With the mini-project [*Medieval Music, Big Data and the Research Blend*](http://www.southampton.ac.uk/music/news/2014/04/11_medieval_music_big_data_and_research_blend.page) we attempt to address the question of the function of the conductus, a corpus of almost 900 thirteenth-century Latin poems variably set to monophonic and polyphonic music. As a matter of fact, this repertory does not seem to have a clear place in the medieval liturgy and, although the themes are mostly devotional, the texts set to music cover a wide range of topics. The known manuscript sources of the conductus (i.e. organised collections of music and poetry) do not provide much information about the significance and scope of the genre.

Yet, recent work by team members of the [*Cantum pulcriorem invenire*](http://www.southampton.ac.uk/music/research/projects/cantum_pulcriorem_invenire.page)research project has detected the presence of conductus text within rather unconventional sources: the poem ‘Naturas Deus regulis’, for instance, is mentioned in the twelfth-century chronicles of the Benedictine abbey in Abingdon, as part of a description of the miraculous expulsion of the Danes from the abbey’s monastic refectory in the late 860s. This accidental discovery was made by manual searches of text portions on the World Wide Web and prompted a series of research questions: how many other conducti appear in unconventional contexts? How does this contribute to our present understanding of the function of the conductus? And most importantly, can we develop a digital tool that, based on the digital edition of the 900 poems available on the *[Cantum pulcriorem invenire](http://catalogue.conductus.ac.uk/" \l "m-columnbrowser@||m-informationcontrol@url=html/home.php)* database, would perform automated searches on the World Wide Web in essentially no time?

This is where Kieran White entered the picture. After a long talk in Southampton and extensive email exchange between Kieran, Mark, and me, Kieran developed a HTML report that did the job for us. In order to index and search Latin documents, Kieran depended essentially on Lucene and the universal stemmer Stempel. With these tools, he then tokenised and stemmed a JSON export of the whole conductus poetry collection. More than 65,000 search engine queries were generated. Each query corresponded to a trigram of stemmed terms in the Conductus and was composed of multiple morphological variations of the trigam. In this way documents containing various inflections of the terms were included in the list of returned results. For instance, the trigram ‘mundi+pro+salute’ from ‘Ad cultum tue laudis’ was associated with a query seeking all phrases that could be generated from the following three groups of terms:

mundi, mundo + pro + salute, salutis, salutem, salus

This enables the report to identify not only identical concordances but also possible variations of a given trigram, considerably limiting the possibility of missing out on relevant data. In order to filter out certain known URLs (such as, of course, the CPI database itself!), I provided Kieran with a blacklist of sites that were not to be included in the automated searches. Queries generated as described above were submitted to Bing and relevant documents were downloaded and indexed. Plain text versions of the documents were also saved in case the originals were removed from the Web.

The final HTML report provides full text editions of the 900 conductus poems, and all lines in each poem link to a matching list of results (textual excerpts that identify concordances are highlighted). Obviously, lines containing a single word were not considered sufficiently discriminatory. Therefore for these lines two phrases were generated, one where the single word line was appended to its preceding line and the other where it prefixed its succeeding line. These two phrases then comprised a single query.

The report was completed in early January, and since then I have been using it extensively in order to identify documents displaying concordances comparable to the Abingdon Chronicle case. So far, I have been through almost 600 poems and have already detected some interesting concordances. In particular, the poem ‘Aristippe quamvis sero’ is quoted, and labelled ‘cantilena’ (i.e. ‘song’), within the twelfth-century Anglo-Norman ‘[Glossae in Sidonium](http://www.archive.org/stream/medievallatinlyr00alleiala/medievallatinlyr00alleiala_djvu.txt)’; the conductus ‘Deus pacis et dilectionis’ has revealed textual concordances with an ‘[oratio post cibum](http://en.wikipedia.org/wiki/Jesus_College,_Cambridge)’ (i.e. prayer after meal) that seems to have been recited at Jesus College, Cambridge, after formal dinners and is certainly older than the college itself, which was established in the fifteenth century. These kinds of discoveries are certainly promising, as they will allow us to address the question of the function and significance of the conductus repertory in its time; this would indeed be the final aim of the mini-project. I will complete the review of all 900 poems in the HTML report by mid-March, and Mark and I will then carry in-depth investigations of relevant cases comparable to those mentioned above; we are confident that our work will broaden the present understanding of the conductus.

Dr Gregorio Bevilacqua