The British Journal for the History of Science (2021), **54**, 213–231
doi:10.1017/S0007087421000224

RESEARCH ARTICLE

Genes go digital: *Mendelian Inheritance in Man* and the genealogy of electronic publishing in biomedicine

Michael F. McGovern*

Program in History of Science, Department of History, Princeton University
*Corresponding author: Michael F. McGovern, Email: mmcgovern@princeton.edu

Abstract

Mendelian Inheritance in Man (*MIM*), a computerized catalogue of human genetic disorders authored and maintained by cardiologist and medical genetics pioneer Victor A. McKusick, played a major part in demarcating between a novel biomedical science and the eugenic projects of racial betterment which existed prior to its emergence. Nonetheless, it built upon prior efforts to systematize genetic knowledge tied to individuals and institutions invested in eugenics. By unpacking the process of digitizing a homespun cataloguing project and charting its development into an online database, this article aims to illuminate how the institution-building efforts of one individual created an 'information order' for accessing genetic information that tacitly shaped the norms and priorities of the field toward the pursuit of specific genes associated with discernible genetic disorders. This was not by design, but rather arose through negotiation with the catalogue's users; it accommodated further changes as biomedical research displaced the Mendelian paradigm. While great effort was expended toward making sequence data available to investigators during the Human Genome Project, *MIM* was largely taken for granted as a 'legacy system', McKusick's own labour of love. Drawing on recent histories of biomedical data, the article suggests that the bibliographical work of curation and translation is a central feature of value production in the life sciences meriting attention in its own right.

'No other organism is so thoroughly and extensively phenotyped, and the result is bewildering', wrote cancer biologist Alfred G. Knudson in 1967. One might presume that Knudson was referring to the fruit fly, or perhaps the mouse, canonical model organisms with a wealth of genetic markers derived from what historian Bruno Strasser terms 'live museums'.[1] But he was actually talking about *Homo sapiens*, and reviewing *Mendelian Inheritance in Man* (*MIM*), a catalogue of the existing literature on clinically observed

[1] Alfred Knudson, review of *Mendelian Inheritance in Man*, by Victor A. McKusick, *Quarterly Review of Biology* (1 September 1967) 42(3), p. 426. Knudson's characterization was somewhat hyperbolic, as the number of known *Drosophila* genes was by a number of counts greater, even into the 1970s, and according to similar catalogues of the time. Theodosius Dobzhansky, review of *Genetic Variations of Drosophila melanogaster*, by Dan L. Lindsley and E.H. Grell, *Science* (29 November 1968) 162(3857), p. 993; Bentley Glass, review of *Mendelian Inheritance in Man*, by Victor A. McKusick, *Quarterly Review of Biology* (1 December 1975) 50(4), pp. 456–7. On the relationship between experimental biomedicine and natural history, with particular attention to data practices, see Bruno J. Strasser, *Collecting Experiments: Making Big Data Biology*, Chicago: The University of Chicago Press, 2019. For older, canonical accounts of the laboratory 'breeder reactor', standardization and the rise of model organisms see Robert E. Kohler, *Lords of the Fly: Drosophila Genetics and the Experimental Life*, Chicago: The University of Chicago Press, 1994; Karen A. Rader, *Making Mice: Standardizing Animals for American Biomedical Research, 1900–1955*, Princeton, NJ: Princeton University Press, 2004.

genetic disorders compiled by Johns Hopkins cardiologist Victor A. McKusick. Knudson's statement is perhaps less surprising upon further inspection. After all, historians have demonstrated how eugenics research and state-administered policies refined the information practices central to the rise of modern bureaucracies.[2] *MIM* was a different enterprise, an index in the most literal sense, but also figuratively. It pointed to a transition under way: the rebranding of genetics as a medical speciality.[3] Though Knudson saw no need to comment on this point, its two closest antecedents had emerged from the British Eugenics Society and Nazi racial-hygiene projects. *MIM* channelled the modest aspiration of helping laboratory researchers, physicians and genetic counsellors alike to demarcate their science from white-supremacist projects of racial betterment.

Likened to an encyclopedia, even a bible, *MIM* was a bibliographic initiative that organized biomedical abstracts under specific genes – a printed card catalogue maintained actively by a leading authority in the field.[4] Over the course of its twelve print editions from 1966 to 1998 and eventual move online, *MIM* has grown from 1,486 to more than 25,000 entries and established a widely used classification system based on numerical IDs.[5] McKusick was the sole attributed author for the majority of its editions, and *MIM* was central to nearly every one of his many institution-building efforts: from a well-attended medical genetics summer school at Bar Harbor, Maine, to his clinical fellowship programme at Johns Hopkins, to the pivotal gene-mapping workshops he co-sponsored at Yale that formed the basis for the Human Genome Project (HGP).[6] Consequently, it became the standard reference guide to human genes, helping non-geneticists and a growing profession of genetic counsellors follow the leading edge of a field in rapid advance.[7]

In light of growing scholarship on the relationship between genetics and computing, scientific publishing and information practices in the sciences more broadly, *MIM* warrants a history of its own. It shaped how countless investigators encountered and made

[2] See, inter alia, Garland E. Allen, 'The Eugenics Record Office at Cold Spring Harbor, 1910–1940: an essay in institutional history', *Osiris* (1986) 2, pp. 225–64; Alexandra M. Stern, *Eugenic Nation: Faults and Frontiers of Better Breeding in Modern America*, Berkeley: University of California Press, 2005; Phillip Thurtle, *The Emergence of Genetic Rationality: Space, Time, & Information in American Biological Science, 1870–1920*, Seattle: University of Washington Press, 2007; Theodore M. Porter, *Genetics in the Madhouse: The Unknown History of Human Heredity*, Princeton, NJ: Princeton University Press, 2018.

[3] On the incomplete transition from eugenics to medical genetics see Nathaniel Comfort, *The Science of Human Perfection: How Genes Became the Heart of American Medicine*, New Haven, CT: Yale University Press, 2012; Jenny Bangham, *Blood Relations: Transfusion and the Making of Human Genetics*, Chicago: The University of Chicago Press, 2020.

[4] On the long history of bibliographic indices in the sciences see Alex Csiszar, 'How lives became lists and scientific papers became data: cataloguing authorship during the nineteenth century', *BJHS* (March 2017) 50(1), pp. 23–60. On card catalogues as a distinctive technology of modernity underlying the rise of modern computing see Markus Krajewski, *Paper Machines: About Cards & Catalogs, 1548–1929* (tr. Peter Krapp), Cambridge, MA: MIT Press, 2011.

[5] *OMIM: Online Mendelian Inheritance in Man*, at www.omim.org, accessed 3 March 2021. For accounts of MIM's history and present challenges by McKusick and the subsequent team responsible for maintaining it see Victor A. McKusick, '*Mendelian Inheritance in Man* and its online version, OMIM', *American Journal of Human Genetics* (April 2007) 80(4), pp. 588–604; Joanna Amberger, Carol Bocchini and Ada Hamosh, 'A new face and new challenges for Online Mendelian Inheritance in Man (OMIM®)', *Human Mutation* (1 May 2011) 32(5), pp. 564–7.

[6] On McKusick's institution building see M. Susan Lindee, *Moments of Truth in Genetic Medicine*, Baltimore: Johns Hopkins University Press, 2005; Andrew J. Hogan, *Life Histories of Genetic Disease: Patterns and Prevention in Postwar Medical Genetics*, Baltimore: Johns Hopkins University Press, 2016; Soraya de Chadarevian, *Heredity under the Microscope: Chromosomes and the Study of the Human Genome*, Chicago: The University of Chicago Press, 2020.

[7] On genetic counselling see Karen-Sue Taussig, *Ordinary Genomes: Science, Citizenship, and Genetic Identities*, Durham, NC: Duke University Press, 2009; Alexandra M. Stern, *Telling Genes: The Story of Genetic Counseling in America*, Baltimore: Johns Hopkins University Press, 2012.

https://doi.org/10.1017/S0007087421000224 Published online by Cambridge University Press

sense of genetics, whether on paper or online.[8] For these very reasons, *MIM*'s contribution is often cited, although rarely scrutinized.[9] This article sheds new light on the intersection of genetics and computing by tracing the origins, digitization, reception and eventual online distribution of this catalogue. Following Simon Schaffer, I employ C.A. Bayly's notion of an 'information order', in which formal and informal knowledge operate together within a particular social formation, to explore *MIM*'s role in the informational economy of modern genetics.[10]

To be sure, a number of scholars have explored the convergence of genetics with information technology, but most are primarily concerned with the manipulation and circulation of sequences: strings of nucleotides and amino acids represented by a cipher of letters.[11] Recent studies argue that preoccupation with the vision of an immutable genetic text effaces the bumpy cellular cartography forged in labs and clinics throughout the post-war period with the goal of refining the diagnosis and pathophysiology of well-known human hereditary disorders.[12] Whether in government-sponsored workshops or samizdat newsletters, these investigators developed their own strategies for sharing information to coordinate a research programme that formed the basis of the HGP.[13] Focusing on sequencing at the expense of these developments risks making only the most visible efforts to accommodate information technology to genetics appear as

[8] For a broad, historiographical consideration of handbooks in the sciences, describing a conference in which this paper was first presented, see Angela N.H. Creager, Mathias Grote and Elaine Leong, 'Learning by the book: manuals and handbooks in the history of science', *BJHS Themes* (2020) 5, pp. 1–13. On catalogues and the political economy of biomedicine in particular, see Mathias Grote, 'Total knowledge? Encyclopedic handbooks in the twentieth-century chemical and life sciences', *BJHS Themes* (2020) 5, pp. 187–203.

[9] Philosopher of biology Rachel Ankeny has analysed how *MIM* was an early promoter of 'geneticization', or the tendency to reduce observable disease entities to isolated genetic components. Rachel A. Ankeny, 'Geneticization in MIM/OMIM®? Exploring historic and epistemic drivers of contemporary understandings of genetic disease', *Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine* (August 2017) 42(4), pp. 367–84. My account, based on extensive work in McKusick's archives, focuses on the information technology behind *MIM* and situates it within a broader discussion about publication and economies of credit in biomedicine.

