e-Science and Musicology

Richard Lewis

Goldsmiths College, University of London

e-Science for Musicology Workshop, Edinburgh, 1 July 2009

Goldsmiths
UNIVERSITY OF LONDON

Outline

Goldsmiths
UNIVERSITY OF LONDON

Outline

Goldsmiths
UNIVERSITY OF LONDON

e-Science *n.*

- Become a bit of a buzzword
- Can help secure funding
- Same as/related to 'cyberinfrastructure' (common in US), 'e-Research' (common in Australia)
- Various domain-specific derivations: e-Social Science, e-Humanities

Technology for Research

- Using technology to enhance existing research methods
- Using technology to enable new research methods
- But not research into new technologies

Engineering

- Many e-Science projects are somewhat computer science engineering focused
- Often involve developing tools with which to do science
- (But the tools don't represent breakthroughs in computing research)

What Technologies?

- Distributed computing; make large scale computational resources available to scientists
- Distributed access to large scale datasets; visualisation tools
- Collaboration over digital networks

Who Likes e-Science?

- Often technologies are embraced by working scientists
- Rather than being handed down by science managers
- Potential partnerships with industry and with open source

Outline

Goldsmiths
UNIVERSITY OF LONDON

Three Modes of Knowledge Creation

- Computational discovery
- Comparative research
- Digital library browsing

Automation in Scientific Work

- Many accounts of (future) e-Science describe increased automation
- Cataloguing samples used in experiments
- Automated booking of experimental equipment (telescopes, microscopes)
- Automated configuration of experimental equipment based on previous similar experiments
- Monitoring and recording experimental procedures (potential for re-play)
- Automated notification of results to colleagues
- Automated retrieval of results of similar experiments
- Automated organisation of conference calls to discuss early stage conclusions

Hypothesis Formation

- In the e-Science just described, the technology is simply providing the scientists with access to knowledge and tools
- The *computational discovery* mode of e-Science would afford the technology a role in forming hypotheses

Expert Systems

- Existing AI techniques such as expert systems already make this possible (within limited domains)
- A knowledge base consists of *facts* and *rules*
- An inference engine applies logic to infer new facts from the existing facts and rules
- A natural extension of e-Science's already available large datasets

Comparative Research

- Second mode of knowledge creation is comparative research
- Executing the same (often simple) comparative operation against large numbers of data
- Knowledge comes in the form of statistical trends

Enabling Comparative Research

- Comparative research requires existence of large datasets
- Bandwidth, intuitive interfaces, and authentication mechanisms to access data (provided by sound e-Science funding)
- Mechanism to execute large number of comparison operations (e.g. distributed, parallel computing)

Digital Libraries

- Third mode of knowledge creation is use of digital libraries
- Explicitly human agents interacting with digitally encoded and delivered knowledge
- Being able to make novel connections through ease of access to diverse sources

e-Scientists

- Hints at an important consequence of e-Science
- It will require scientists to become e-Scientists
- They will have to understand the possibilities and limitations of the technology
- Concede some knowledge creation tasks to the technology

So What is e-Science?

- e for ‘enhanced’ (more than for ‘electronic’)
- Three modes of knowledge creation
- One enabled by technology (computational discovery)
- Two enhanced by technology (comparative research, digital library browsing)

Outline

Goldsmiths
UNIVERSITY OF LONDON

Humanities Computational Discovery

- Computational linguistics has been a core part of the AI programme from the beginning
- Results include natural language processing, computer modelling of grammars
- However, little that emulates the common formulations of humanistic research

Musicological Computational Discovery

- Similarly for musicology
- Work carried out in groups such as ISMS gives us models of human music cognition, measures of melodic similarity, or tools to identify functional harmony
- But it can't give us a feminist reading of Beethoven 5
- Or even a Schenker graph (yet!)

Humanities Comparative Research

- Corpus linguistics has a long and quite successful history
- Originally a sub-discipline, now more fully accepted
- Image comparison in archaeology (and paleography)
- Social science based heavily on quantitative (and qualitative) research enhanced by technology

Musicological Comparative Research I

- A potentially very interesting new direction for musicology
- Or maybe not so new?
- Musicology, from its inception in the C19th as a discipline concerned with the reconstruction of the music of the past through techniques borrowed from philology, ...
- ... up until the 2nd World War, was very interested in comparison
- The works of the newly formed canon were analysed for their conformance to universal rules
- And music from other cultures was analysed for its conformance (or not) to those rules
- All this was carried out through zealous comparison

