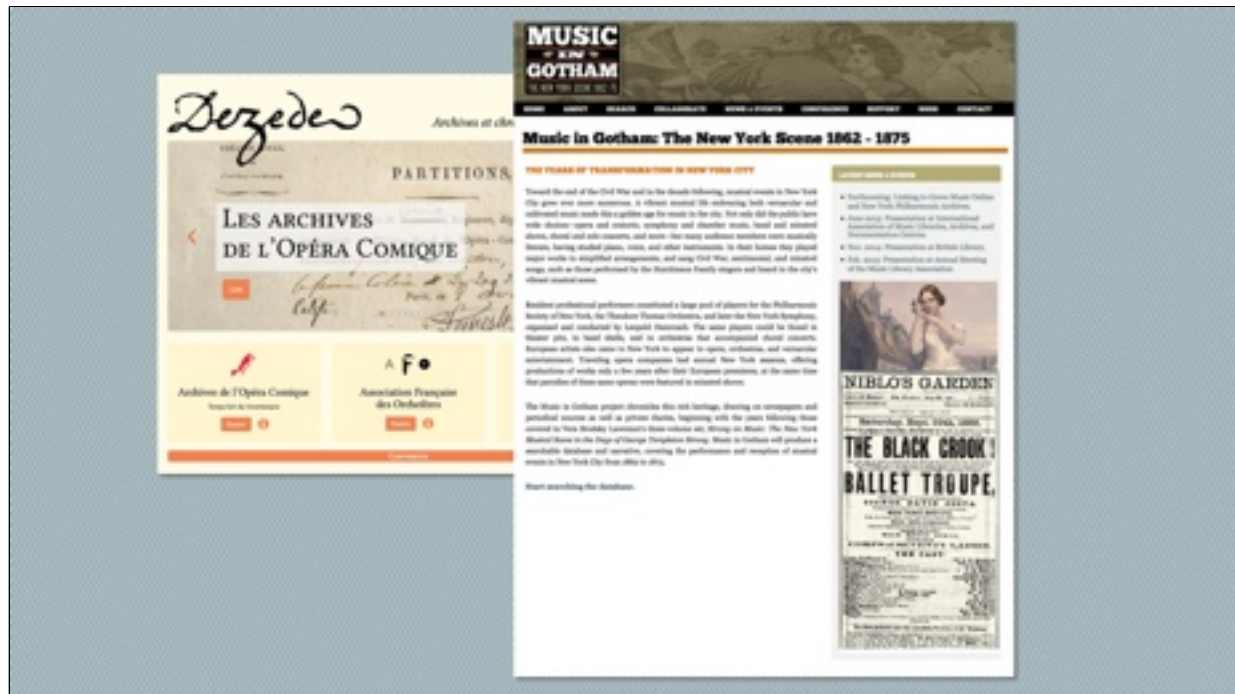
The potential research value of librettos and other forms of performance documentation is generally acknowledged within the music library and musicology communities. Numerous print and online reference sources that document the repertory of dramatic music were compiled, in part, using information found in these primary sources.



and the “Opening Night!” index hosted by Stanford University.



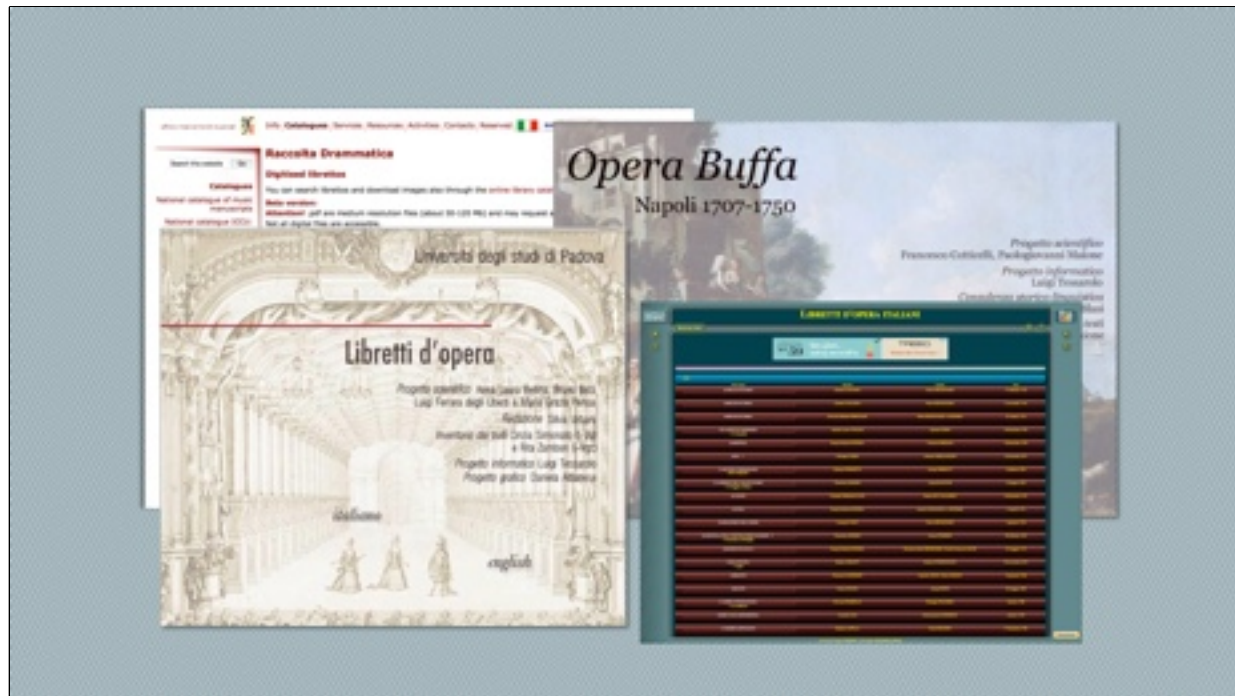
A growing number of performance indexes such as “Dezede”



the “Music in Gotham” project



And AusStage focus on chronologies.