[10] C.A. Bayly, *Empire and Information: Intelligence Gathering and Social Communication in India, 1780–1870*, Cambridge: Cambridge University Press, 1996, p. 3; Simon Schaffer, 'Newton on the beach: the information order of *Principia Mathematica*', *History of Science* (1 September 2009) 47(3), pp. 243–76.

[11] Lily Kay's history of molecular biology and Cold War information discourse has provided much fodder for scholarly engagement: Lily E. Kay, *Who Wrote the Book of Life: A History of the Genetic Code*, Stanford, CA: Stanford University Press, 2000. Further scholarship has complicated this account by considering the relationship between computerization and the databasing and archiving efforts that preceded it: Timothy Lenoir, 'Shaping biomedicine as an information science', in Mary Ellen Bowden, Trudi Bellardo Hahn and Robert Virgil Williams (eds.), *Proceedings of the 1998 Conference on the History and Heritage of Science Information Systems*, Medford, NJ: American Society for Information Science and the Chemical Heritage Foundation, 1999, pp. 27–45; Miguel García-Sancho, *Biology, Computing, and the History of Molecular Sequencing: From Proteins to DNA, 1945–2000*, Basingstoke: Palgrave Macmillan, 2012; Joseph November, *Biomedical Computing: Digitizing Life in the United States* (Baltimore: Johns Hopkins University Press, 2012); Hallam Stevens, *Life out of Sequence: A Data-Driven History of Bioinformatics*, Chicago: The University of Chicago Press, 2013; Strasser, op. cit. (1). For a current perspective on the history of genetic-information-sharing policies and technologies see Robert Cook-Deegan, Rachel A. Ankeny and Kathryn Maxson Jones, 'Sharing data to build a medical information commons: from Bermuda to the Global Alliance', *Annual Review of Genomics and Human Genetics* (31 August 2017) 18(1), pp. 389–415.

[12] On the importance of cells, chromosomes and other biological substrates see María Jesús Santesmases and Edna Suárez-Díaz, 'A cell-based epistemology: human genetics in the era of biomedicine', *Historical Studies in the Natural Sciences* (1 February 2015) 45(1), pp. 1–13; de Chadarevian, op. cit. (6); Mathias Grote *et al.*, 'The molecular vista: current perspectives on molecules and life in the twentieth century', *History and Philosophy of the Life Sciences* (4 February 2021) 43(1), p. 16 (preprint).

[13] On informal networks of information sharing see Christopher M. Kelty, 'This is not an article: model organism newsletters and the question of "open science"', *BioSocieties* (June 2012) 7(2), pp. 140–68.

https://doi.org/10.1017/S0007087421000224 Published online by Cambridge University Press

the motive force of the entire enterprise. Pursuing an alternative view of the genetic-information revolution from the perspective of a key bibliographic resource – setting aside the aims and intentions of genome boosters who have shaped what we know about genetic databases – highlights a different set of technical and human constraints than the scalar concerns of making sequences available.[14]

In the post-war political economy of biomedical research, clinicians struggled to keep up to date while competing amongst themselves for grants. *MIM* helped them comprehend the disorderly production of genetic knowledge but in a way that was not as systematic as it appears at first glance. Sceptical of so-called 'information overload' as a uniquely modern predicament, historians of science interested in data practices more broadly have interrogated the relationship between technology, scholarship and institutional norms.[15] *MIM* seems to fit nicely into Jon Agar's schema for early computerization projects, 'attempted [only] in settings where there already existed material and theoretical computational practices and technologies'.[16] That McKusick's own hard-fought access to computer time for parallel projects made the enterprise possible lends credence to Agar's account.

However, *MIM* provides a different portrait of what was at stake in leveraging computer power. Mathias Grote calls the encyclopedic handbook a 'paradoxical medium of scientific modernity': relied upon for comprehensiveness, yet always (already) out of date.[17] McKusick had a provisional solution to this problem. The impressive amount of information gathered in *MIM* was a palimpsest rather than a fully fledged systematization project, coming from particular sources like abstract digests, and sometimes including personal communications and observations. Its information order not only helped centre the gene as an object of biomedical scrutiny, but also reinforced a particular temporality of knowledge. Time is money, so they say, and Alex Csiszar argues that catalogues should be understood as 'technologies of valuation'; rather than merely extending access as unbranded containers of the latest findings, they promote specific visions of who gets to produce knowledge, and get repurposed in unintended ways.[18] Though edited by a dedicated staff, *MIM*'s successor, *Online Mendelian Inheritance in Man* (*OMIM*), bears traces of McKusick's own process and priorities – a 'legacy system' that remains integral to the organization of genetic knowledge.[19] I suggest that *MIM* shaped the norms of publication around individual genes before the sequencing revolution, but only after abandoning a focus on the individual author as the basis for organization, as well as the very ontology of Mendelian inheritance upon which its underlying technology was based.

This article begins by situating *MIM* within a genealogy of genetic compendia before recounting how McKusick came to the project as director of Hopkins's genetics clinic.

[14] On speed, access and the politics of genomic information see Michael Fortun, 'Projecting speed genomics', in Michael Fortun and Everett Mendelsohn (eds.), *The Practices of Human Genetics*, Dordrecht: Springer, 1999, pp. 25–48; Stephen Hilgartner, *Reordering Life: Knowledge and Control in the Genomics Revolution*, Cambridge, MA: MIT Press, 2017.

[15] On the question of 'information overload' see Ann Blair, *Too Much to Know: Managing Scholarly Information before the Modern Age*, New Haven, CT: Yale University Press, 2010; Nick Levine, 'The nature of the glut: information overload in postwar America', *History of the Human Sciences* (1 February 2017) 30(1), pp. 32–49. For many nuanced discussions of data practices in the history of science see Elena Aronova, Christine von Oertzen and David Sepkoski (eds.), *Data Histories*, *Osiris*, 2nd series (2017) 32.

[16] Jon Agar, 'What difference did computers make?', *Social Studies of Science* (1 December 2006) 36(6), pp. 869–907, 872.

[17] Grote, op. cit. (8), p. 202.

[18] Csiszar, op. cit. (4), p. 25.

[19] By way of demonstration, Rachel Ankeny notes that almost 70 per cent of *MIM* and *OMIM* can be attributed to McKusick, though without noting that he compiled abstracts directly from specific sources like the Institute for Scientific Information's *Current Contents*, discussed below. Ankeny, op. cit. (9), p. 374.

https://doi.org/10.1017/S0007087421000224 Published online by Cambridge University Press

It then turns to consider how the primary record locator of the *MIM* software shifted from the author to the allele. Summarizing changes over the years within a representative set of disorders, the article highlights broader shifts in *MIM*'s bibliographic information order throughout the rise of genetic sequencing. It concludes by sketching the development of *OMIM* in the context of national funding for medical-library infrastructure and the rise of information-sharing resources through the HGP.

Eugenics and the cataloguing imperative

McKusick did not hesitate to call *MIM* 'an encyclopedia of genes', invoking the Enlightenment project of Diderot and D'Alembert: 'a barometer of knowledge in the making, to be constantly revised as experience and knowledge were enlarged and refined'.[20] He also drew parallels to the *Oxford English Dictionary*, since *MIM* was compiled using 'a historical or diachronic approach rather than by the descriptive or synchronic method used by most encyclopaedias and textbooks'. Yet he was not the first to catalogue human genes, and such rhetorical gestures conceal the work's antecedents, discussed in this section.[21] In the *Treasury of Human Inheritance* (1912, 1948), Julia Bell compiled data in order to argue for distinct patterns of genetic inheritance. Otmar Freiherr von Verschuer's *Genetik des Menschen* (1959) was more of a catalogue but organized its information along the lines of medical speciality and remained committed to a normal/pathological distinction that McKusick would eschew. *MIM* addressed itself to medical geneticists as an impartial literature review that presented information at the level of individual genes, though it inherited aspects of both.

The Galton Laboratory for National Eugenics was established when the biology of heredity was still in flux. Founding documents defined the institution primarily as 'a storehouse for statistical material bearing on the mental and physical conditions in man and the relation of these conditions to inheritance and environment'.[22] For a time, adherents of biometry, a discipline pioneered by Charles Darwin's cousin Francis Galton, carried the flag of eugenics, and such statistical material was amenable to their needs. Biometricians worked to derive laws of inheritance from observable, continuous human characteristics like intelligence and stature, analysing their distribution across populations and within families using mathematical techniques.[23] In a well-studied episode, Cambridge biologist William Bateson took the biometrical conception of heredity – championed by Karl Pearson and W.F.R. Weldon, his former teacher – to task after successfully replicating the work of Gregor Mendel, becoming Mendel's leading English proponent and naming his science 'genetics'.[24] Mendelians would eventually lay claim to the soil of the new

[20] Victor A. McKusick, *Mendelian Inheritance in Man: Catalogs of Autosomal Dominant, Autosomal Recessive, and X-Linked Phenotypes*, 10th edn, Baltimore: Johns Hopkins University Press, 1992, p. xxxiv.

[21] McKusick directly compared his own enumeration to Verschuer's in 1962 while referring to Bell's work only in passing; human geneticist and historian Peter Harper has considered Bell's influence in the cataloging enterprise apart from her own work on linkage analysis. Victor A. McKusick, 'On the X chromosome of man', *Quarterly Review of Biology* (1 June 1962) 37(2), pp. 69–175; Peter S. Harper, 'Julia Bell and the *Treasury of Human Inheritance*', *Human Genetics* (1 April 2005) 116(5), pp. 422–32.

[22] '[Laboratory Founding Document]', *c.*1904, UCL Special Collections, Galton Laboratory/1/1, Business Papers, Folder 2 of 2, p. 2.

