Archimedes

NEW STUDIES IN THE HISTORY AND PHILOSOPHY OF SCIENCE AND TECHNOLOGY

VOLUME 30

Archimedes has three fundamental goals; to further the integration of the histories of science and technology with one another: to investigate the technical, social and practical histories of specific developments in science and technology; and finally, where possible and desirable, to bring the histories of science and technology into closer contact with the philosophy of science. To these ends, each volume will have its own theme and title and will be planned by one or more members of the Advisory Board in consultation with the editor. Although the volumes have specific themes, the series itself will not be limited to one or even to a few particular areas. Its subjects include any of the sciences, ranging from biology through physics, all aspects of technology, broadly construed, as well as historically-engaged philosophy of science or technology. Taken as a whole, *Archimedes* will be of interest to historians, philosophers, and scientists, as well as to those in business and industry who seek to understand how science and industry have come to be so strongly linked.

For further volumes:
http://www.springer.com/series/5644

Jed Z. Buchwald
Editor

A Master of Science History

Essays in Honor of Charles Coulston Gillispie

Springer

Chapter 4
Charles Gillispie in the Digital Age

Jane Maienschein and Manfred D. Laubichler

> *There are three principal means of acquiring knowledge available to us: observation of nature, reflection, and experimentation. Observation collects facts; reflection combines them; experimentation verifies the result of that combination. Our observation of nature must be diligent, our reflection profound, and our experiments exact. We rarely see these three means combined; and for this reason, creative geniuses are not common.*
> *Denis Diderot*

Denis Diderot was in more than one way a predecessor of Charles Gillispie, and this widely quoted passage does not, of course, apply only to knowledge about nature. It also describes how we acquire our understanding of history. Given the enormity of the world and the almost endless wealth of observations, it is beyond the reach of any single scholar to achieve anything close to a complete overview. Therefore, "profound reflection" and "exact experimentation" are all the more important. And so, we might add, are the organized and guided efforts of large communities of scholars and scientists working towards a common goal. Only then can we even approach collective "creative genius."

Charles Coulston Gillispie has been a worthy successor to Diderot on all these counts. A half century ago, Princeton University Press published *The Edge of Objectivity. An Essay in the History of Scientific Ideas.* This volume cost $7.50 in hardback at the Yale Co-op, where one of us (JM) bought it in 1970 for undergraduate courses taught by historians of science Larry Holmes and Martin Klein. Considered a standard for the sort of survey courses in the humanities for which it was written, the book was also on the reading list for JM's graduate study at Indiana University in 1972. Two decades later, it was still one of the first books that ML (then a biology graduate student at Yale) purchased in the same bookstore and read in what became a five year intensive

J. Maienschein (✉)
Center for Biology and Society, Arizona State University, Tempe, AZ 85287-4501, USA

Marine Biological Laboratory, Woods Hole, MA 02543, USA
e-mail: Maienschein@asu.edu

tutorial in history of science with Larry Holmes and he encountered it, as well as its author, again when he began his graduate studies in history of science at Princeton. The book remains in print, though in 2011 costing $46.95 for the paperback version.

How can one author's book that spans the history of ideas in science from Aristotle into the 20th century remain relevant for so long? And what might we do today with such a project? Our contention is that the lasting value of the approach and the book lies precisely in its focus on ideas and the fact that it is the product of "profound reflection" and "diligent observation" and that the way it probed the transformations of and connections between ideas represents a form of "exact experimentation." But, like his predecessor Diderot, Gillispie realized that such an endeavor can only be sustainable in the long run if it also involves a whole community of scholars (or a network in 21st century-speak). Today such a community would include the kind of digital and computational approaches that some of us, including his long-time Princeton colleague Robert Darnton, have begun to champion (Darnton, Robert, 2008; 2010). In fact, Gillispie's own editing of the marvelous *Dictionary of Scientific Biography* offers an early hint at what is now becoming possible with digital publishing, the "Web 2.0" and beyond.