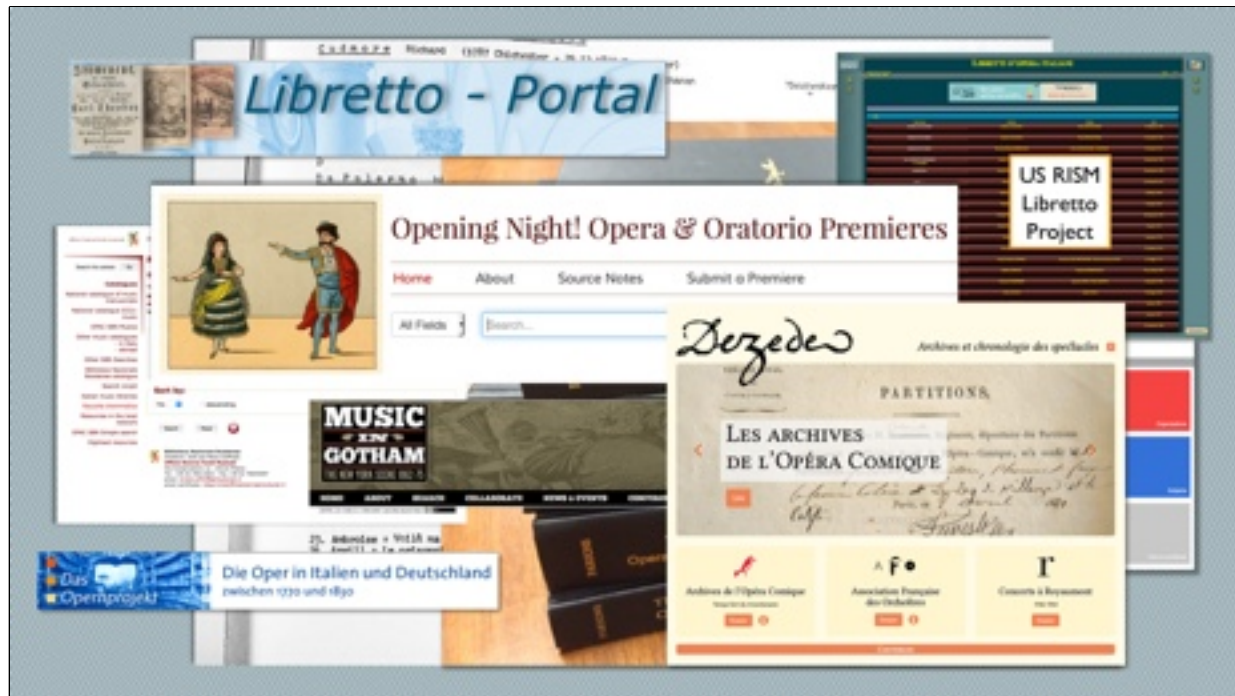
Other noteworthy indexes and catalogs devoted specifically to opera librettos include Claudio Sartori's *I libretti Italiani*, several websites devoted to Italian libretto collections,



the “Libretto Portal,” and “Das Opernprojekt.”



We should also acknowledge the US RISM Libretto Project and current efforts to link those catalog records to scans of the corresponding librettos.



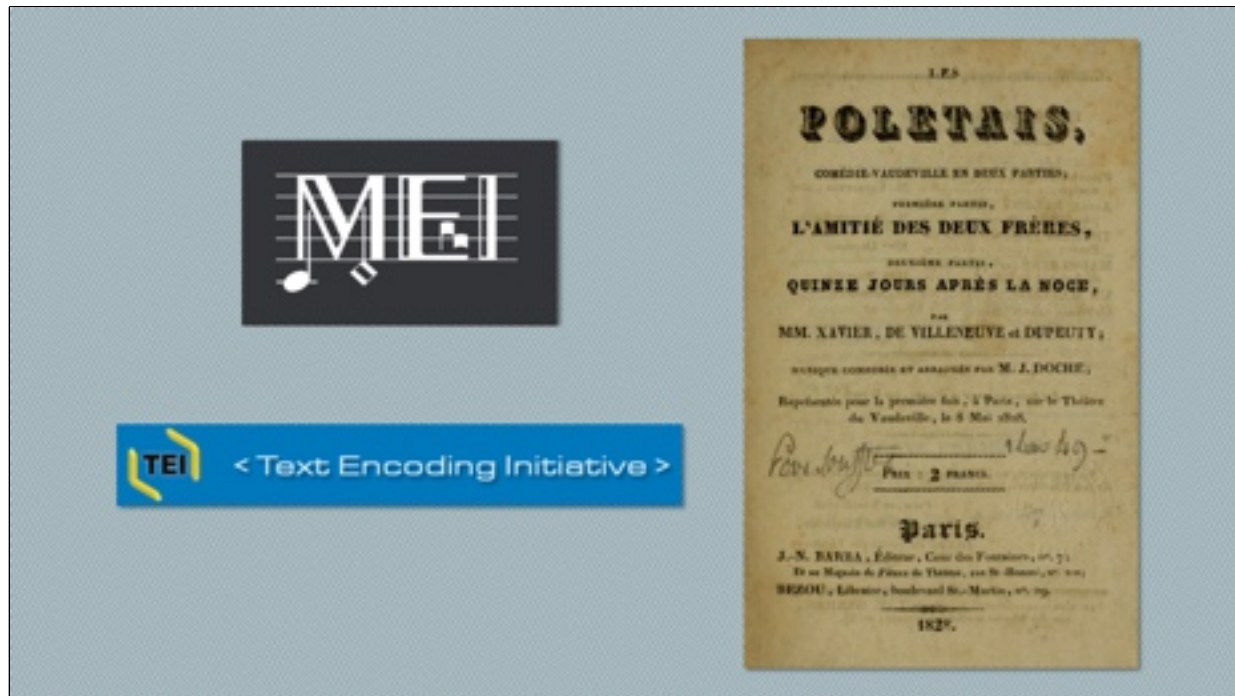
All of these resources help advance traditional scholarship, yet all must be accessed independently and there is currently no means to combine their data for computer indexing and analysis as practiced in the digital humanities. They also lack an advanced collective query that could facilitate automated answers to sophisticated research questions.



The IAML Working Group on Access to Performance Ephemera was initiated in 2003 in large part to explore the possibility of organizing the information recorded IN these kinds of documents. To-date the Working Group has achieved noteworthy progress by identifying major institutional holdings in the database “Concert Programs.”



In more recent years, the IAML conferences at Dublin in 2011 and Antwerp in 2014 hosted special forums on the topic of librettos, in which various ideas on how to promote library holdings were discussed, but without any coordinated effort being established.



In my presentation today, I would like to revisit this discussion and report on two encoding schemas that might now be used to exploit the full research potential of librettos and other forms of performance documentation.

IMPORTANCE OF LIBRETTOS AND OTHER PERFORMANCE DOCUMENTATION

Casts with specific roles
Premier information
Performance venues
Associated creators
etc.



These documents are noteworthy in many regards. They frequently include casts of performers associated with specific roles. They may include useful information on premiers, performance venues, and associated creators such as scene and costume designers.

IMPORTANCE OF LIBRETTOS AND OTHER PERFORMANCE DOCUMENTATION

Casts with specific roles
Premier information
Performance venues
Associated creators
etc.



For those who have long advocated more detailed description and analysis of these sources, I believe the objective has been to facilitate answers to advanced research questions like the following:

DATA ENABLED RESEARCH QUESTIONS

- Who are the artists that performed a given role?
- What roles did a given artist perform?



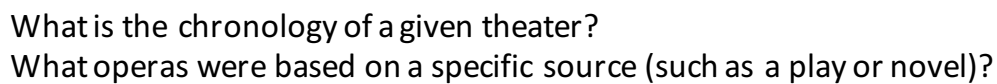
Who are the artists that performed a given role?
Conversely, what roles did a given artist perform?

- How did the characters in a opera evolve or transform at different venues?
- What was the popularity of a opera during a given time period or at geographical location?



What was the level of popularity of an opera during a given time period or at a geographical location?

- What is the chronology of a given theater?
- What operas were based on a given source (such as a play or novel)?