[23] Peter S. Harper, *A Short History of Medical Genetics*, Oxford: Oxford University Press, 2008, pp. 65–9.

[24] Historian Daniel Kevles calls this well-worn conflict study 'one of the most vitriolic disputes in the history of science', though later analyses have been more measured. Daniel J. Kevles, *In the Name of Eugenics: Genetics and the Uses of Human Heredity*, 2nd edn, Cambridge, MA: Harvard University Press, 1997, p. 44. Porter argues that Pearson the positivist never rejected Mendelism *a priori* and largely saw it as compatible, while accounts by Harper and Müller-Wille and Rheinberger both argue that these views were never necessarily incompatible, nor did Mendelism ever fully supplant biometrical methods. Theodore M. Porter, *Karl Pearson: The Scientific*

https://doi.org/10.1017/S0007087421000224 Published online by Cambridge University Press

field, as well as the culpability for promoting eugenic legislation, both in Britain and the United States.[25]

The key distinction at stake between Mendelism and biometry was whether traits were inherited in a continuous or a discrete manner. This could be difficult to sort out in practice, as conspicuous traits like eye colour appeared to result from blending both parents' contributions. Before the role of the X chromosome in sex determination was sorted out, this issue also caused trouble for the Mendelian camp, which struggled to explain how the human sex ratio could be maintained if some sex traits were dominant, manifesting more commonly, and others recessive and merely carried. As research on generations of experimentally bred fruit flies and pea plants resolved some of these questions, another lasting contribution to heredity research came from tabulating clinical observations.

While the Galton Laboratory produced a number of signature publications – the journals *Biometrika* and the *Annals of Eugenics*, for example – one of its most enduring documents was the meticulous *Treasury of Human Inheritance*, a project begun in 1909 under Pearson's directorship. As geneticist-cum-historian Peter Harper has argued, many of the early issues, later collected as volumes, were more compilations of data than analyses: pedigrees and photographs of congenital abnormalities aimed toward standardizing eugenics in the clinic.[26] The demands of the First World War made it difficult for Pearson to keep up with the work, and in the 1920s physician Julia Bell assumed the mantle of the *Treasury*. With support from Britain's Medical Research Council and Pearson's successor, R.A. Fisher, whose wartime work on blood groups laid the groundwork for more robust study of human heredity, Bell's assiduous efforts turned the endeavour decisively toward a Mendelian approach to inheritance.[27]

The section 'Dystrophia myotonica and allied diseases' best captures Bell's approach and doubles as an overview of the disorder I discuss below. A condition originally described in 1830 by neurologist Charles Bell (no relation), myotonic dystrophy is characterized by a lack of control over facial musculature and, as of her writing up in the 1930s and 1940s, had been classified into five putatively distinctive subtypes.[28] Bell provided a detailed historical overview of the various grouped disorders as well as clinical notes furnished by a neurologist. She analysed 223 pedigrees culled from the literature or contributed directly by neurologist Otto Maas, all of which are published in detail and annotated following an extensive bibliography. Throughout numerous tables that tabulate and collate the pedigree data, Bell singled out Maas's data because he treated two of the conditions, Thomsen's disease and the more prominent dystrophia myotonica, as the same disorder. She apologized that her 'conclusions and mode of procedure sometimes diverge conspicuously from his expressed and authoritative views', but insisted on distinguishing between the two on the basis of her own observation of patients at the London Hospital.[29]

Life in a Statistical Age, Princeton, NJ: Princeton University Press, 2004, p. 269; Harper, op. cit. (23), p. 66; Staffan Müller-Wille and Hans-Jörg Rheinberger, *A Cultural History of Heredity*, Chicago: The University of Chicago Press, 2012, pp. 114–15. On gendered labor in the experimental culture of Bateson's plant studies see Marsha L. Richmond, 'Women in the early history of genetics: William Bateson and the Newnham College Mendelians, 1900–1910', *Isis* (1 March 2001), 92(1), pp. 55–90.

[25] Kevles, op. cit. (24), p. 105. The camps were eventually unified over a 1918 paper by R.A. Fisher that demonstrated both continuous and discontinuous approaches as compatible with an underlying Mendelian basis. Harper, op. cit. (23), p. 69.

[26] Harper, op. cit. (21), pp. 424–6.

[27] Jenny Bangham pays particular attention to the material culture and coding schemes circulated with blood collected for the war effort. Bangham, op. cit. (3).

[28] Julia Bell, *The Treasury of Human Inheritance*, vol. 4: *Nervous Diseases and Muscular Dystrophies* (ed. Ronald Aylmer Fisher and Lionel Sharples Penrose), Cambridge: Cambridge University Press, 1948, pp. 343–6.

[29] Bell, op. cit. (28), p. 365. Harper argues that Bell's rejection of nosological 'lumping' is a testament to her analytical rigour, especially compared to the individual investigators upon whose data she relied. Harper, op. cit. (21), p. 428.

https://doi.org/10.1017/S0007087421000224 Published online by Cambridge University Press

Part of the difficulty in carrying out the work was dealing with extra information within the pedigrees, such as notes on cataracts and other mental conditions, not reported consistently by investigators.[30] Extracting tabular data from pedigrees was by no means straightforward – particularly vexed by inconsistent ages of onset – but Bell ventured to propose a dominant pattern of inheritance that awaited confirmation.[31]

Though outside the anglophone context, the distinction of the first tallied-up compendium of known human genetic disorders belongs to German geneticist Otmar Freiherr (baron) von Verschuer, whose role in the genocide at Auschwitz and subsequent redemption by mainstream medicine has been documented by prior historians.[32] His 1959 *Genetik des Menschen: Lehrbuch der Humangenetik* updated an earlier textbook, *Erbpathologie* (Hereditary Pathology) initially published in 1936, with descriptions of 412 known human mutant genes and a decreased emphasis on applications of genetics, perhaps to distance himself from associations with Nazism.[33]

The book represents a halfway point between Bell's *Treasury* and McKusick's early work. Verschuer's text is divided into three major parts: a textbook-style 'Allgemeine Genetik des Menschen' (General genetics of man), that relates research on human heredity to the latest cellular and organismal research, and two catalogue-like sections on 'Spezielle Genetik des Menschen' (Specialized genetics of man), one of 'Normale Eigenschaften' (Normal qualities) and a far more substantial section of 'Krankhafte Eigenschaften' (Pathological qualities).[34] Verschuer counted fifty-one 'normal' phenotypes in all, largely blood groups and features like hair colour, and frequently engaged in speculations on racial difference.[35] The 'pathological' section is organized by medical speciality, consisting of literature reviews alongside reproduced pedigrees and images, without much further analysis. Finally, Verschuer maintained an extensive author index, which McKusick would also prioritize.

As the need for compendia to keep track of the field became clear in the 1950s, intra-speciality synthesis was de rigueur. McKusick's catalogue took a different tack. He abandoned Verschuer's organization by medical speciality, and distinctions between normal and pathological altogether. With the accelerating tempo of research driven by new heredity clinics and interest in the effects of radiation, McKusick thought a more general tool would help stave off a veritable avalanche of printed phenotypes.[36]

[30] For a history of the pedigree see Robert G. Resta, 'The crane's foot: the rise of the pedigree in human genetics', *Journal of Genetic Counseling* (December 1993) 2(4), pp. 235–60.

[31] Bell, op. cit. (28), p. 364.

[32] Benno Müller-Hill, 'The blood from Auschwitz and the silence of the scholars', *History and Philosophy of the Life Sciences* (1999) 21(3), pp. 331–65; Sheila Faith Weiss, 'The loyal genetic doctor, Otmar Freiherr von Verschuer, and the Institut für Erbbiologie und Rassenhygiene: origins, controversy, and racial political practice', *Central European History* (December 2012) 45(4), pp. 631–68.

[33] Arno G. Motulsky, review of *Genetik des Menschen*, by Otmar Freiherr von Verschuer, *American Journal of Human Genetics* (March 1960) 12(1), pp. 139–40. Harper discusses how the German medical genetics community courted respectability in the postwar years. Harper, op. cit. (23), pp. 420–2.

[34] Otmar Freiherr von Verschuer, *Genetik des Menschen: Lehrbuch der Humangenetik*, Munich: Urban & Schwarzenberg, 1959, pp. ix–xi.

[35] Verschuer, op. cit. (34), p. 139. The numbers in his final table do not add up and he refused to attribute dominant or recessive inheritance to any 'normal' features. The lack of enumeration in the text is complicated by the fact that there was no standard format for an entry: some could go on for a number of paragraphs while others got merely a mention.

[36] To borrow a phrase from Ian Hacking, 'Biopower and the avalanche of printed numbers', *Humanities in Society* (1982) 5(3–4), pp. 279–95. For a review of the American development of medical genetics out of eugenics that rejects Kevles's 'reform' narrative and emphasizes instead continuities between eugenic organizations and medical research see Comfort, op. cit. (3).

https://doi.org/10.1017/S0007087421000224 Published online by Cambridge University Press

McKusick's catalogue

Victor A. McKusick came to medical genetics as an internist, and his career as the cornerstone of what would become the Moore Clinic at Johns Hopkins University cemented his status as a founding father of the field. McKusick's research and writing were centred on the clinical encounter. Trained as a cardiologist, he established his reputation with the textbook *Heritable Disorders of Connective Tissue*. Upon taking over Hopkins's genetics clinic in 1956, he began to organize what would become the new discipline's central text. In this section, I discuss McKusick's own ambitions for the cataloguing project in order to frame the technical aspects of its computerization.

In their work on McKusick, Susan Lindee and Andrew Hogan both emphasize his 'cataloging imperative' that inspired other clinicians to pool knowledge and see one-off abnormalities as disorders awaiting characterization.[37] As its sole credited author into the 1990s, when the writing and editing were scaled up and allocated among a team of science writers, McKusick exerted decisive influence over *MIM*.[38] While the catalogue undoubtedly helped foster a newfound focus on individual human genes, this change was evolutionary rather than revolutionary. Unlike Margaret O. Dayhoff's *Atlas of Protein Sequence and Structure*, as discussed by Strasser, *MIM* did not contain valuable information about sequences that could readily be exploited and analysed through the purchase of computer tape.[39] But like Dayhoff's *Atlas*, it formed a kind of publication system in itself: employing an early bibliographic retrieval computer program, it organized the biomedical literature and established a classificatory system for expanding them through a cheap and widely used publication.

