# Big Musicology: A Framework for Transformation

Anna Kent-Muller
University of Southampton
UK
alkm1g12@soton.ac.uk

## ABSTRACT

Musicology is in need of a transformation. The field still largely conforms to the humanities lone-scholar stereotype, examining manuscripts and treatises with a fundamental lack of digital adoption. By considering biology's human genome project, I propose that musicology requires a change in its ethos – moving to a collaborative culture. These changes will provide the required conditions to execute a 'big musicology' project. In musicology, this project should be the digitisation of music and its associated documents. In turn, this will enable musicology to see digital technology as a tool, and not an exiled research group.

## CCS CONCEPTS

- Applied Computing → Arts and Humanities

## KEYWORDS

Big Musicology, Big Science, Transformation, Interdisciplinary, Collaboration, Digital Musicology

## 1 The Human Genome Project

Prior to the human genome project (HGP), the field of biology was highly observational, as exemplified by Charles Darwin's theory of evolution. In 1985, a workshop organised by Robert Sinsheimer debated the merits of sequencing the human genome. Concern was raised over the project's concept of 'big science' (a large-scale collaborative project, formed of an interdisciplinary team, collectively working towards one aim). The main concern was that the project would be a waste of time and money (to the detriment of 'real science'), due to the project not being a hypothesis-driven enterprise [1-3]. Some researchers have suggested that although there is no explicit hypothesis, this type of research has a broad implicit hypothesis: that the generation and assessment of data will develop models that identify new knowledge. Importantly, this type of research enables the examination of large data sets, thus providing specific leads for further study and massively increasing our knowledge of the world, instead of proving or refuting existing notions [4-6].

The project has enabled incredibly impactful research that could not have been predicted prior to the sequencing, and has transformed our understanding of evolutionary biology and medicine [7]. In addition to the huge medical advances made by this project, it has heightened awareness of alternative international strategies to research [3]. The field of biology now embraces 'big science' research, empowering researchers to build on tools and knowledge to enable objective unforeseen research.

## 2 What can Musicology learn from HGP?

Many computational musicological research projects aim to 'transform' the field, however none have yet had the significance of HGP.[1] The success of HGP can be seen through the utilisation of a 'big science' approach, with the collaborative culture in the field of biology provided the necessary conditions for HGP to be successful. Musicology (like the humanities generally) has exhibited a reticence to this collaborative research ethos, often involving a 'lone researcher' working with little or no input from others, other than perhaps during the peer-review process, or dissemination at conferences. This lone-scholar ethos is reflected in the musicology projects funded by UK research councils since 2006: of the 272 research-council funded projects, 57% had only one investigator.[2] Furthermore, only 31% can be said to be 'interdisciplinary', as the remaining projects feature investigators from a single discipline. The average number of investigators per funded project has seen a gradual increase from 1.1 in 2006 to 2 in 2017 (with a high of 4 in 2014). But still, in 2017 nearly half of the projects had a single investigator. This results in highly specialised research that addresses niche (rather than collective) issues. In contrast, in biology more than 64% projects funded by UK research councils in 2017 have 2 or more investigators.

As a result of musicologists working in sub-disciplinary silos, there is no consensus on what the 'big' issues are. Consequently, effort and scarce funding are spread too thinly to have the transformative effect digital musicology projects repeatedly promise. By identifying a project of ubiquitous importance and executing it as a big musicology project, I suggest that we could see our desired transformation. In this regard, note that a collaborative culture is already in part found in computational



[1] There are 20 musicology projects from 2006-2017 funded by UK Research Councils that state they are going to 'transform' musicology; see the Research Councils UK 'Gateway to Research' at http://gtr.rcuk.ac.uk/.

[2] Statistics generated from a search of 'Gateway to Research' for the terms 'music' and 'musicology' http://gtr.rcuk.ac.uk/.

musicology projects, which alone account for 50% of the projects funded by UK research councils with more than 2 researchers.

However, to complete such a 'big musicology' project we not only need to see the transformation of musicology to collaborative research, but also the acceptance of computational and digital techniques as a democratised tool, ubiquitous to all fields of musicology. The field of biology welcomes the use of digital tools to further knowledge and develop sophisticated methods for new research, for example GenBank and UCSC Genome Browser. Biologists have learnt, as Hallan Steven's [8] discussed, that computational processes bring new tools and questions, which completely reshape the way biological research is done. All biological research groups regard the digital tool as a method utilizable by them [9], whereas in musicology the digital is routinely regarded as a sub-discipline: 'digital musicology'.

## 3 Music's Human Genome Project

All areas of musicology require access to musical data – thus I propose that musicology's equivalent of the HGP should be the creation of musical sources in machine-readable formats, enabling computational search and analysis at scale. The project should aim to develop a system for exploring and analyzing large scale music collections, in a variety of formats – score, audio, and metadata. This is similar to the aims of the Digital Music Lab.[3] This project has seen slow progress since its initial funding in 2014, perhaps because it is missing the HGP's large interdisciplinary cross-institutional team. Music's HGP should be based on the work of the Digital Music Lab but the project would be enhanced by a larger collaborative team, utilizing other existing datasets such as those of CHARM's 'Mazurka Project' and SALAMI. Existing databases curators should agree on a machine-readable format that will enable us to link and analyse across these large datasets.

