Studies in History and Philosophy of Biological and Biomedical Sciences xxx (2015) 1–9

Contents lists available at ScienceDirect

Studies in History and Philosophy of Biological and Biomedical Sciences

journal homepage: www.elsevier.com/locate/shpsc

Documenting genomics: Applying archival theory to preserving the records of the Human Genome Project

Jennifer Shaw*

Wellcome Library, 215 Euston Road, London, NW1 2BE, UK

ARTICLE INFO

Article history:
Available online xxx

Keywords:
Human Genome Project
Archival theory
Scientific collecting

ABSTRACT

The Human Genome Archive Project (HGAP) aimed to preserve the documentary heritage of the UK's contribution to the Human Genome Project (HGP) by using archival theory to develop a suitable methodology for capturing the results of modern, collaborative science. After assessing past projects and different archival theories, the HGAP used an approach based on the theory of documentation strategy to try to capture the records of a scientific project that had an influence beyond the purely scientific sphere. The HGAP was an archival survey that ran for two years. It led to ninety scientists being contacted and has, so far, led to six collections being deposited in the Wellcome Library, with additional collections being deposited in other UK repositories. In applying documentation strategy the HGAP was attempting to move away from traditional archival approaches to science, which have generally focused on retired Nobel Prize winners. It has been partially successful in this aim, having managed to secure collections from people who are not 'big names', but who made an important contribution to the HGP. However, the attempt to redress the gender imbalance in scientific collections and to improve record-keeping in scientific organisations has continued to be difficult to achieve.

1. Introduction

The failure to preserve the records of scientific endeavour is intermittently bemoaned by scientists, historians and archivists. In some cases, this is followed by activity to try to remedy the situation. While American geneticist L. C. Dunn was writing his *Short History of Genetics* he became concerned about the loss of material due to the: "Failure of working scientists to preserve their correspondence and other unpublished papers and to bequeath them to an appropriate repository" (Dunn, 1965). Dunn led by example by depositing his papers at the American Philosophical Society and actively encouraged other scientists to do the same.[1]

Similar concerns to those expressed by Dunn can also be seen from the archival perspective, with the lack of training and interest shown by archivists in scientific material both raised as issues. In 1969 during his presidential address at the Society of Archivist's annual general meeting Roger H. Ellis stated that the records of science and technology: "Do not at present figure largely in our holdings, and up to now we have not been taught anything specific about them in our Diploma Courses" (Ellis, 1970, p. 87). Despite Ellis's hope that there was the prospect of change for the latter, the records of science and technology still do not feature as a significant aspect of any of the post-graduate training courses for archivists in the UK and Ireland. There are currently seven universities offering record-keeping courses accredited by the Archives and Records Association (UK and Ireland).[2] The majority of modules on these

* Tel.: +4420 7611 8253
E-mail address: Je.shaw@wellcome.ac.uk.

[1] L. C. Dunn Papers held by The American Philosophical Society: http://www.amphilsoc.org/guides/glass/dunn.htm. There is a Guide to the Genetics Collections at the American Philosophical Society: http://www.amphilsoc.org/guides/glass/ (both accessed 8 May 2015).

[2] At present accredited courses are offered by: Aberystwyth University; Northumbria University; University College Dublin; University of Dundee; University of Glasgow; University of Liverpool; and University College London. Archives and Records Association website: http://www.archives.org.uk/careers/careers-in-archives.html (accessed 8 May 2015).

http://dx.doi.org/10.1016/j.shpsc.2015.08.005
1369-8486/

courses focus on the theoretical basis of record-keeping or the key skills required in this area, such as managing digital records or palaeography. The need to equip archivists with the necessary skills for an evolving workplace, in particular the need to manage records in both analogue and digital formats, has seen a gradual reduction in the number of subject specific modules and a greater emphasis on the overarching principles that apply across the management of archives. Although now in the minority, a few subject specific modules remain on corporate records, ecclesiastical archives, education archives, house history and military archives. The records of science and technology still do not feature on the post-graduate courses and, unless there is a reversal of the current trend, a dedicated module is unlikely in the foreseeable future.

Against this backdrop of concern that not enough was being done by all sides to preserve the records of science and technology in general, in the late 2000s key organisations involved in the Human Genome Project (HGP) started to worry about these records specifically. In part this was prompted by the retirement of key personnel, but also the format of the records. The era in which the HGP happened meant that a large proportion of the primary material was created in born-digital format making it more vulnerable to loss than its analogue equivalents. In response to this concern, a meeting was held at Cold Spring Harbor Laboratory, New York in 2009 to discuss the importance of the HGP, raise awareness amongst scientists and ascertain what action needed to be taken to preserve its documentary heritage.[3] The initial idea was for a single, international project to locate and preserve records relating to the HGP, but it quickly became clear that the logistics and funding for this type of project were impractical. Instead, it was decided that national projects would be established to undertake the majority of the work in each country whilst maintaining close collaboration and communication with each other. The aspiration remains to unite the work of these national projects in the future. The UK project, known as the Human Genome Archive Project (HGAP), was launched by the Wellcome Library, funded by the Wellcome Trust, in January 2012 to preserve the documentary heritage of the HGP in the UK.[4] This paper will tell the story, from an archivist's perspective, of how archival theory was applied to modern collaborative science to develop and implement a suitable survey methodology by the UK project between January 2012 and December 2013.