MIM grew out of a modest exercise. In 1957, McKusick began holding a monthly journal club at his home for the Moore Clinic fellows to prepare annual reviews of medical genetics for the *Journal of Chronic Diseases*. Participants summarized the literature on a field with an index card for each relevant article: the full reference on one side, relevant disorders and points of interest on the other.[40] In the words of one geneticist, these reviews were the most one could ask for in a rapidly changing field whose 'current state cannot be put into a textbook fast enough – by the time such a text were written and printed, it would be out of date'.[41] The following year, McKusick published a monograph-length article in the *Quarterly Review of Biology* on X-linked traits, sorting them based on how conclusive the evidence of linkage was.[42] Dealing with the other twenty-two chromosomes begat an information management problem that emerged in tandem with visions of a full human genome map.[43]

MIM allegedly began in earnest when David Bolling, McKusick's own 'computer man', observed his ritual of sifting through hundreds of notecards to update his reviews and suggested that using the computer to store it on magnetic tape would vastly simplify the process of adding entries for the annual review.[44] However, it is clear that other

[37] Lindee, op. cit. (6), p. 81; Hogan, op. cit. (6), p. 35.

[38] The personal element of McKusick's authorship is underscored by a dedication to his wife, Dr Anne B. McKusick, at the front of every edition.

[39] Strasser, op. cit. (1), pp. 135–46.

[40] Krishna R. Dronamraju and Clair A. Francomano (eds.), *Victor McKusick and the Development of Medical Genetics*, New York: Springer, 2012, pp. 54–55.

[41] William B. Bean, 'Medical genetics 1958–1960', *JAMA* (1 December 1962) 182(9), p. 971.

[42] McKusick, op. cit. (21). X-linkage was among the best-known genetic markers due to distinctive patterns of inheritance that affected male offspring far more frequently than females.

[43] Andrew J. Hogan, 'The "morbid anatomy" of the human genome: tracing the observational and representational approaches of postwar genetics and biomedicine', *Medical History* (July 2014) 58(3), pp. 315–36.

[44] Victor A. McKusick, 'Genealogic and bibliographic applications of computers in human genetics (unpublished)', 1966, Victor Almon McKusick Collection, the Alan Mason Chesney Medical Archives of the Johns

https://doi.org/10.1017/S0007087421000224 Published online by Cambridge University Press

factors made the project interesting – and feasible. McKusick had already sought the use of electronic computers for related work. At IBM's 6th Medical Symposium in October of 1964, he discussed two such initiatives: one to aid in the calculation of linkages between different genetic loci using human pedigrees in which multiple traits were present, and another to use computers to record more detailed information on isolated populations to aid in longitudinal studies.[45]

This romanticized version of McKusick trading his notecards for punched cards also differs from the account in the forewords of *MIM*. According to McKusick writing in 1966, a full catalogue of recessive disorders was essential to his studies of the Amish.[46] Even when McKusick limited the Amish project to rare, distinctive recessive phenotypes rather than potential chronic conditions that occurred more commonly than one in a thousand people, by 1963 'the complexity of the recessives catalog prompted exploration of computer methods for assembling, revising, and indexing'.[47] Retroactive constructions of the affinity between the computer catalogue and the shaky state of knowledge in the field underestimate the magnitude of the work involved, as well as its centrality to McKusick's own research programme.

We gain more from understanding the catalogue as a tool for executing McKusick's broader vision for medical genetics. To the extent that diseases could be classified as Mendelian, they were typically rare conditions of little consequence to the health of the general population. Therefore medical geneticists had to justify their knowledge as serving a more general project.[48] In contrast to Verschuer, McKusick presented his work as a comprehensive picture of the human genetic make-up built from pathologies: 'Discovery of "new" genetic diseases is not mere "stamp-collecting" ... The catalogs of simply inherited genetic traits in man are like photographic negatives from which a positive picture of the normal genetic constitution of man can be constructed.'[49] Increasing the resolution of these negatives required McKusick to convince his medical colleagues that genetic classification had to aim for the exclusive rigour of cleanly delineated phenotypes. In a 1969 article on 'lumping' and 'splitting' in genetic nosology, McKusick displayed his keen awareness of the politics of taxonomy and implicitly declared himself a splitter.[50]

Hopkins Medical Institutions (subsequently VAM), Box 509179, Folder 533, 'Computers in research'. McKusick's taxonomic strategies invite comparison to those of Linnaeus, who kept quarto sheets of species with their references in the literature, enabling him to experiment with orders by shuffling them. Staffan Müller-Wille and Isabelle Charmantier, 'Natural history and information overload: the case of Linnaeus', *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* (March 2012) 43 (1), pp. 4–15.

[45] Victor A. McKusick, 'Some computer applications to problems in human genetics (draft)', 1964, VAM, Box 509169, Folder 527, 'IBM'; McKusick, 'Some computer applications to problems in human genetics', *Methods of Information in Medicine* (1965) 4(4), pp. 183–9.

[46] For an expansive look at McKusick's Amish research see Lindee, op. cit. (6).

[47] Victor A. McKusick, *Mendelian Inheritance in Man: Catalogs of Autosomal Dominant, Autosomal Recessive, and X-Linked Phenotypes*, 1st edn, Baltimore: Johns Hopkins University Press, 1966, p. vii.

[48] Others, such as James Neel, saw this focus on rare disorders as a handicap for the field, once asking rhetorically in a lecture, 'in our concern for the individual, have we forgotten to set up the team which has as its concern the species as a whole?' James V. Neel, *Physician to the Gene Pool: Genetic Lessons and Other Stories*, New York: J. Wiley, 1994, p. 32.

[49] Victor A. McKusick, John A. Hostetler and Janice A. Egeland, 'Genetic studies of the Amish', *Bulletin of the Johns Hopkins Hospital* (1964) 115, pp. 203–22, 215. On 'stamp collecting' as a pejorative see Kristin Johnson, 'Natural history as stamp collecting: a brief history', *Archives of Natural History* (1 October 2007) 34(2), pp. 244–58; Strasser, op. cit. (1), p. 257.

[50] Victor A. McKusick, 'On lumpers and splitters, or the nosology of genetic disease', *Perspectives in Biology and Medicine* (1969) 12(2), pp. 298–312. On lumping and splitting as a battleground of scientific credibility see Jim Endersby, *Imperial Nature: Joseph Hooker and the Practices of Victorian Science*, Chicago: The University of Chicago Press, 2008.

https://doi.org/10.1017/S0007087421000224 Published online by Cambridge University Press

A comprehensive nosology of Mendelian traits allowed him to leverage an exponentially growing number of discrete clinical studies in favour of this perspective.[51] He envisioned each chromosome with its own catalogue, which underlines the sway of the chromosomal perspective that would not give way to a genomic one until the late 1980s.[52] The map, as the saying goes, would not be the territory.

By organizing the literature into discrete genetic entities, provisional as they were, McKusick helped instil a gene-centred view of hereditary disorders before human DNA sequences were even available. However, this path was neither straightforward nor inevitable, and involved translating a bibliographic endeavour organized around medical specialities into a database of distinct genetic disorders, with a unique and stable numerical ID for each gene.

Going digital

Early reviews of *MIM* evince commingled excitement and anxiety over computerized management of biomedical knowledge. One reviewer praised the self-consciously provisional character of the printed-out electronic database (Figure 1) as offering 'the hope that the mechanics for keeping it up-to-date are at hand'.[53] Another was less sanguine about such 'heavy reliance on the computer', citing poor editorial oversight – typos and duplicate entries abounded – and insufficient critical attention that allowed speculative publications to be reified into putative biological entities.[54] Nonetheless, *MIM* was hailed as a model for leveraging technology to manage an unwieldy literature:

> The entire work, which would be impossible without the aid of the computer, is an instance of the organization and presentation of scientific information that must occur in many fields ... for the mountainous masses of knowledge that have accumulated must somehow be rendered accessible to would-be users and new investigators.[55]

Taking concerns about 'information overload' seriously does not mean taking them at face value. The mundane organization techniques of this emergent information order offer a glimpse at a computerized political economy of knowledge in the making.

This section considers how *MIM* was compiled with the equipment available to the Moore Clinic. Geneticists working today might be most familiar with *MIM* for providing the ID numbers used to organize all known human genes and their alleles. The catalogue was alphabetical, organized into three major sections – dominant, recessive, and X-linked traits, indicated by the first numeral of the ID – with long indexes of authors and disorders. However, a closer investigation of source material reveals that this scheme was a contingent feature that emerged over time. Both in McKusick's mind and in the bibliographic software used to compile *MIM*, *author* was the primary record locator and organizing principle. Unpacking how *MIM* was created from this early computer system allows us to see how, in contrast to McKusick's elegant conceptualizations of genetic nosology,

[51] On McKusick's nosological approach see Andrew J. Hogan, 'Locating genetic disease: the impact of clinical nosology on biomedical conceptions of the human genome (1966–1990)', *New Genetics and Society* (2013) 32(1), pp. 78–96.