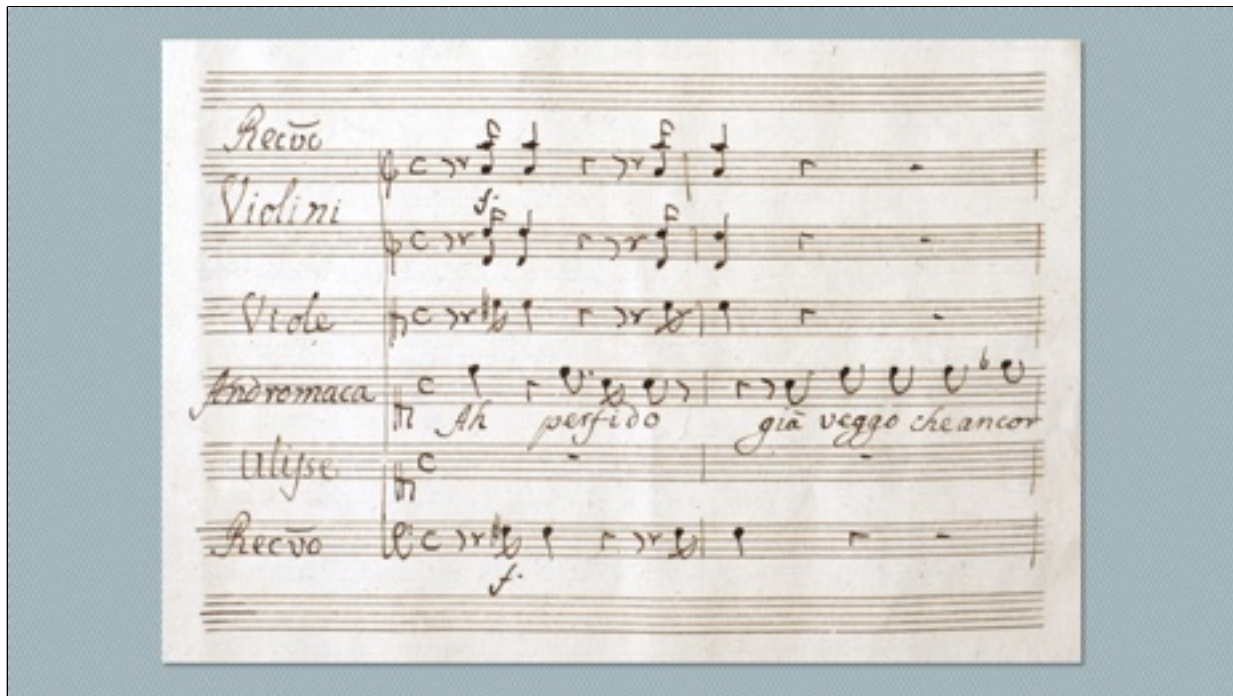
DATA ENABLED RESEARCH QUESTIONS

- What was the popularity of a given plot source over time and in different locations?
- What topical themes were common in operas and other dramatic music, when and where?

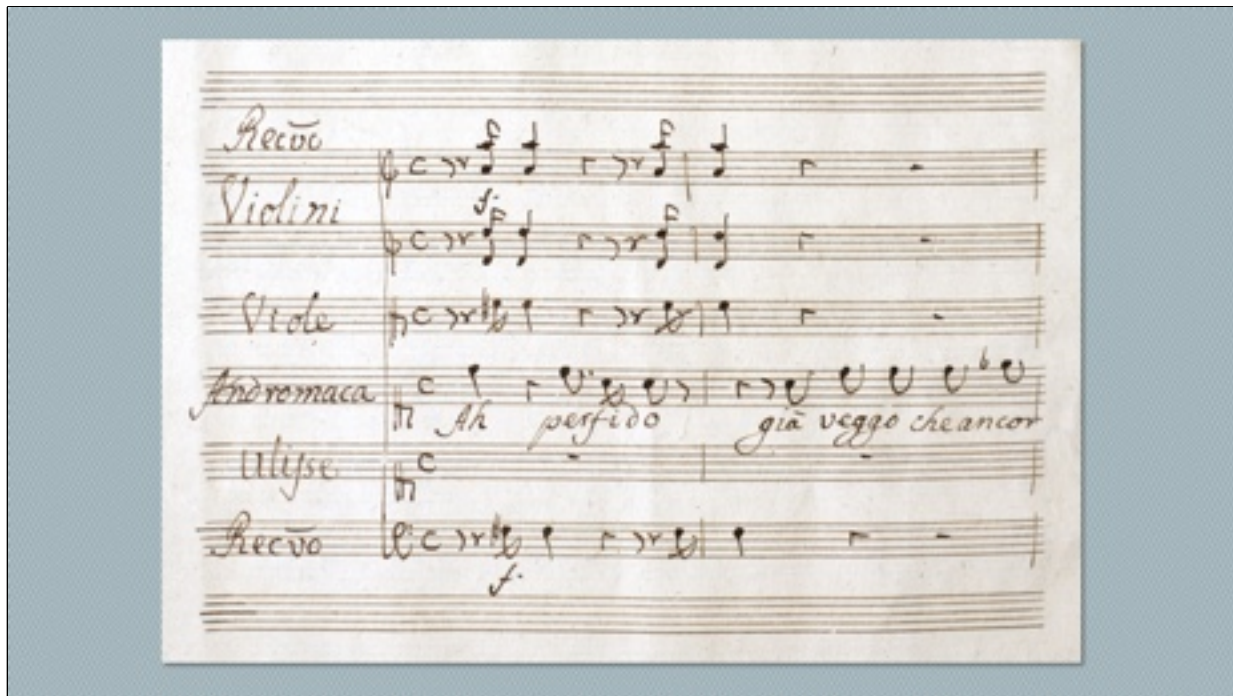


What was the popularity or longevity of a plot source over time and in different locations.

What topical themes were common in operas and other forms of dramatic music, when and where?



A well-designed data analysis system could result in additional benefits such as an index to lyric incipits. A more advanced system might also include music incipits.



Such a design could facilitate research on what musical themes appear repeatedly in different works and how this use of musical quotation helped convey or underscore the plot?

WHY NOW?

If the potential to use these documents to answer such broadly based questions has been acknowledged in the past, and yet we have not succeeded in creating a system or consortia for the desired results, you might ask why revisit the issue now? I believe there are two important and related factors that should prompt our re-examination of what can be done.



One development that should catch our attention is the growing body of librettos, concert programs, and the musical press currently accessible in full text online. My own efforts to index opera and ballet primary sources show that more than 40,000 librettos alone are available.



While not all of these scanned images are conveniently converted to full text, and while the metadata corresponding to these documents is widely varied, the potential body of data useful for advanced research invites a fresh look at the possibilities.



- Context of large corpus of full text librettos, programs, and the musical press



A second development that could help organize and query these documents is the ongoing progress of the TEI and MEI encoding schemas. It was my awareness of the first factor (that is the vast body of online resources) that prompted my interest in the potential of TEI and MEI.



- Context of large corpus of full text librettos, programs, and the musical press



For those who may not be familiar with the basics of the Text Encoding Initiative (TEI) and the Music Encoding Initiative (MEI), these are two schemas or systems of XML tags that can be used to mark or identify characteristics of words in a text or elements of musical notation.




- Context of large corpus of full text librettos, programs, and the musical press
- Ability to tag key elements of a text

In the context of a large corpus of full text librettos, it seems feasible that a project might be undertaken to mark up these texts using TEI tags to identify names, titles, dramatic topics, venues, dates, and so on. From the beginning I assumed that any efforts following this model would require some kind of linking to substantial authority files.



- Context of large corpus of full text librettos, programs, and the musical press
- Ability to tag key elements of a text

When using a TEI model to analysis full texts, some have argued that the resulting encoded documents would still be mostly unstructured. And it might require complicated programing to create a query interface capable of answering the research questions mentioned earlier.