Goldsmiths
UNIVERSITY OF LONDON

Musicological Comparative Research II

- After the 2nd World War this work was seen as too nationalistic
- Ethnomusicology advocated the study of world musics only in the context of their cultures
- Cross-cultural comparison was not valid
- Music analysis sought interest in individual works
- Cross-work comparison was not valid
- Now new techniques are making comparative studies possible again
- Especially comparative studies of performance through analysis of recordings
- However, musicologists have not yet formulated a programme of musical scholarship which requires or takes advantage of large scale comparative techniques

Goldsmiths
UNIVERSITY OF LONDON

Humanities Digital Libraries

- One area in which the humanities have benefited greatly from technology
- Much humanities scholarship is concerned with engaging at a very detailed level with literature
- Easy access to a good proportion of existing literature is now possible
- Projects like the Text Encoding Initiative enhance the semantic richness of this literature

Musicology's Digital Libraries

- For musicologists, a lot of scholarship is literature based
- So the same advantages apply
- However, access to musical materials (scores and recordings) has been slower to become available
- This is where techniques and efforts to produce digital editions of music are key
- Similarly, projects concerned with describing available music resources (publishing *metadata*)

Investigation Outline

- Identify research questions
- Identify information requirements
- Acquisition
- Analysis
- Evaluation
- Presentation

Identify Research Questions

- Includes both traditional research questions answered using new techniques (*enhancement*)
- And new research questions made possible by new techniques (*enabling*)
- Corpus musicology provides a good example
- As suggested earlier, formulated good research questions in the context of technologised musicology still provides fertile ground

Identify Information Requirements

- What information would you need to provide evidence to (dis)prove your hypothesis?
- How to represent and encode the information needed to answer your question?
- Collaboration; working with experts in different fields to generate research data and analytical methods
- Management of acquired data

Digression: Music from the Computer's Point of View I

- Syntax/semantics divide
- Computers can only manipulate syntax, they don't *understand* what they're dealing with
- Adding semantics to their results is up to the user
- Features: segments with semantics
- So syntactic segments can be extracted from musical data using some well-define technique
- And a semantic interpretation allows them to be interpreted as features of the music
- What do features tell us about music?
- How can we use them to answer questions about music?

Goldsmiths
UNIVERSITY OF LONDON

Digression: Music from the Computer's Point of View II

- The problem of tacit knowledge
- An experienced music analyst comes to a given music example with a wealth of prior experience of the musical genre
- Computers have no such prior experience
- *Except* when they are presented with a corpus; providing a context for the music example

Digression: Music from the Computer's Point of View III

- Music as information exists in different domains:
- 'Auditory', 'acoustic', 'graphemic' (after Babbit).
- Plus musical knowledge contained in texts, images, etc.
- How far can these domains be made available to the computer?

Acquisition

- Using digital music libraries (*metadata* again)
- Access to digital editions of music
- Optical music recognition
- Automatic transcription
- Motion capture of performers
- Experimentation; gathering evidence from subjects in listening tests, for example
- Surveys

Analysis

- Analysis, searching, visualisation of musical features
- Statistical analysis of surveys
- Algorithms for performing music theoretic operations on musical data

Evaluation

- Appropriate use of statistics
- Large amounts of data necessitate summarisation; patterns and trends
- It's important that those trends are credible
- Choosing helpful and unambiguous visualisations
- Applicability of findings to original research questions