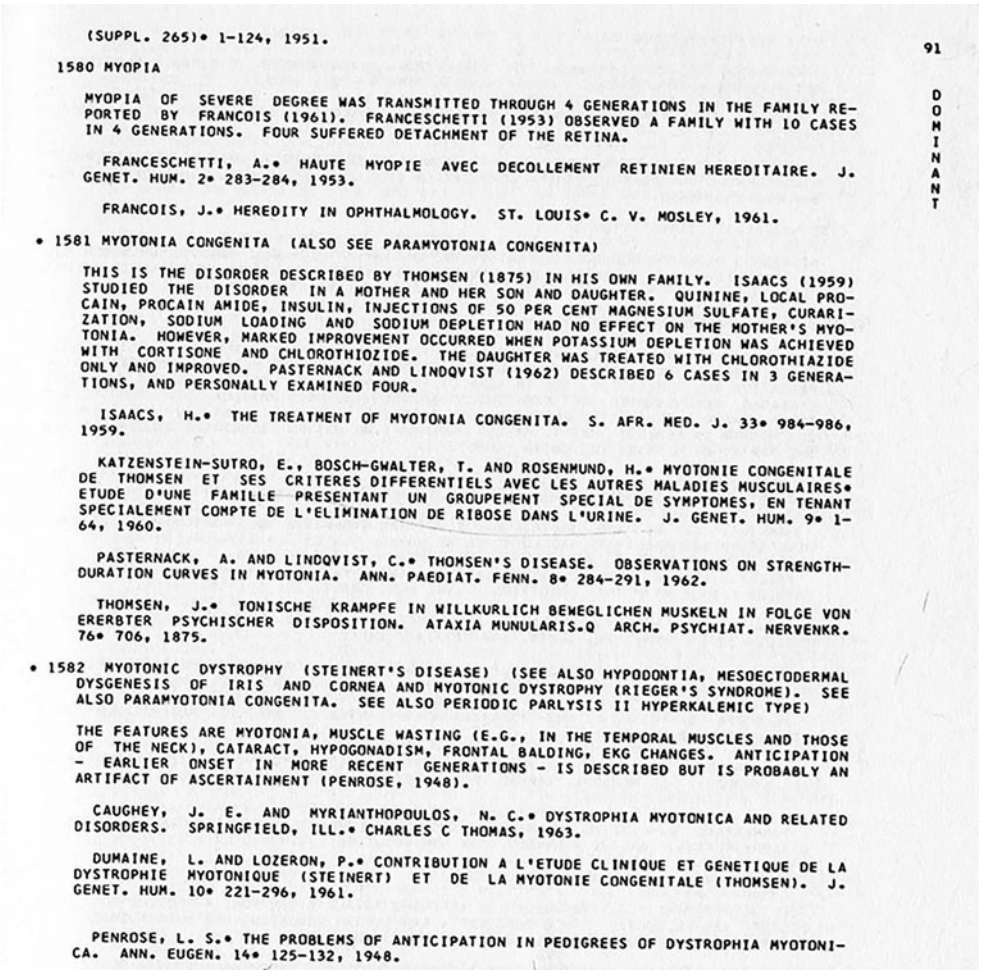
[52] Victor A. McKusick, *Mendelian Inheritance in Man: Catalogs of Autosomal Dominant, Autosomal Recessive, and X-Linked Phenotypes*, 11th edn, Baltimore: Johns Hopkins University Press, 1994, p. xix.

[53] Knudson, op. cit. (1), p. 426.

[54] H. Eldon Sutton, review of *Mendelian Inheritance in Man*, by Victor A. McKusick, *JAMA* (24 April 1967) 200(4), p. 351.

[55] Glass, op. cit. (1), p. 456.

https://doi.org/10.1017/S0007087421000224 Published online by Cambridge University Press



(SUPPL. 265)• 1-124, 1951.

91

DOMINANT

1580 MYOPIA

MYOPIA OF SEVERE DEGREE WAS TRANSMITTED THROUGH 4 GENERATIONS IN THE FAMILY REPORTED BY FRANCOIS (1961). FRANCESCHETTI (1953) OBSERVED A FAMILY WITH 10 CASES IN 4 GENERATIONS. FOUR SUFFERED DETACHMENT OF THE RETINA.

FRANCESCHETTI, A.• HAUTE MYOPIE AVEC DECOLLEMENT RETINIEN HEREDITAIRE. J. GENET. HUM. 2• 283-284, 1953.

FRANCOIS, J.• HEREDITY IN OPHTHALMOLOGY. ST. LOUIS• C. V. MOSLEY, 1961.

• 1581 MYOTONIA CONGENITA (ALSO SEE PARAMYOTONIA CONGENITA)

THIS IS THE DISORDER DESCRIBED BY THOMSEN (1875) IN HIS OWN FAMILY. ISAACS (1959) STUDIED THE DISORDER IN A MOTHER AND HER SON AND DAUGHTER. QUININE, LOCAL PROCAIN, PROCAIN AMIDE, INSULIN, INJECTIONS OF 50 PER CENT MAGNESIUM SULFATE, CURARIZATION, SODIUM LOADING AND SODIUM DEPLETION HAD NO EFFECT ON THE MOTHER'S MYOTONIA. HOWEVER, MARKED IMPROVEMENT OCCURRED WHEN POTASSIUM DEPLETION WAS ACHIEVED WITH CORTISONE AND CHLOROTHIOZIDE. THE DAUGHTER WAS TREATED WITH CHLOROTHIAZIDE ONLY AND IMPROVED. PASTERNACK AND LINDQVIST (1962) DESCRIBED 6 CASES IN 3 GENERATIONS, AND PERSONALLY EXAMINED FOUR.

ISAACS, H.• THE TREATMENT OF MYOTONIA CONGENITA. S. AFR. MED. J. 33• 984-986, 1959.

KATZENSTEIN-SUTRO, E., BOSCH-GWALTER, T. AND ROSENMUND, H.• MYOTONIE CONGENITALE DE THOMSEN ET SES CRITERES DIFFERENTIELS AVEC LES AUTRES MALADIES MUSCULAIRES• ETUDE D'UNE FAMILLE PRESENTANT UN GROUPEMENT SPECIAL DE SYMPTOMES, EN TENANT SPECIALEMENT COMPTE DE L'ELIMINATION DE RIBOSE DANS L'URINE. J. GENET. HUM. 9• 1-64, 1960.

PASTERNACK, A. AND LINDQVIST, C.• THOMSEN'S DISEASE. OBSERVATIONS ON STRENGTH-DURATION CURVES IN MYOTONIA. ANN. PAEDIAT. FENN. 8• 284-291, 1962.

THOMSEN, J.• TONISCHE KRAMPFE IN WILLKURLICH BEWEGLICHEN MUSKELN IN FOLGE VON ERERBTER PSYCHISCHER DISPOSITION. ATAXIA MUNULARIS.Q ARCH. PSYCHIAT. NERVENKR. 76• 706, 1875.

• 1582 MYOTONIC DYSTROPHY (STEINERT'S DISEASE) (SEE ALSO HYPODONTIA, MESOECTODERMAL DYSGENESIS OF IRIS AND CORNEA AND MYOTONIC DYSTROPHY (RIEGER'S SYNDROME). SEE ALSO PARAMYOTONIA CONGENITA. SEE ALSO PERIODIC PARLYSIS II HYPERKALEMIC TYPE)

THE FEATURES ARE MYOTONIA, MUSCLE WASTING (E.G., IN THE TEMPORAL MUSCLES AND THOSE OF THE NECK), CATARACT, HYPOGONADISM, FRONTAL BALDING, EKG CHANGES. ANTICIPATION - EARLIER ONSET IN MORE RECENT GENERATIONS - IS DESCRIBED BUT IS PROBABLY AN ARTIFACT OF ASCERTAINMENT (PENROSE, 1948).

CAUGHEY, J. E. AND MYRIANTHOPOULOS, N. C.• DYSTROPHIA MYOTONICA AND RELATED DISORDERS. SPRINGFIELD, ILL.• CHARLES C THOMAS, 1963.

DUMAINE, L. AND LOZERON, P.• CONTRIBUTION A L'ETUDE CLINIQUE ET GENETIQUE DE LA DYSTROPHIE MYOTONIQUE (STEINERT) ET DE LA MYOTONIE CONGENITALE (THOMSEN). J. GENET. HUM. 10• 221-296, 1961.

PENROSE, L. S.• THE PROBLEMS OF ANTICIPATION IN PEDIGREES OF DYSTROPHIA MYOTONICA. ANN. EUGEN. 14• 125-132, 1948.

Figure 1. Victor A. McKusick, *Mendelian Inheritance in Man: Catalogs of Autosomal Dominant, Autosomal Recessive, and X-Linked Phenotypes*, 1st edn, Baltimore: Johns Hopkins University Press, 1966, p. 91.

the gene was secondary to a bibliographical information order. Lest these organizational changes appear trivial, I conclude by reviewing some of the systems and publications linked to *MIM*'s ID system.

At the September 1966 International Congress of Human Genetics in Chicago, where the community focused on issues of standardization, McKusick highlighted the Moore Clinic's bibliographic use of computers rather than his statistical linkage mapping research.[56] The computer program retained features of McKusick's manual abstract-compiling process: article abstracts were assigned paragraph numbers and index words, and references tied to the number of the paragraph where they were discussed, allowing an alphabetical list of authors and an index to be generated and stored on magnetic tape. While Bell's *Treasury* synthesized clinical data alongside relevant medical histories, McKusick *et al.*'s paragraph-length summaries compiled the latest research findings in a

[56] John L. Hamerton, 'Chicago conference 1966: standardization in human cytogenetics', *Journal of Medical Genetics* (September 1967) 4(3), p. 226.

https://doi.org/10.1017/S0007087421000224 Published online by Cambridge University Press

largely additive fashion. Because clinical geneticists worked in a system in which credit accrued to individuals, due to their tight social network – 'one recalls that X.Y.Z. Jones described a peculiar syndrome in 1931' – the author remained the most important piece of information. A virtue of the system, as McKusick saw it, was the 'automatic updating of the numbers in the indices each time a new item [was] added to the catalogs ... changing the numbering in the text'.[57] References to stable genetic entities were not built into its design.