< Text Encoding Initiative >

- Context of large corpus of full text librettos, programs, and the musical press
- Ability to tag key elements of a text
- Potential textual analysis

On the other hand, TEI encoding of full texts may have merit for certain forms of deep learning, neural networks, and other textual analysis and should be fully explored.



- Context of large corpus of full text librettos, programs, and the musical press
- Ability to tag key elements of a text
- Potential textual analysis
- Text incipit indexes

The creation of lyric or text incipit indexes might also require some use of a libretto's full text. Furthermore, it is true that TEI may have certain advantages when working with concert programs that usually list more than one work in a given performance.



- Context of large corpus of full text librettos, programs, and the musical press
- Ability to tag key elements of a text
- Potential textual analysis
- Text incipit indexes

More advanced textual analysis of full texts might some day be achieved in a combination of MEI and TEI. It is also possible that MEI could be further developed to deal with multi-work events described in concert programs. There are also many new approaches to the study of a text, such as network text analysis and numerous software products. These also should be investigated.



- Context of large corpus of full text librettos, programs, and the musical press
- Ability to tag key elements of a text
- Potential textual analysis
- Text incipit indexes

However, as a starting point, MEI, specifically the metadata side of MEI (which deals with text, rather than notation) is already well developed for the description of musical works and their performances. It may be better suited than other approaches to answer the data-enabled research questions outlined earlier.



- Context of large corpus of full text librettos, programs, and the musical press
- Ability to tag key elements of a text
- Potential textual analysis
- Text incipit indexes

If the end objective is to facilitate answers to data-dependent, but predictable research questions, it is probably more efficient to organize the data in a structured manner. In other words, rather than focusing on the full texts of these documents, it might be better to start with the metadata or cataloging for these same sources.

UNNECESSARY WORK?

I realize there could be an immediate outcry from this audience: Why re-catalog tens of thousands of documents and why not resort to existing bibliographic systems including OCLC, Internet Culturale, Europeana, RISM, and others?

UNNECESSARY WORK?

Earlier in this week we learned about Triple-I-F, which appears to promise the sort of cross system searching that could bring together most of the relevant records. Why not use it?

UNNECESSARY WORK?

You might also be asking, Why not just use the existing MARCXML?

UNNECESSARY WORK?

Perhaps the obvious answer to these questions is that none of these systems are currently capable of facilitating a “big data” research environment. That is my understanding. If I am wrong, you can let us all know at the end of my presentation.

ADVANTAGES TO MEI ENCODING

Beyond what I think is obvious, my response to this anticipated challenge and justification for promoting MEI is four-fold.

ADVANTAGES OF MEI ENCODING

- I. It is not difficult to convert MARC XML to MEI

MARC XML ↔ MEI

First, it is not that difficult to convert MARC and other flavors of XML to MEI. There are already well-developed crosswalks in place.

ADVANTAGES OF MEI ENCODING

- I. It is not difficult to convert MARC XML to MEI

MARC XML ↔ MEI

It is conceivable that if a cooperative effort could be established in which catalogs rich in librettos and other performance documentation were able and willing to provide downloads of their data in MARC XML, these data sets could be converted to MEI and then enhanced in ways necessary to achieve the desired research environment.

ADVANTAGES OF MEI ENCODING

- I. It is not difficult to convert MARC XML to MEI

MARC XML ↔ MEI

So, let me emphasize that I am not suggesting all of these documents be re-cataloged, rather I am questioning if it is feasible to consolidate data from existing catalogs and convert them to a uniform style and schema.

ADVANTAGES OF MEI ENCODING

1. It is not difficult to convert MARC XML to MEI
2. MEI is tailored to works and performances



My second defense of MEI tagging is that its metadata schema is richly developed in a manner specifically tailored to musical works and performances. These advances in the MEI schema can be credited to several projects including the Carl Nielsen online thematic catalog and the Detmold Court Theater project.

ADVANTAGES OF MEI ENCODING

1. It is not difficult to convert MARC XML to MEI
2. MEI is tailored to works and performances
3. Ability to encode both text and musical notation

```
<define name="data.ACCIDENTAL.IMPLICIT">
  <choice>
    <value>s</value>
    <a:documentation xmlns:rng="http://relaxng.org/ns/structure/1.0"
      xmlns:sch="http://purl.oclc.org/dsdl/schematron"
      xmlns:a="http://relaxng.org/ns/compatibility/annotations/1.0">sharp</a:documentation>
    <value>f</value>
    <a:documentation xmlns:rng="http://relaxng.org/ns/structure/1.0"
      xmlns:sch="http://purl.oclc.org/dsdl/schematron"
      xmlns:a="http://relaxng.org/ns/compatibility/annotations/1.0">flat</a:documentation>
  </choice>
</define>
```

A third defense of MEI, is the ability to encode both text and musical notation in a single system. As I mentioned earlier, in an ideal environment, we would have the ability to research musical themes in relation to topical content. It would also be ideal to generate an aria or excerpt index that combined both text and music incipits.

ADVANTAGES OF MEI ENCODING

1. It is not difficult to convert MARC XML to MEI
2. MEI is tailored to works and performances
3. Ability to encode both text and musical notation
4. FRBR is inherent to the MEI metadata schema



The fourth advantage is that FRBR is inherent to the design of the MEI metadata schema. A FRBR approach to structuring information about performance documents is preferable due to the hierarchy of works, expressions, manifestations, and items fundamental to dramatic music. I might add this is especially important when dealing with documentation of performances.

QUICK REVIEW

At this point let's pause and review a few key points.

QUICK REVIEW

- I. Tradition of using librettos and other performance documents for research

First, there is a long tradition of using librettos and other performance ephemera to document and analyze the repertory of dramatic music.

QUICK REVIEW

1. Tradition of using librettos and other performance documents for research
2. IAML's interest and activities to-date

Second, IAML, as manifest in the activities of the Working Group on Access to Performance Ephemera and recent forums, has shown a desire to facilitate greater access to and manipulation of the information found in these resources.

QUICK REVIEW

1. Tradition of using librettos and other performance documents for research
2. IAML's interest and activities to-date
3. Rapid growth of digital collections

Third, rapid growth in digital collections that include the relevant documents begs our attention and reconsideration of the possibilities.

QUICK REVIEW

1. Tradition of using librettos and other performance documents for research
2. IAML's interest and activities to-date
3. Rapid growth of digital collections
4. A well-developed MEI schema to facilitate advanced research

And fourth, the MEI metadata schema now provides a means to encode the information researchers want in a way that can facilitate advanced queries.



For the truly optimistic, and enthusiastic devotees of dramatic music, it would appear, that at last, the stars are aligned! But, the pragmatists among us will question the practicality of everything I have suggested. For sure, the obstacles are real and sometimes very complicated, but I am optimist and believe they can be overcome.

CHALLENGES & SOLUTIONS

Let's consider a few of the most obvious challenges and how they might be addressed.

CHALLENGES & SOLUTIONS

I. Lack of a consortia of shared records, or XML downloads

Perhaps the first obstacle is the fact that although many library catalogs and larger consortia databases already included a wealth of bibliographic description for the sources we want, very few provide a means to export the full results of a search to MARC or some other flavor of XML that could then be converted to MEI.