2. Approaches to scientific collecting

The first question that the UK project faced was; where do you start when trying to survey contemporary, collaborative science? The most obvious starting point appeared to be a review of how scientific collecting had been approached in the past. This work mainly focused on the UK because the archival landscape can vary greatly from country to country, often influenced by legislation and funding mechanisms. However, some approaches from overseas were also investigated, but only in Anglophone literature.

There has been regular activity in the UK to preserve the records of individual scientists since the early 1960s. The majority of this work can trace its origins back to when Lord Evershed, chairman of the Royal Commission on Historical Manuscripts (HMC), asked Commissioner Roger Quick to form a plan for addressing what he saw as a neglected area in the UK's archive provision. This eventually led to the establishment of a pilot project organised by the Royal Society and the HMC in 1967, which surveyed three distinguished and recently deceased Fellows of the Royal Society, namely pharmacologist and physiologist Sir John Gaddum, physicist Sir Francis Simon and geologist Professor Lawrence Wager.[5] Their records were collected, sorted, listed and deposited with the Royal Society's library.[6] The project successfully demonstrated tangible results with limited resources. It also set a number of precedents that have been applied to scientific collecting in the UK ever since.

Prior to the pilot it had been imagined that a physical archive would be established for scientific collections, but as a result of the pilot this changed. Instead, the idea for a processing centre was proposed where collections would be collected and catalogued before a suitable permanent archival home could be found for them. The idea of operating a processing centre had many aspects in its favour, but a key benefit was that it would be much more economical to operate as it would lack the burden of long-term storage costs. As a result, in 1973 the Contemporary Scientific Archives Centre was established in Oxford. This was initially a three year trial with the Royal Society acting as the main financial supporter (Gowing, 1979). Over the years the processing centre evolved with changes in funding, staffing and location, but the method of working remained stable. In 1987 the Centre moved to Bath and became known as the National Cataloguing Unit for the Archives of Contemporary Scientists (NCUACS). The Unit continued to collect scientific papers, catalogue them and then find a suitable archival home for them until 2009 when the Royal Society withdrew its funding and the Unit was forced to close.[7] Following the closure of the NCUACS, the Centre for Scientific Archives was established as a processing centre for scientific collections based at the Science Museum's site in Wroughton, near Swindon.[8] Its work is broadly similar, but it does not have any core funding; instead it receives educational grants for the cataloguing of specific collections.[9]

There has certainly been a high level of continuity in the approach used to preserve scientific archives in the UK with the methodology developed as part of the pilot in the 1960s still being applied nearly fifty years later. This approach has undoubtedly yielded tangible and positive results. It also achieves these results in a highly economical way with relatively low resource requirements. Although it takes time to catalogue archive collections, by far the biggest costs are the long-term care and storage of these collections

[3] See the Cold Spring Harbor Laboratory annual report for 2009, p.368: http://www.cshl.edu/images/stories/about_us/annual_reports/363-400BanburyCenter_09annrep.pdf (accessed 8 May 2015).

[4] This paper focuses on the methodology used for the UK project. The other national projects have had to create their own methodologies in response to their own environments and challenges. For instance, the French project looks at contributions to the HGP from the start of the Centre d'Etude du Polymorphisme Humain's (Center for the Study of Human Polymorphisms) research on the human genome; the Japanese project looks at where Japan made significant contributions including work on individual chromosomes and the development of new technology; and the US project involves initiatives led by Cold Spring Harbor Laboratory and by the National Human Genome Research Institute. In November 2014 an initial meeting to establish a Chinese project was held.

[5] For a more detailed account of the Royal Society/HMC pilot project see Ellis (1970).

[6] Sir John Gaddum (reference JHG) and Sir Francis Simon collection (reference FS) both held by the Royal Society archives: http://www.aim25.ac.uk/cgi-bin/vcdf/detail?coll_id=5976&inst_id=18&nv1=browse&nv2=person (for Gaddum); http://www.aim25.ac.uk/cgi-bin/vcdf/detail?coll_id=5973&inst_id=18&nv1=search&nv2=person (for Simon). The Lawrence Wager collection was subsequently returned to the family before being deposited again. It is now held by Oxford University, Natural History Museum (reference **NCUACS 84.5.99):** http://discovery.nationalarchives.gov.uk/details/rd/7444c6ae-f937-4c12-b18a-196a1f827d04 (all accessed 8 May 2015).

[7] NCUACS website preserved by the UK Web Archive: http://www.webarchive.org.uk/wayback/archive/20100204232541/http://www.bath.ac.uk/ncuacs/ (accessed 8 May 2015).

[8] Centre for Scientific Archives website: https://sites.google.com/site/centreforscientificarchives/(accessed 8 May 2015).

[9] The Charity Commission holds the 2011 annual report for the Centre for Scientific Archives: http://apps.charitycommission.gov.uk/Accounts/Ends03/0001138303_ac_20111130_e_c.pdf (accessed 8 May 2015).

Please cite this article in press as: Shaw, J., Documenting genomics: Applying archival theory to preserving the records of the Human Genome Project, Studies in History and Philosophy of Biological and Biomedical Sciences (2015), http://dx.doi.org/10.1016/j.shpsc.2015.08.005