Presentation

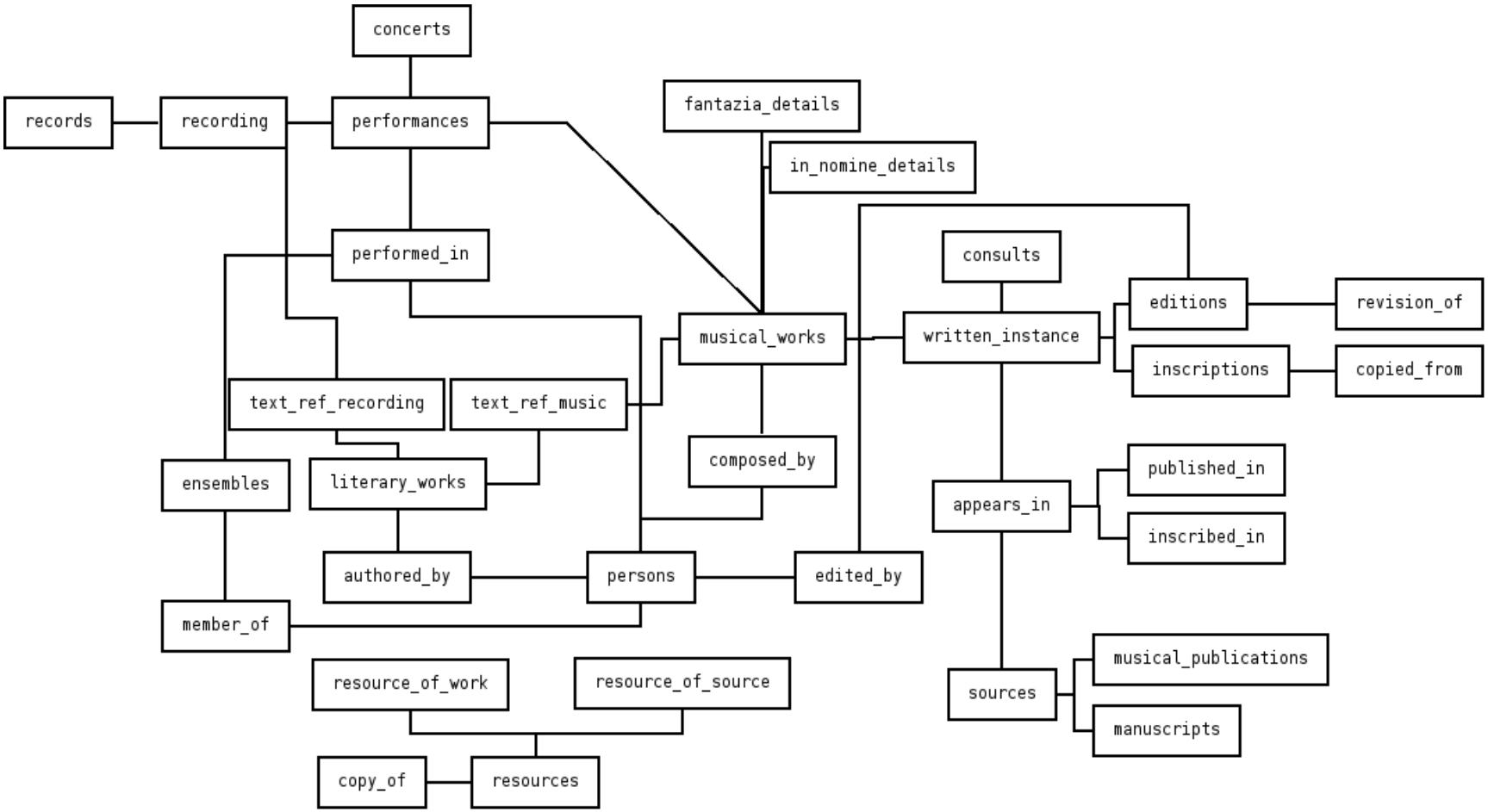
- Communicating results to a musicological audience
- Justifying your choice of statistically derived trends
- Acknowledging potential shortcomings of your method, data, analysis, etc.
- Integration of results into (public) datasets

A Mutli-domain Dataset

- The Purcell Plus project at Goldsmiths aims to apply some e-Science principles to the study of a small corpus
- The works are the *Fantazias* and *In Nomines* of Henry Purcell
- The dataset includes numerous audio recordings of each work
- Digital editions of each work
- And textual commentaries from various sources including concert programmes, concert reviews, record sleeves, record reviews, and analytical literature
- This dataset gives us the chance to apply some comparative research techniques
- The musicologist has designed the database schema for holding all the metadata for these representations

Goldsmiths
UNIVERSITY OF LONDON

Schema



database entity relationships

Purce

Goldsmiths
UNIVERSITY OF LONDON

Cross Domain Linking

- Another technological requirement we are investigating is how best to encode relationships between objects in different domains
- For example, between a concert view and the work performed, and the recording of the performance
- Or between an analytical text with specific or (more difficult) vague references to parts of a work
- We are hoping to find some standards-bases solution such that our data is more easily usable for further analysis by others

Text Analysis

- We hope to attempt a computer analysis of the literature with the aim of being to assert links to works automatically
- And to be able to analyse the kinds of vocabularies that different communities of musical commentators employ
- Examples of the computational discovery mode of knowledge creation

Disciplinary Impact

- Finally, the project hopes to be able to assess the impact e-Science may have on musicology
- This workshop is part of that process!
- Along with my Ph.D thesis

References I

Nicolas Cook.
The Compleat Musicologist?
ISMIR, 2005, 2005.

H. De Jong et al.
The computer revolution in science: steps towards the realization of computer-supported discovery environments
Artificial Intelligence 91:2 1997.

Dave De Roure et al.
The Semantic Grid: a future e-Science infrastructure
In F. Berman et al. (eds) *Grid Computing. Making the Global Infrastructure a Reality* 2003.

Goldsmiths
UNIVERSITY OF LONDON

References II

Paul Wouters.
What is the matter with e-Science? – thinking aloud about informatisation in knowledge creation.
The Pantaneto Forum, 23 July 2006.

Goldsmiths
UNIVERSITY OF LONDON