CHALLENGES & SOLUTIONS

- I. Lack of a consortia of shared records, or XML downloads

To the best of my knowledge, the various performance indexes also do not have a convenient means of sharing their data, which is often uniquely structured.

CHALLENGES & SOLUTIONS

I. Lack of a consortia of shared records, or XML downloads

I have a few ideas on how to overcome this obstacle. In one approach, Triple-I-F or other methods of query might help provide a means to export large quantities of records. From another angle, perhaps IAML could use its influence to bring together appropriate parties and simply request that they share their data for the greater good. If this first obstacle could be overcome, it would open a door to the creation of a vast data set.

CHALLENGES & SOLUTIONS

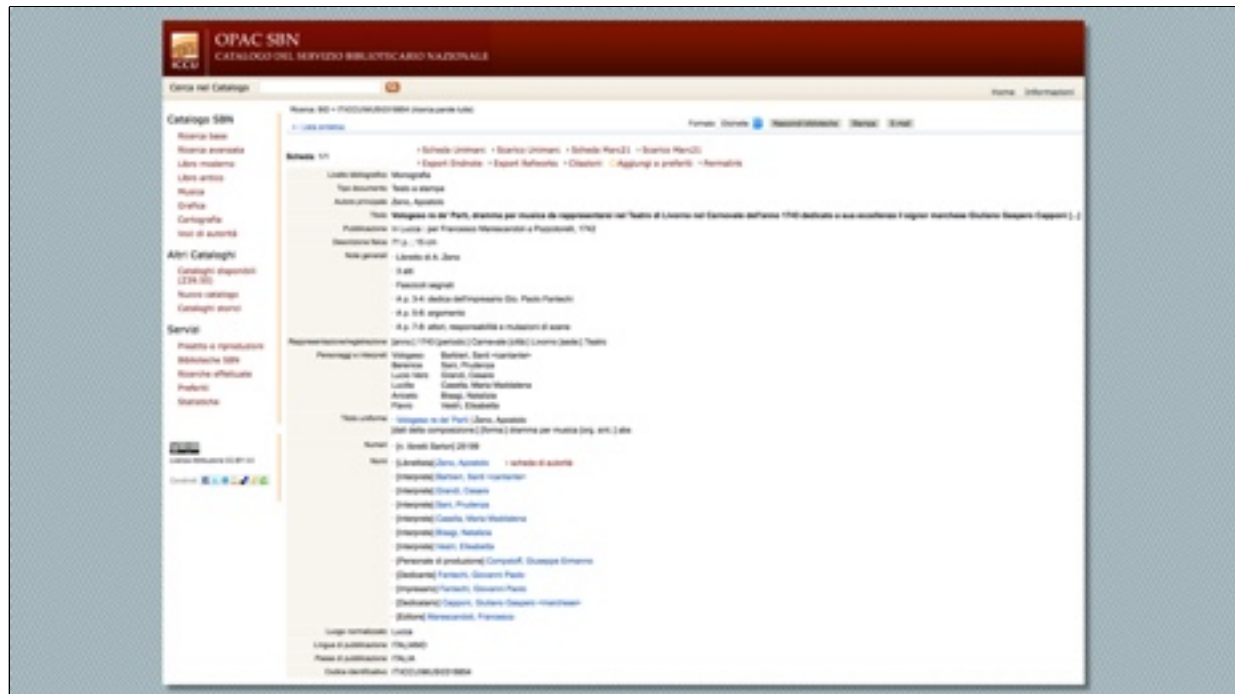
I. Lack of a consortia of shared records, or XML downloads

These consolidated records could then be converted to MEI and harmonized, edited, and enhanced to facilitate the desired results. It is even conceivable that monumental print reference sources might also be converted to full text and treated in a similar fashion.

CHALLENGES & SOLUTIONS

1. Lack of a consortia of shared records, or XML downloads
2. Lack of unified descriptive standards

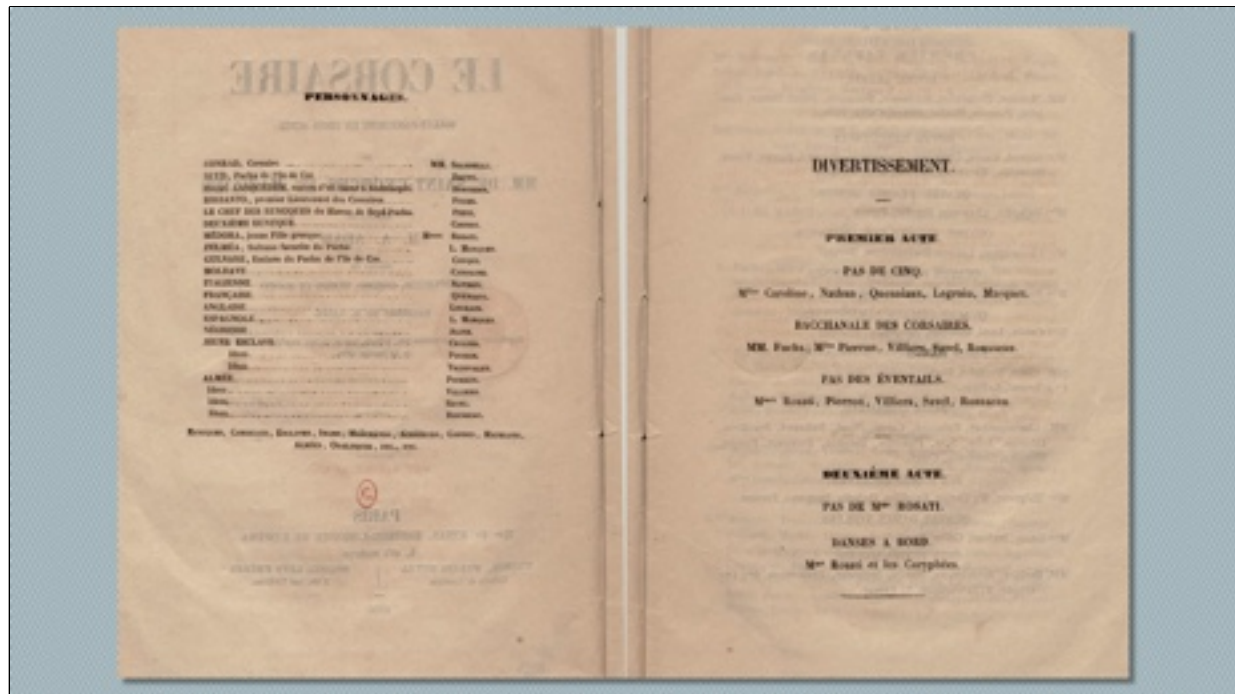
A second obstacle to face in organizing all of this data is the weakness of unified descriptive standards. Even though the bulk of metadata available is MARC based, in practical terms, the data is far from standardized and compatible. Time will not permit a detailed look at all the problems, but I will illustrate the point with two examples.