The Moore Clinic shared a system developed by Jane Olmer and Robert Rich for Hopkins's Applied Physics Laboratory (APL) using an IBM 1401 computer. Though the bibliographic program was coded to manage the APL's literature, interest from users like McKusick led them to envision a system 'general and flexible enough to handle a wide variety of applications in a multiplicity of input and output formats'.[58] Magnetic tape was the storage medium of choice for a multiple-user system, since the tape could be moved between units for processing or printing.[59] All documents were 'key punched and fed directly to the library master tape', which made 'searching a matter of selection with no sorting required'.[60] Though more sophisticated searching was possible, Olmer's technique used a set of thesaurus terms for each user that were appended to individual entries, demarcated by slash marks to set them apart from free-standing text. The approach privileged general usability and portability to a better system, and was chosen over viable alternatives to satisfy the most users and work best with the 1401.

In early negotiations with the publisher, Johns Hopkins University Press, one of McKusick's staff argued that the book be compiled straight from computer printouts to speed up the proofing process, and claimed that reviewers who encountered the text in this format failed to note the drawbacks of all-caps text and limited characters.[61] Photo-offset printing of computer output had been pioneered in the early 1960s by the US Government Printing Office, but was deemed so unreadable that they commissioned an adaptation of Linotype technology for magnetic tape.[62] For McKusick and his collaborators, convenience and control superseded any aesthetic concerns.

McKusick insisted that scientific advances, rather than the economics of publishing, determine when the catalogue was released and how it was distributed. He wanted to print up to ten thousand copies, but his editor insisted that they limit the first run to three thousand bound copies and two thousand sets of unbound sheets, and suggested that he consider planning for supplements.[63] McKusick bristled at the suggestion; he preferred to publish a new volume 'even if it means dumping a thousand copies in the bay', ensuring that 'advances in genetic nosology [were] the only consideration in republishing, not the backlog of an unsold prior edition'.[64] Nonetheless, in 1966, three thousand copies

[57] McKusick, op. cit. (44), pp. 7–8.

[58] Jane Olmer and Robert P. Rich, 'A flexible direct file approach to information retrieval: text edit, search or select and print on an IBM 1401', in *Proceedings of the American Federation of Information Processing Societies 1963 Fall Joint Computer Conference*, Baltimore: Spartan Books, 1963, pp. 173–82, 173.

[59] International Business Machines Corporation, *IBM 1401 Systems Summary*, Endicott, NY: IBM Product Publications, 1963, p. 8. Hard disks and magnetic drums were also in use at this time.

[60] F.L. Kennedy and M.E. Brown, *The Applications of Computers to the APL Storage and Retrieval System*, TG-669, Silver Spring, MD: Johns Hopkins University, Applied Physics Laboratory, 1965, pp. 2–4, 7.

[61] Richard H. Shepard to Jack G. Goellner, March 1966, VAM, Box 509269, Folder 'Mendelian Inheritance in Man'.

[62] Richard L. Worsnop, 'Computers in publishing', in *Editorial Research Reports 1968*, vol. 2, Washington, DC: Congressional Quarterly, Inc., 1968, pp. 505–22.

[63] Jack G. Goellner to Victor A. McKusick, 20 April 1966, VAM, Box 509269, Folder 'Mendelian Inheritance in Man'.

[64] Victor A. McKusick to Jack G. Goellner, 26 April 1966, VAM, Box 509269, Folder 'Mendelian Inheritance in Man'; McKusick, op. cit. (44), p. 8.

of the 368-page catalogue were bound and sold for only eight dollars, due to low production costs. New editions every few years would be sent to press in March to be bound by July, just in time for the annual course on medical genetics McKusick ran in Bar Harbor, Maine.

The catalogue's popularity meant that the collaboration with the Applied Physics Laboratory would not continue indefinitely. In 1971, the group acquired its own microcomputer, a Datapoint 2200, that allowed them to bring their editing process in-house and eliminate many of the earlier method's steps.[65] The new machine did not help them appease their editor, eagerly awaiting 'the day when computers will print both upper and lower case'.[66] Having an in-house computer did, however, allow McKusick and his team to respond better to user suggestions.

Recalling that McKusick saw the program's ability to constantly update entry numbers as a feature, rather than a bug, it is noteworthy that it took until the third edition for him to change his mind. 'Some geneticists expressed a desire to use the numbering system of these catalogues as the basis of a diagnostic and bibliographic filing system and were distressed by the change of numbering between the first and second editions', he noted in his introduction.[67] McKusick quickly saw the benefits of having stable IDs, an innovation that helped enable cross-referencing and mapping human genes on a large scale. The schema allowed him to construct tables and appendices of valuable biological and clinical markers for quick consultation, which would not be rendered useless by constant re-enumeration. One such appendix was the map of human chromosomes emerging from the first Human Gene Mapping conference at Yale in 1973, updated in all editions following the fourth. *MIM* became an integral resource for coordinating the mapping work using somatic cell genetics; an appendix was even added listing genes with mutant cell lines available from a repository in New Jersey to facilitate further research.[68]

Though seemingly trivial, the implementation of a standard enumeration scheme for *MIM* turned this printed database into a publication infrastructure for medical genetics, linking up cutting-edge mapping work with protein biochemistry research and the filing cabinets of clinicians worldwide.[69] *MIM* incorporated others' databases before the hyperlinked online version, including information from Dayoff's *Atlas* on protein sequences for

[65] The Datapoint 2200 was born of a failed collaboration between the Computer Terminal Corporation and Intel that supported the development of the seminal Intel 8008 microchip. Lamont Wood, 'Forgotten PC history: the true origins of the personal computer', *Computerworld*, 8 August 2008, at www.computerworld.com/article/2532590/computer-hardware/forgotten-pc-history--the-true-origins-of-the-personal-computer.html.

[66] Jack G. Goellner to Victor A. McKusick, 22 January 1969, VAM, Box 509269, Folder 'Mendelian Inheritance in Man'. This goal was only satisfied a few years later using photo-offset images of a computer monitor rather than actual printouts. Victor A. McKusick, *Mendelian Inheritance in Man: Catalogs of Autosomal Dominant, Autosomal Recessive, and X-Linked Phenotypes*, 5th edn, Baltimore: Johns Hopkins University Press, 1978, p. xxiv.

[67] Victor A. McKusick, *Mendelian Inheritance in Man: Catalogs of Autosomal Dominant, Autosomal Recessive, and X-Linked Phenotypes*, 3rd edn, Baltimore: Johns Hopkins University Press, 1971, p. xlv.

[68] Victor A. McKusick, *Mendelian Inheritance in Man: Catalogs of Autosomal Dominant, Autosomal Recessive, and X-Linked Phenotypes*, 6th edn, Baltimore: Johns Hopkins University Press, 1983, p. ix. On somatic cell genetics see Hannah Landecker, *Culturing Life: How Cells Became Technologies*, Cambridge, MA: Harvard University Press, 2007.

[69] A number of foreign-language editions of *MIM* were published: a Spanish edition for Mexico in 1976 (translated by Rudolfo Guzmán Toledano), a Russian edition also in 1976 (translated by E.K. Gentera and V.I. Ivanova) and an eventual two-volume Mandarin edition in 1996 (translated by Wilson H.Y. Lo and others). McKusick, op. cit. (5), p. 589. A catalogue of teratogens made use of the *MIM* software. Thomas H. Shepard, *Catalog of Teratogenic Agents*, 6th edn, Baltimore: Johns Hopkins University Press, 1989.

https://doi.org/10.1017/S0007087421000224 Published online by Cambridge University Press

haemoglobin and a World Health Organization report on glucose-6-phosphate dehydrogenase diversity in the global population.[70] It even became the basis for a hospital network computer database for clinical genetics, MEDGEN, run by the University of California, San Francisco.[71] A classification system developed to keep track of an exploding literature helped reorient biomedical research around distinct gene entities, providing the basis for other systems of storage, circulation and interconnection.

A palimpsest of pathology

While there is little doubt that *MIM* saw wide circulation and was central to the work of human genetics, one still might ask how its information order shaped the norms of the emerging field, or even how it was put to use. While the latter question remains largely elusive, following changes in an exemplary set of disorders – the myotonic dystrophy family discussed above – through each edition reveals important features of how the catalogue worked, with broad import for the historical analysis of reference material.

First, centralization is never neutral. Cataloguing is a human effort, often bound up in institutions with interests, constraints and values. While McKusick always encouraged readers to submit information, many of the citations in *MIM* came from his browsing of the Institute for Scientific Information's *Current Contents*, a weekly digest of cover pages throughout the biomedical sciences spearheaded by Eugene Garfield, a linguist regarded as one of the founders of scientometrics. While contributions came from hundreds of journals, an internal analysis in 1994 revealed that two-thirds of all entries came from just twenty-one high-impact journals.[72] Though 'largely a bibliographic task', the mark of the Moore Clinic and its steady rotation of international and domestic fellows was apparent in the inclusion of 'unpublished observations' derived from their own research projects; further, 'much judgment based on personal experience was necessary in selecting items for inclusion and in deciding the manner in which they should be treated'.[73] Whether used as filters or merely presented alongside published results, McKusick's own assumptions and clinical data assumed the mantle of fact. The write-up of myotonic dystrophy in the fourth edition discussed the results of a study from his lab without direct attribution, and this very segment was duplicated within the entry in a later edition.[74]

Second, not all entries are created equal. Subjects of controversy tend to attract the most attention, and this is particularly the case with the genetic loci found in *MIM*. The first edition of *MIM* (1966) included four variants of myotonic dystrophy (a fifth would later be added; two are shown in Figure 1), and while the entries themselves would remain, the number of references they contained (Figure 2) and their genetic status would come to differ greatly.[75] When mapping efforts expanded during the 1980s, the

[70] Victor A. McKusick, *Mendelian Inheritance in Man: Catalogs of Autosomal Dominant, Autosomal Recessive, and X-Linked Phenotypes*, 2nd edn, Baltimore: Johns Hopkins University Press, 1968, pp. 85–113, 398–401. On Dayhoff's *Atlas* see Strasser, op. cit. (1).

[71] J.A. Mitchell, W.D. Loughman and C.J. Epstein, 'GENFILES: a computerized medical genetics information network. II. MEDGEN: the clinical genetics system', *American Journal of Medical Genetics* (1980) 7(3), pp. 251–66.