There have been developments in the digitisation of music with sites such as IMSLP[4] providing free access to PDF scans of sheet music. However, due to the data not being machine-readable, the process of interacting with the documents is still the same as with non-digital copies. Influence should be taken from the Google books project, which uses Optical Character Recognition to encode data. Further investment in Optical Music Recognition would enable the automatic encoding of music in MEI or MusicXML formats. Encoding data in this way will allow further developments in the analysis of large musical data sets. Current investment in computational methods of score analysis (such as the work of Alan Marsden [10] and Jonathan Bragg et al. [11]) could then be transformed to enable analysis at scale. In contrast, audio files are already machine-readable formats; the work of the MIR community will enable the analysis of these files [12].

Unlike HGP, musicology has to contend with music copyright. In many jurisdictions special copyright regulations exist that enable the research and publication of derived data – data about data [13]. The Digital Music Lab allows users to run experiments without direct access to the audio file, enabling research on copyrighted data. Music's HGP could embrace this methodology.

## 4 CONCLUSIONS

By adopting 'big musicology' we could see the transformation of musicology. However, musicologists first need to collaborate to create the necessary conditions to enable a 'big musicology' project. In musicology this project should be the digitisation of musical documents, so as to enable computational analysis of musical data. This will enable research that is currently unforeseeable, the development of new technology to aid musicologists, along with the possible development of ontologies to enable objective discussions about music.

As David Meredith [14] writes, 'it is neither possible nor desirable to prevent technology from being used, not merely to assist with the tasks that can be done by humans, but to do entirely new things that were not possible without it' (p.5). Some researchers are potentially unaware of how technology could impact their research – a situation exacerbated by the existence of the sub-discipline of 'digital musicology'. If as musicologists we begin to see computational methods as a democratised research tool, musicologists in turn can then help shape the technology.

## REFERENCES


[1] Francis S. Collins, Michael Morgan and Aristides Patrinos. 2003. The Human Genome Project: Lessons from Large-Scale Biology. *Science* 300, 5617 (April, 2003), 286-290.

[2] José Esparaza and Tadataka Yamada. 2007. The discovery value of "Big Science". *The Journal of Experimental Medicine* 204, 4 (April 2007), 701-704. DOI: http://dx.doi.org/ 10.1084/jem.20070073

[3] Charles DeLisi. 1988. The Human Genome Project: The ambitious proposal to map and decipher the complete sequence of human DNA. *American Scientist* 76, 5 (Sept-Oct 1988), 488-493.

[4] Steven Wiley. 2008. Hypothesis-Free? No Such Thing. *The Scientist* (2008).

[5] Eric D. Green, James D. Watson and Francis S. Collins. 2015. Twenty-five years of big biology. *Nature* 526, 7571 (Oct, 2015), 29-31. DOI: http://dx.doi.org/10.1038/526029a

[6] Paul van Helden. 2013. Data-driven hypothesis. *EMBO reports,* 14, 2 (2013). DOI: http://dx.doi.org/10.1038/embor.2012.207

[7] Leroy Hood and Lee Rowen, 2013. The Human Genome Project: big science transforms biology and medicine. *Genome Medicine* 5, 79 (2013). DOI: http://dx.doi.org/10.1186/gm483

[8] Hallm Stevens. 2013. *Life out of sequence: a data-driven history of bioinformatics.* University of Chicago Press, Chicago.

[9] Florian Markowetz. 2017. All biology is computational biology. *PLoS Biol* 15, 3 (March, 2017), 1-4. DOI: https://doi.org/10.1371/journal.pbio.2002050

[10] Alan Marsden. 2010. Schenkerian Analysis by Computer: a proof of concept. *Journal of New Music Research* 39, 3 (Nov, 2010) 269-289. DOI: http://dx.doi.org/10.1080/09298215.2010.503898

[11] Jonathan Bragg, Elaine Chew and Stuart Shieber. 2011. Neo-Riemannian Cycle Detection with Weighted Finite-State Transducers. In *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR 2011).* 399-404.

[12] Stephen Downie. 2003. Music Information Retrieval. *Annual Review of Information Science and Technology,* 37, Chapter 7 (2003) 295-340. DOI: http://doi.org/10.1002/aris.1440370108

[13] Tilman Weyde et al. 2014. Big Data for Musicology. In *Proceedings of the 1st International Digital Libraries for Musicology Workshop (DLfM 2014).* 1-3.

[14] David Meredith. 2009. Introduction. *Modern Methods for Musicology: Prospects, Proposals, and Realities.* Routledge, Oxford. 1-6.


[3] http://dml.city.ac.uk/
[4] http://imslp.org/