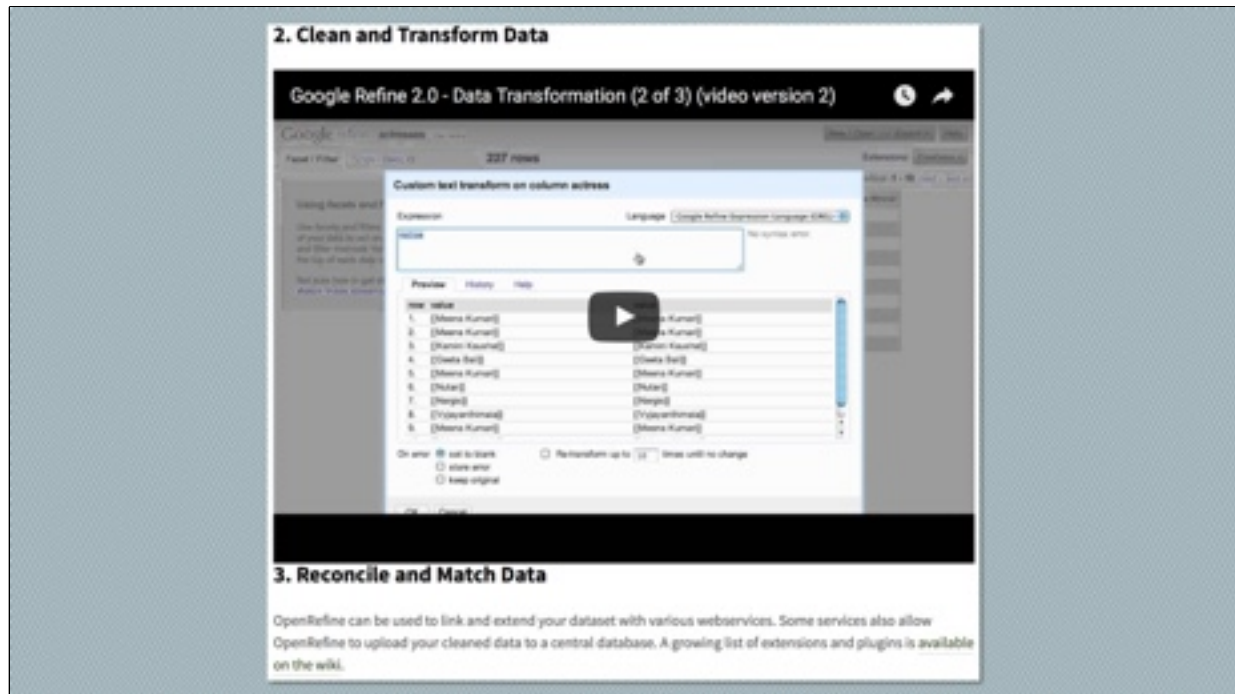
This example of beautifully complete cataloging is taken from the OPAC SBN. In this case there is an extensive listing of artists and related creators.

Personaggi e interpreti	Vologeso	Barbieri, Santi <cantante>
	Berenice	Sani, Prudenza
	Lucio Vero	Grandi, Cesare
	Lucilla	Casella, Maria Maddalena
	Aniceto	Bisagi, Natalizia
	Flavio	Vestri, Elisabetta
Titolo uniforme	[titolo] [n. libretti Sartori] [data] [forma:] dramma per musica [org. sint.] abs	
Numeri	[n. libretti Sartori] 25199	
Nomi	- [Librettista] Zeno, Apostolo > scheda di autorità - [Interprete] Barbieri, Santi <cantante> - [Interprete] Grandi, Cesare - [Interprete] Sani, Prudenza - [Interprete] Casella, Maria Maddalena - [Interprete] Bisagi, Natalizia - [Interprete] Vestri, Elisabetta - [Personale di produzione] Compstoft, Giuseppe Ermanno - [Dedicante] Fantechi, Giovanni Paolo - [Impresario] Fantechi, Giovanni Paolo - [Dedicatario] Capponi, Giuliano Gaspero <marchese> - [Editore] Marescandoli, Francesco	

The artists are also associated with their specific roles.



Even though this information is described in great detail in the libretto itself.

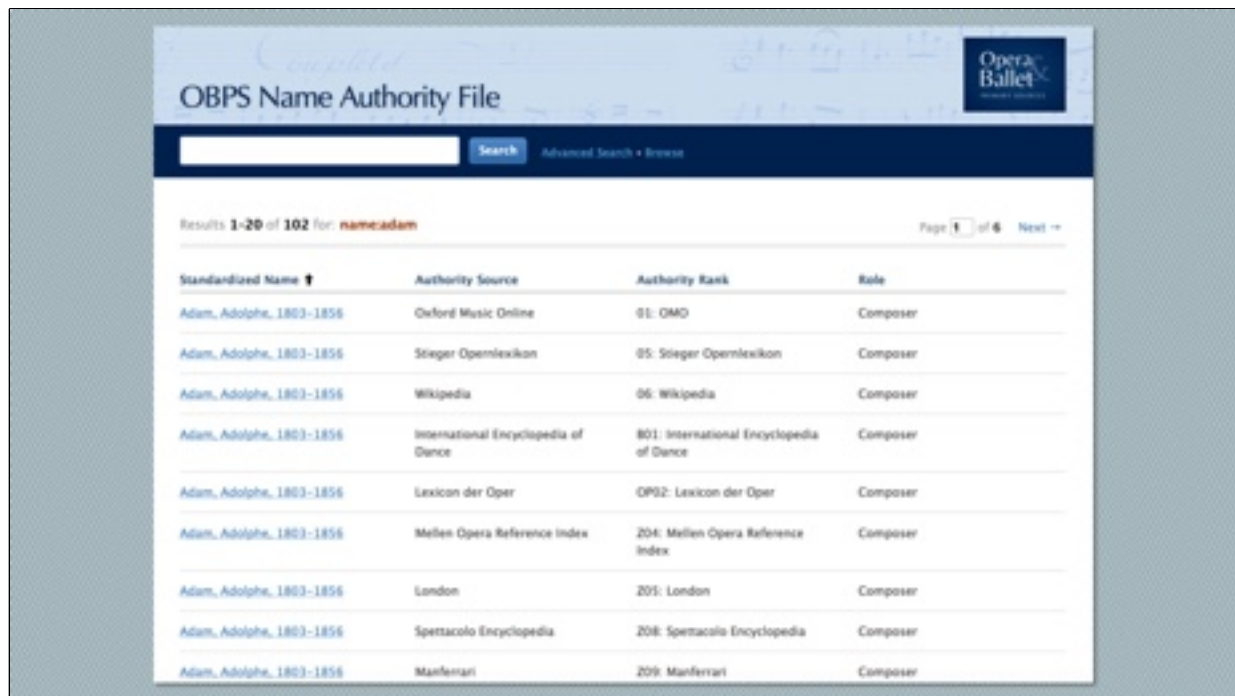


The problems of inconsistent and incomplete metadata can be overcome through standardization and enhancement of minimal records. The process of harmonization can be aided with relatively new tools like Google Refine. The enhancement of minimal records, however, would be a more costly and time-consuming proposition.

CHALLENGES & SOLUTIONS

1. Lack of a consortia of shared records, or XML downloads
2. Lack of unified descriptive standards
3. Lack of name and title authority control

A third obstacle to full utilization of all the texts that are available is the lack of name and title authority control. While major composers and librettists are generally well established in authority files such as VIAF and LCNA, once you reach a level of coverage that includes more obscure works, name authority is problematic.



The screenshot shows the OBPS Name Authority File interface. At the top, there is a search bar with the text "Search" and a link to "Advanced Search + Filters". Below the search bar, the results are displayed as a table. The table has four columns: "Standardized Name", "Authority Source", "Authority Rank", and "Role". The results are for the search term "name:adam" and show 102 results, with the first 20 displayed. The results list various authority sources such as Oxford Music Online, Steier Opernlexikon, Wikipedia, International Encyclopedia of Dance, Lexicon der Oper, Mellen Opera Reference Index, London, Spettacolo Encyclopedia, and Manferriani, all identifying the role as "Composer".