[72] Victor A. McKusick, 'A multiauthored OMIM medical genetics knowledgebase', 1 November 1994, VAM, Box 511334, Folder '1995', p. 7.

[73] McKusick, op. cit. (44), p. 6.

[74] Victor A. McKusick, *Mendelian Inheritance in Man: Catalogs of Autosomal Dominant, Autosomal Recessive, and X-Linked Phenotypes*, 4th edn, Baltimore: Johns Hopkins University Press, 1975, p. 224; McKusick, *Mendelian Inheritance in Man: Catalogs of Autosomal Dominant, Autosomal Recessive, and X-Linked Phenotypes*, 7th edn, Baltimore: Johns Hopkins University Press, 1986, p. 511.

[75] All of these conditions were marked with an asterisk, McKusick's way of noting a disorder for which 'the particular mode of inheritance is considered quite certain'. McKusick, op. cit. (44), p. xvii.

https://doi.org/10.1017/S0007087421000224 Published online by Cambridge University Press

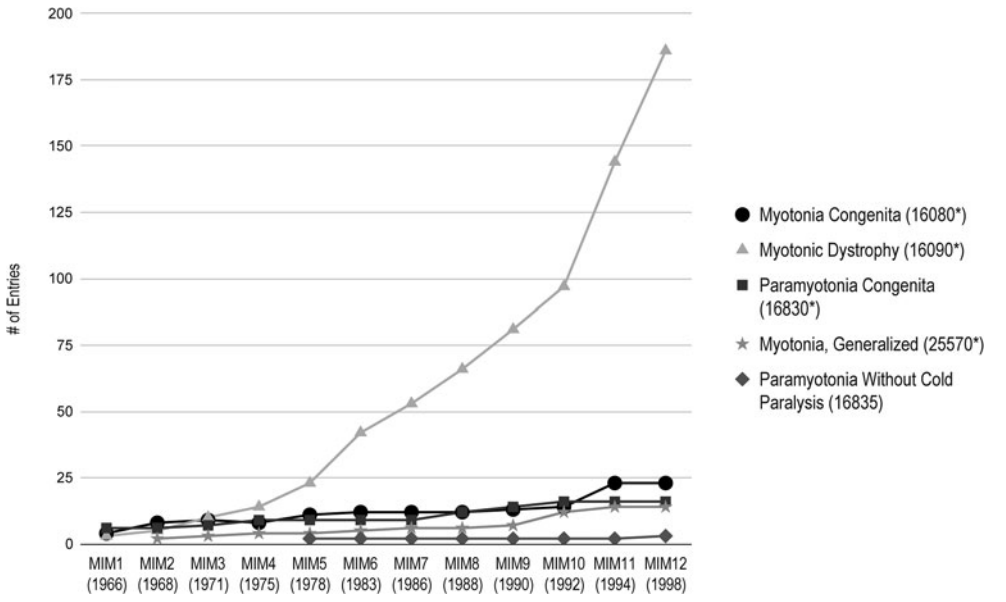


Figure 2. References included in various myotonia entries, compiled by the author. Numbers taken from third-edition designations.

literature on myotonic dystrophy grew exponentially as researchers came closer to pinning down its location within the human genome.[76] The entry became increasingly unwieldy, taking up nearly an entire page; new information was tacked on as it came in, and no distinctions were drawn between the kinds of studies discussed. Finally, McKusick brought on board a team of medical writers who consulted with various specialists and reorganized such entries into sizable review essays.[77] Other variants did not change substantially over the course of three decades.

Finally, knowledge is sticky. Paying attention to changing entries across editions shows that McKusick largely worked in an additive fashion – a feature of the early bibliographic-entry software – and entries were often more chronological than prescriptive. Only a major upheaval could compel him to make a substantial revision. A fascinating example of this was a problem known as 'anticipation': the notion that a genetic disorder could get worse and have an earlier onset in successive generations. Even before *MIM*'s first edition, McKusick had authoritatively cited a prominent article claiming this phenomenon as a statistical artefact.[78] However, molecular studies showed that successive generations did, in fact, accumulate repeat mutations in the gene, and McKusick was eventually forced to amend the entry.[79] This raises a more general problem. *MIM* began as a catalogue of *phenotypes* – the physical manifestation of an underlying *genotype*, or presumably unique genetic variant observed in the clinic. But what would happen to entries once

[76] McKusick, op. cit. (72), pp. 366–8.

[77] Myotonic dystrophy (160900) was completely rewritten with major sections on mapping, molecular genetics, population genetics, diagnosis, clinical management, animal models and history. McKusick, op. cit. (52), pp. 985–9.

[78] Victor A. McKusick, *Medical Genetics, 1958–1960: An Annotated Review*, St Louis: Mosby, 1961, p. 79.

[79] McKusick, op. cit. (20), p. 747. Cases of cytogenetic and molecular research like this disrupted the idea of genetic traits passed on as stable, informational entities. Judith E. Friedman, 'Anticipation in hereditary disease: the history of a biomedical concept', *Human Genetics* (1 December 2011) 130(6), pp. 705–14.

https://doi.org/10.1017/S0007087421000224 Published online by Cambridge University Press

a coding sequence for a protein involved in more general pathways was discovered to be responsible? Rather than collapse entries, they decided to link them using a hash tag to indicate phenotypes caused by another mutation.[80]

Eventually, molecular studies led the catalogue to abandon its numbering scheme separating dominant and recessive disorders. Having already appended a sixth digit to all entries and allowed for decimals in order to account for molecular variants, they adopted a new digit, 6, to prefix all new entries.[81] *Mendelian Inheritance in Man* ceased to be truly Mendelian as it developed in tandem with knowledge in the field, and it increasingly relied on pointers between different databases as knowledge about underlying molecular entities accrued.[82] Technical fixes like this were the tip of the iceberg; reimagining *MIM* for a new era required accommodating its information order to work alongside other resources, and without McKusick.

The centre cannot hold

The transition between the physical *MIM* and its online counterpart entailed more than porting a computerized file to a singular computer network, such as the Internet. Rather, it was bound up in negotiations over how to make all manner of new genetic information available over different networks. In the realm of sequences, this took the form of a valuable public–private contract vied for by different biotech start-ups at the outset of the genomic age.[83] However, the bibliographic information compiled for *OMIM* had a different trajectory, traced in this final section.

In 1986, the director of Hopkins's Welch Medical Library, Nina Matheson, approached McKusick with the offer of creating an online version of *MIM* that would serve as a testbed for a contextual search engine, IRX (Information Retrieval Experiment), being developed by the National Library of Medicine (NLM).[84] Editors at Hopkins Press stood their ground against making the catalogue available through a computer connection, but the Howard Hughes Medical Institute funded the joint venture the following year.[85] This was a time of rapid growth for medical informatics. The Medical Library Assistance Act of 1965 had substantially broadened the NLM's purview, allowing it to invest in computing technology in addition to serving regional medical library needs.[86] A 1987 report declared the urgent need for a National Center for Biotechnology Information (NCBI), established the

[80] For example, paramyotonia congenita (16830) took on a hash tag in the tenth edition on the basis of 'evidence that this disorder is due to mutation in the SCN4A gene (170500), the gene coding for the same sodium channel that is mutant in hyperkalemic periodic paralysis (HYPP)'. McKusick, op. cit. (20), p. 825.

[81] Victor A. McKusick, *Mendelian Inheritance in Man: Catalogs of Autosomal Dominant, Autosomal Recessive, and X-Linked Phenotypes*, 9th edn, Baltimore: Johns Hopkins University Press, 1990, p. ix.

[82] On how the gene mapping workshops required creative database solutions see de Chadarevian, op. cit. (6), pp. 165–71.

[83] On the development of the Internet out of heterogeneous networks for defence research see Janet Abbate, *Inventing the Internet*, Cambridge, MA: MIT Press, 1999. On another effort to bring bibliographic resources and sequence data online in a private–public collaboration, namely through the biotech start-up IntelliGenetics, see November, op. cit. (11), pp. 267–8; Stevens, op. cit. (11), pp. 157–68; Strasser, op. cit. (1), pp. 213–14.

[84] D. Harman, Dennis Benson, Larry Fitzpatrick, Rand Huntzinger and Charles Goldstein, 'IRX: an information retrieval system for experimentation and user applications', *SIGIR Forum* (May 1988) 22(3–4), pp. 2–10.

[85] 'OMIM's 50th anniversary symposium and message from past Welch director, Nina Matheson', *Welch Medical Library Blog*, at http://blog.welch.jhmi.edu/node/381, accessed 27 February 2018; Robert Cook-Deegan, *The Gene Wars: Science, Politics, and the Human Genome*, New York: W.W. Norton, 1994, p. 123.

[86] 'Public Law 89-291 – an Act to amend the Public Health Service Act to provide for a program of grants to assist in meeting the need for adequate medical library services and facilities', 22 October 1965, National Library of Medicine, John E. Fogarty Papers, at https://profiles.nlm.nih.gov/101743404X405. Original repository: Legislative Records, Phillips Memorial Library, Special and Archival Collections at Providence College. This

https://doi.org/10.1017/S0007087421000224 Published online by Cambridge University Press

following year, to meet the demands of the growing genetics community.[87] Constant collaboration between the Welch and the NLM, bolstered by their geographical proximity, supported this multi-institutional effort to keep *OMIM* available as a public resource, despite changes at the NCBI and in the cost of maintaining the staff and servers to support it.