Standardized Name ↑	Authority Source	Authority Rank	Role
Adam, Adolphe, 1803-1856	Oxford Music Online	01: OMO	Composer
Adam, Adolphe, 1803-1856	Steier Opernlexikon	05: Steier Opernlexikon	Composer
Adam, Adolphe, 1803-1856	Wikipedia	06: Wikipedia	Composer
Adam, Adolphe, 1803-1856	International Encyclopedia of Dance	01: International Encyclopedia of Dance	Composer
Adam, Adolphe, 1803-1856	Lexicon der Oper	002: Lexicon der Oper	Composer
Adam, Adolphe, 1803-1856	Mellen Opera Reference Index	204: Mellen Opera Reference Index	Composer
Adam, Adolphe, 1803-1856	London	205: London	Composer
Adam, Adolphe, 1803-1856	Spettacolo Encyclopedia	208: Spettacolo Encyclopedia	Composer
Adam, Adolphe, 1803-1856	Manferriani	209: Manferriani	Composer

As a semi-scientific assessment, the OPBS name authority file includes about 9,000 names of composers and librettists. Only about 45 percent of these names or less than half are currently found in VIAF. It would be even more challenging to establish name authority for the casts and other associated creators.

CHALLENGES & SOLUTIONS

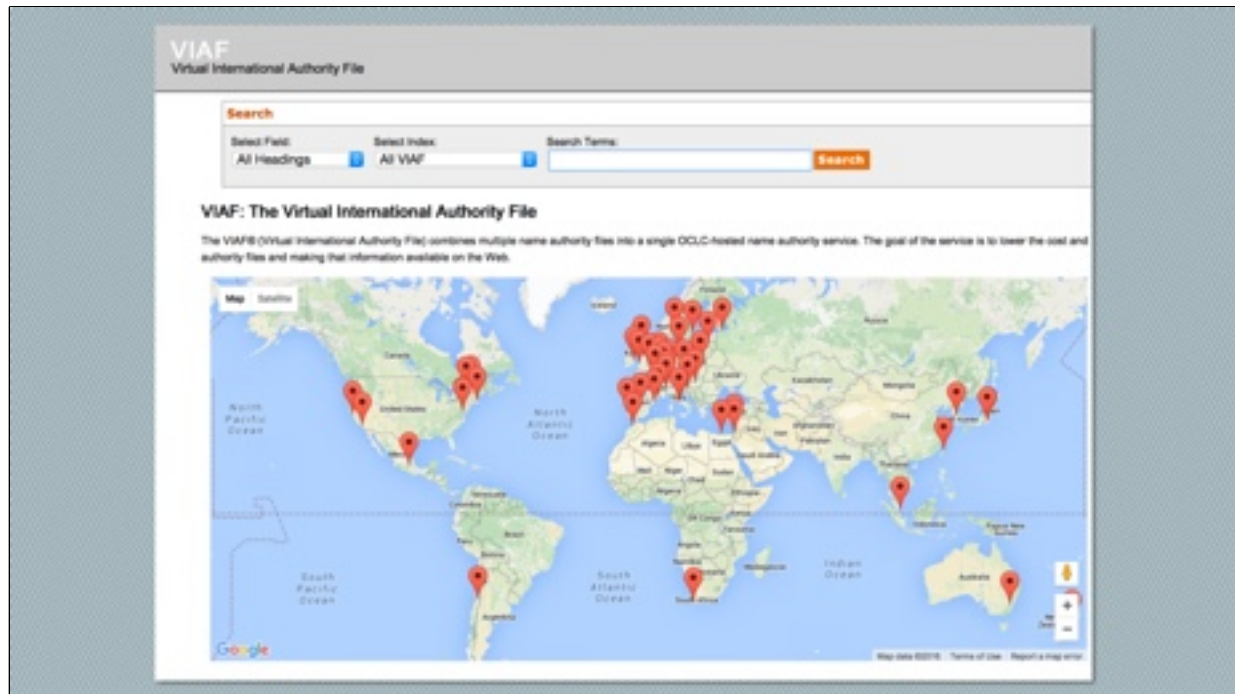
1. Lack of a consortia of shared records, or XML downloads
2. Lack of unified descriptive standards
3. Lack of name and title authority control

When it comes to title authority, dramatic works are inherently complicated. Even a premier title can appear differently among the various sources associated with the first performance including the, libretto, scores (possibly both print and manuscript), the program announcement, and potentially multiple press releases.

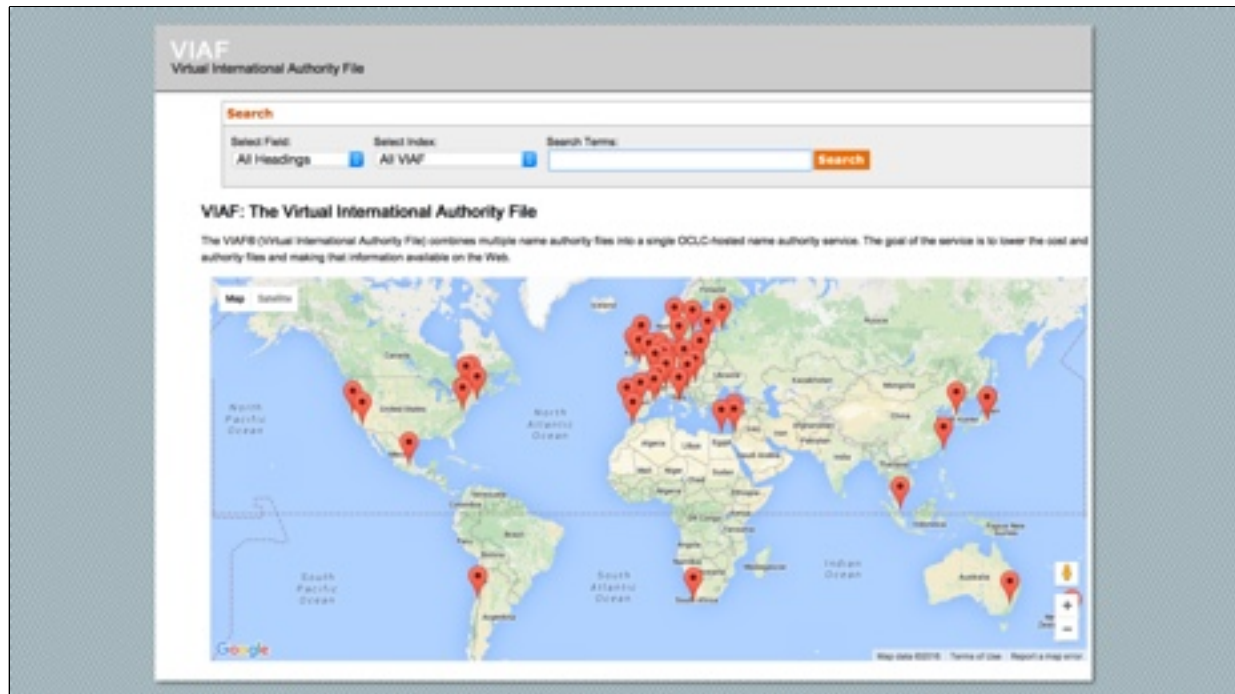
CHALLENGES & SOLUTIONS

1. Lack of a consortia of shared records, or XML downloads
2. Lack of unified descriptive standards
3. Lack of name and title authority control

Beyond establishing the title of the premier, most works will have variants stemming from productions at different theaters and revisions of the work. Again, title authority for the major established repertory is advanced, but control for the more obscure titles is wanting.