Throughout the next few years, McKusick remained largely responsible for the content of *MIM* and its online counterpart. 1993 saw the beginning of major changes to the project: entries and names were updated to better interact with the HGP-proximal Genome Data Base (GDB) project, and a full editorial board was implemented to handle different subject areas.[88] As automatic sequence submission systems were put in place, the *OMIM* project argued that expert-based data curation was more valuable than ever. It had vocal allies. Stanford human geneticist and eventual scientific director of 23andMe, Uta Francke, and Phyllis J. McAlpine, director of the Human Genome Organization's nomenclature committee, both claimed that *OMIM* entries often replaced a literature search for overworked researchers struggling to put together grant proposals.[89] A review in the first issue of *Genome Biology* declared *OMIM*'s annotations 'second to none … a result of their policy of manual curation', and evinced anxiety over 'the coming deluge of data'.[90] Users of the platform valued its connectedness and curation equally.

Yet there were always tensions over the project's ownership. Managing a team of science writers and coordinating efforts to stay in sync with other digital resources proved difficult for subsequent directors, who had to contend with McKusick's ongoing presence even as he took the back seat. Throughout his correspondence during these tumultuous years of growth, he expressed constant concern that technical choices and personnel changes would derail his vision: 'There is concern by many, not only within my group but in the genetics community at large, that we will screw up [in the margins: 'destroy?'] what is now a very successful operation. OMIM is too valuable to let that happen.'[91] Nonetheless, McKusick stepped back over the years and let a new, dedicated team of science writers – with many of whom he collaborated closely – take charge.

By 2005, the NCBI assumed full control of *OMIM*, by that point fully hooked into its suite of related resources: 'genetic databases such as DNA and protein sequence, PubMed references, general and locus-specific mutation databases, HUGO nomenclature, MapViewer, GeneTests, patient support groups and many others'.[92] Strasser argues that the NCBI's success in uniting publication platforms with sequence databases (and mandatory sequence submission) represented a triumph of open science long resisted by the scientific community and journal publishers.[93] In practice, however, this consolidation proved to be short-lived for *OMIM*. Contract restrictions forced it to separate from the NCBI, landing it back where it began at Hopkins. This allowed its team to be more flexible

push happened around the same time as the NIH invested heavily in computers for biomedical investigators, an oft-neglected history. November, op. cit. (11).

[87] Board of Regents, National Library of Medicine, *Long Range Plan*, Washington, DC: US Dept of Health and Human Services, 1987, at http://hdl.handle.net/2027/mdp.39015012539709.

[88] Peter Pearson, Clair Francomano, Patricia Foster, Carol Bocchini, Peter Li and Victor McKusick, 'The status of *Online Mendelian Inheritance in Man* (OMIM) Medio 1994', *Nucleic Acids Research* (September 1994) 22(17), 3470–3; Ronald Kotulak, 'Hopkins establishes genetic database', *The Sun (1837–1992)* (Baltimore), 24 July 1990.

[89] Uta Francke to Victor A. McKusick, 9 December 1994, VAM, Box 511334, Folder '1995'; Phyllis J. McAlpine to Victor A. McKusick, 14 December 1994, VAM, Box 511334, Folder '1995'.

[90] Colin Semple, 'The thousand doors to disease', *Genome Biology* (18 September 2000) 1, reports2050, at https://doi.org/10.1186/gb-2000-1-3-reports2050.

[91] Cover note *c.* summer 1993, VAM, Box 511128, Folder 'OMIM – Job Description'.

[92] Ada Hamosh , Alan F. Scott, Joanna S. Amberger, Carol A. Bocchini and Victor A. McKusick, '*Online Mendelian Inheritance in Man* (OMIM), a knowledgebase of human genes and genetic disorders', *Nucleic Acids Research* (1 January 2005) 33, pp. D514–17.

[93] Strasser, op. cit. (1), pp. 251–2.

https://doi.org/10.1017/S0007087421000224 Published online by Cambridge University Press

in their information management and website design, but also made the project dependent on grants and donations rather than funded by federal funds earmarked for maintaining open-access scientific resources.[94] Although *MIM* helped usher the genomic information order represented by the NCBI's managed databases into being, providing both a scaffold and an inspiration, today it exists independently.

Conclusion

In this paper, I have argued that *MIM* formed an information order that helped shape the credit system of human genetics around individual genes. This was not one of the project's initial aims – indeed, it had originally been organized by reference to the author of a particular entry – but it emerged as a consequence of evolving technological solutions to problems of information management and the demands of users. In the midst of present debates over open-access journals, government data-rescue initiatives, replication crises, and how to make science more democratic, I concur with Alex Csiszar's claim that historians of science need to 'attend to the ways in which information technologies built to extend access to knowledge can also become technologies of valuation and exclusion'.[95] The history recounted here is an effort to disentangle some of the infrastructure that allows patients, physicians, geneticists and even biotech marketers to stay up to date on genetic research, and to attend to how information orders shape and are shaped by the technologies through which they are realized.

Just as the original *MIM* had been an early example of the use of an electronic digital computer to generate a publication, McKusick called *OMIM* 'one of the first electronic resources to exploit the advantages of the Web'.[96] Rather than matters of bald fact, these priority claims, echoed by others, reflect the self-perception of human geneticists as early adopters of information technology. They also raise the matter of how central McKusick was to the myth-making and community value of *MIM* and *OMIM*. He continued to refer to the project idiosyncratically as a 'knowledgebase', to the irritation of his systems operators.[97] Speaking of the editorial reorganization project, one reviewer quipped that the labour should be thought of in terms of 'Whole McKusick Equivalents', or WMEs.[98] Reconstructing the broken links of this digital history makes *OMIM* appear more and more like a 'legacy system' in a dual sense: somewhat outmoded yet providing shape to an organization, and a tribute to the power of personality that brought such resources together in the first place.

[94] Amberger, Bocchini and Hamosh, op. cit. (5).

[95] Csiszar, op. cit. (4), p. 58.

[96] McKusick, op. cit. (5), p. 591.

[97] There is a long-standing dispute amongst computer scientists and technically inclined philosophers over the epistemological status of 'knowledge' versus 'data', much of which is an inheritance of efforts to create 'expert systems' after the first 'AI winter'. See, for instance, Edward A. Feigenbaum, 'Themes and case studies of knowledge engineering', in Donald Michie (ed.), *Expert Systems in the Micro-electronic Age*, Edinburgh: Edinburgh University Press, 1979, pp. 3–21; Pamela McCorduck, *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence*, San Francisco: W.H. Freeman, 1979. *OMIM* was itself described as an expert system. Daniel F. Schorderet, 'Using OMIM (On-Line Mendelian Inheritance in Man) as an expert system in medical genetics', *American Journal of Medical Genetics* (1 June 1991) 39(3), pp. 278–84. For a discussion of efforts toward 'knowledge engineering' within the biomedical sciences during the 1960s see Strasser, op. cit. (1), p. 165. For broader philosophical and historical discussion of these issues see Sabina Leonelli, *Data-Centric Biology: A Philosophical Study*, Chicago: The University of Chicago Press, 2016; Matthew L. Jones, 'How we became instrumentalists (again): data positivism since World War II', *Historical Studies in the Natural Sciences* (1 November 2018) 48(5), pp. 673–84.

[98] John Burn, 30 November 1994, VAM, Box 511334, Folder '1995'.

https://doi.org/10.1017/S0007087421000224 Published online by Cambridge University Press

Nonetheless, despite its authoritative reputation, *OMIM* continues to parry threats to its existence. Upon visiting omim.org at present, one is greeted with a pop-up urging private donations to sustain the resource, now back at Hopkins. A number of medical writers are still employed to update this knowledge without publication credits. The value of such translational work is only increasing; biotechnology companies continue to invest in it as they grow their consumer bases. As such, *OMIM* relies increasingly on licensing fees from biotechnology platforms that make use of its application program interface (API) to integrate *OMIM* information. With the resource becoming increasingly marginal, we can see a shift in this synthetic and translational work so central to modern biomedicine toward the private sector.[99]

The quality of information available on omim.org makes it one of the best civilian science resources in biomedicine. What would it look like if the biotechnology companies that drew on it were required to make meaningful intellectual contributions? How would grants look if the definition of publications and original research were broadened to include the kind of synthetic work that could only be taken seriously from someone as creditworthy as McKusick? These questions exceed the boundaries of my study, but in emphasizing the connections between genetics, clinical research, capital and information technology, I want to suggest that despite *OMIM*'s 'legacy' status, this kind of reference infrastructure can hardly be a one-person job, and should neither be relegated to charity nor left for the market to determine access.[100]

Acknowledgments. I would like to thank Andy Harrison and Phoebe Evans Letocha at the Alan Mason Chesney Medical Archives of the Johns Hopkins Medical Institutions for their assistance with the Victor McKusick papers, and to the staff of *OMIM*, in particular Joanna Amberger, for their insight. The careful guidance of Angela N.H. Creager was essential at every stage of this paper, inspired by Nick Hopwood's interest in a piece of an earlier project he helped shepherd into being. Participants in the Learning by the Book conference held at Princeton in June 2018 shaped its current iteration, as did insightful feedback from the editor, two anonymous reviewers, Andy Hogan, Alex Csiszar, Erika Milam and her pandemic 'lab group'. Hospitality and intellectual friendship in and around Baltimore made it all the more worthwhile, so thanks in particular to Janel Jin, Ding Xuan Ng, Julie Barzilay, Jon Phillips, Ayah Nuriddin and Heidi Morefield.

[99] On the neoliberalization of biomedical research see, inter alia, Kaushik Sunder Rajan, *Biocapital: The Constitution of Postgenomic Life*, Durham, NC: Duke University Press, 2006; Doogab Yi, *The Recombinant University: Genetic Engineering and the Emergence of Stanford Biotechnology*, Chicago: The University of Chicago Press, 2015; Strasser, op. cit. (1).

[100] For more on databases and the social and economic relationships they entail see Emmanuel Didier, 'Open-access genomic databases: a profit-making tool?', *Historical Studies in the Natural Sciences* (1 November 2018) 48(5), pp. 659–72.

Cite this article: McGovern MF (2021). Genes go digital: *Mendelian Inheritance in Man* and the genealogy of electronic publishing in biomedicine. *The British Journal for the History of Science* **54**, 213–231. https://doi.org/10.1017/S0007087421000224