The task of overcoming the challenges associated with authority control will definitely require considerable effort and coordination. The optimist, however, will see the bright side of the existing accomplishments and infrastructure found in VIAF.

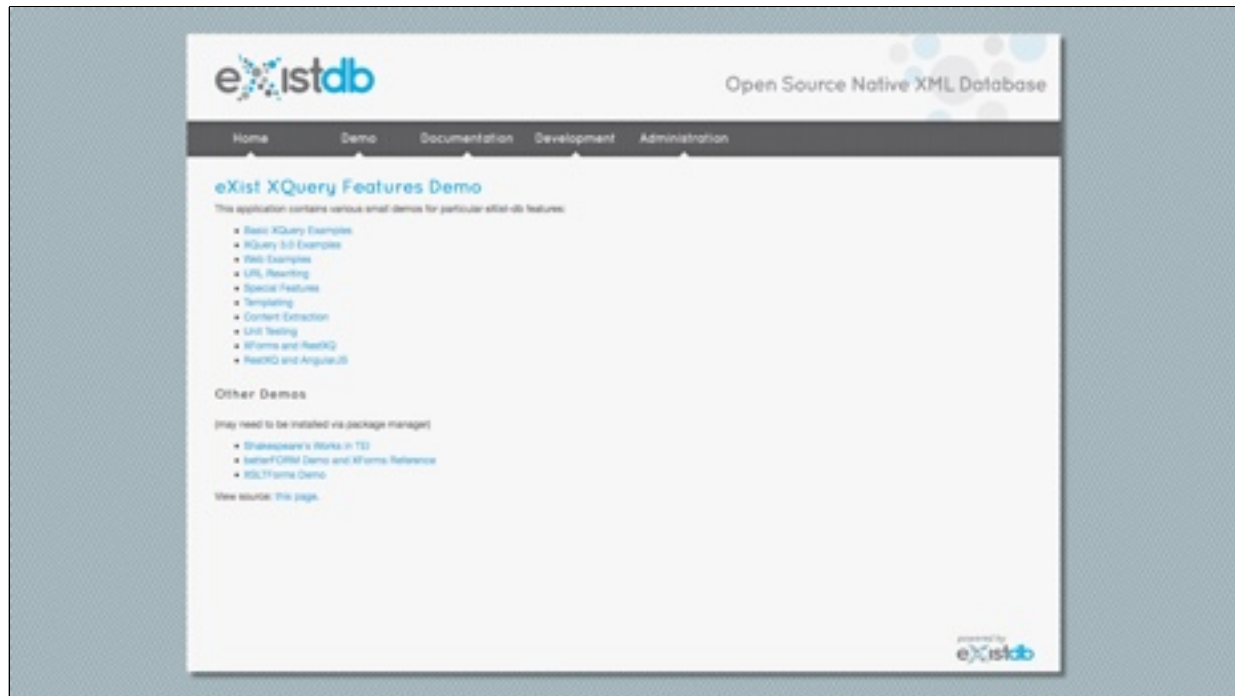


Perhaps the best approach to solving the problem is simply a coordinated effort to enhance VIAF for dramatic music. Such a project might include authority work for titles and concert venues as well as names.



BEYOND

Beyond the obstacles related to data harmonization and authority control, time is too short to even approach the topic of how to build a query interface once the data is ready.



There are, however, open source solutions like Existdb that are available and that have been used in similar applications elsewhere.



BEYOND

Another area that could and should be investigated more thoroughly is how to manage a potential two-track approach to dealing with the full texts. While, as a start, I have advocated a structured approach focused on enhancement of existing metadata, the complementary approach of analyzing full texts should also be investigated in greater detail.

CONCLUSIONS

In conclusion, I hope that the investigation I have presented might stir the imagination of those who believe in the research potential of these documents. I hope that IAML, perhaps RISM, or even just a few interested colleagues, will consider if they can play a role in promoting the tasks of record sharing, data harmonization, and authority work that I have touched on only briefly.

CONCLUSIONS

Even as an optimist, I do not believe it is practical to undertake an ambitious project to do it all. I would, however, propose a modest first step of establishing a Project Group to explore and test some basic propositions. Or as an alternative ask that the existing Study Group consider the same questions:

QUESTIONS AND PROJECTS TO TEST

Here are some questions that could be explored and tested:

QUESTIONS AND PROJECTS TO TEST

I. Amalgamation of metadata from relevant catalogs and indexes

1. Can IAML use its influence to accomplish AN amalgamation of metadata from all or part of the relevant catalogs and indexes. Can IAML help facilitate XML exports of appropriate records? As a possible experiment the US RISM Libretto Project and records for librettos in SBN might be combined to provide a substantial foundation to build on.

QUESTIONS AND PROJECTS TO TEST

1. Amalgamation of metadata from relevant catalogs and indexes
2. Convert relevant print reference sources to full text

2. Can IAML facilitate projects to convert relevant print reference tools (for example Sartori's *I libretti italiani*) to electronic full texts?

QUESTIONS AND PROJECTS TO TEST

1. Amalgamation of metadata from relevant catalogs and indexes
2. Convert relevant print reference sources to full text
3. Identify ways to harmonize data

3. Can IAML or other interested scholars identify ways to harmonize data found in the full spectrum of both print and online resources once they are consolidated?

QUESTIONS AND PROJECTS TO TEST

1. Amalgamation of metadata from relevant catalogs and indexes
2. Convert relevant print reference sources to full text
3. Identify ways to harmonize data
4. Convert consolidated data to MEI

4. Can IAML help facilitate conversion of this refined data to MEI? Are the resources and expertise already available?

QUESTIONS AND PROJECTS TO TEST

1. Amalgamation of metadata from relevant catalogs and indexes
2. Convert relevant print reference sources to full text
3. Identify ways to harmonize data
4. Convert consolidated data to MEI
5. Work with MEI to develop its schema

5. How can IAML work with the MEI community to ensure that the its schema is fully developed to meet the desired research needs? I might add at this point that I believe these first five objectives could be achieved mostly through good will and collaboration. I do not think major outside funding would be a necessity to get started.

QUESTIONS AND PROJECTS TO TEST

1. Amalgamation of metadata from relevant catalogs and indexes
2. Convert relevant print reference sources to full text
3. Identify ways to harmonize data
4. Convert consolidated data to MEI
5. Work with MEI community to develop its schema
6. Promote authority work in VIAF

Finally, how might IAML promote the necessary authority work and make sure all the names and titles in VIAF have identifiers that could be used in an eventual query system? This task might begin with an effort to establish authority for all of the names and titles found in the SBN librettos, or the Opening Night index, or OBPS.



COMMENTS? SUGGESTIONS?

I hope there is time for your feed back ON HOW and if IAML or related organizations like IMS should endeavor to advance this area of research. I welcome your comments and suggestions.



And remember, let's be optimistic!