While Gillispie wrote other books, most notably the magisterial two volumes on *Science and Polity in France*, it is *The Edge of Objectivity* that captures the breadth of ideas as they develop through centuries of individuals, institutions, and intellectual traditions (Diderot, Denis, 1753; Gillispie, Charles Coulston, 1960; 1980). And while some reviewers at the time, and others since, have found things with which to disagree, nearly all have agreed with A. Rupert Hall's review in *Isis*. After fussing about various examples, yet admiring the product as a whole, Hall concluded that "There are in the book perhaps some of the defects as well as the virtues of intellectual brilliance; either way it is an intellectual exercise to read it. And it has much that any historian of science would be justly proud to have written" (Hall, 1960). Indeed, any historian would have been proud to have written such a book a half century ago. We believe that today it is no longer possible to accomplish such a task and that it would be foolish to try. There is just too much scholarship to master, too many additional original sources to consider, and too many enticing new scholarly tools that offer other ways of working. Diligent collection of data has increased dramatically and, in order to accomplish profound refection today we have to explore new forms of exact experimentation enabled by digital and computational approaches, even in the traditional fields of humanistic scholarship.

Diderot and the Enlightenment encyclopedists already recognized that some projects call for collaborations. So did Gillispie with the massive project that emerged over the decade 1970–1980 as the sixteen volumes of the *Dictionary of Scientific Biography*. A considerable army of historians wrote articles that were typically in the five-page range, with a few notable lengthy entries as well. These volumes, plus a two volume Supplement added in 1990, have served as a standard reference source since. The *New Dictionary of Scientific Biography*, edited by Noretta Koertge and published in 2007 included new entries and a set

of updated biographies for major figures. That the project has had such staying power and called for an update demonstrates the value of the original project, as well as its capacity for revision and extension. Charles Scribner's and Sons, with funding from the American Council for Learned Societies, helped make the project possible. Again, a large network of contributors contributed their scholarly interpretations, and the work has been recognized as important with awards and enthusiastic reviews.

Scribner's decided to publish the 2007 *New Dictionary* in print, and also with an on-line option that has a comprehensive index. However, what we can see from the vantage point of 2011 is that we need additional new digital and computational ways of working in order to capture the full advantage of the broad vision that Gillispie offered in both *The Edge of Objectivity* and the *DSB*. Digital and computational approaches begin to offer a vast array of new tools that can help with traditional forms of scholarly pursuits—the modern day descendants of *The Edge of Objectivity*, as well as enable new kinds of scholarship that transcend individual projects. We imagine Charles Gillispie as amused and fascinated by such possibilities. And we hope he agrees that it is a new kind of work about which reviewers will say "And it has much that any historian of science would be justly proud to have written."

4.1 Digital History and Philosophy of Science: The Heritage of Charles Coulston Gillispie

In our subsequent discussions we link history and philosophy of science because they are inseparable dimensions of understanding the transformations of our knowledge of nature and the world. It is also a tribute to the original conception that inspired the formation of the History and Philosophy Program at Princeton University, one of the many ways Gillispie has shaped the intellectual landscape of the last decades. Digital History and Philosophy of Science (HPS) has three dimensions based on the kind of digital objects involved and the computational methods used: (1) found objects, that is the published and archived objects of traditional scholarship available in digital form; (2) new publications written specifically for a digital publication environment that include work that is traditionally scholarly but also provides digital links to other work and sources in ways that make each new object part of a growing complex network of digitally available interconnected objects (texts, images, maps, etc.); (3) newly discovered findings that result from use of new tools, including found relationships among items not known to be linked. This requires the use of computational tools and repositories and represents the truly transformative aspects of Digital HPS. We look at each of these and at some sample efforts to develop each.

1. Found Objects would typically start with existing publications and expand from there. Take Gillispie's *Edge of Objectivity*. Right away he introduces

particular *people*, including Galileo, Paolo Sarpi, Machiavelli, Leonardo, Newton, and Einstein—all in the first ten pages. Then there are particular *works* by the individuals. People work in *places*, so we read also of *institutions* such as Oxford, and of *contextual influences* such as the Roman Catholic Church. He introduces algebra, logic, and eventually the other *fields* of science. These are all relatively defined units that can be linked throughout the work and connected with other uses of the same. We can flag such key words and phrases, to be linked with others directly and through more complex relationships (as discussed further below).

Interpretations are harder. There are *ideas* and *concepts* such as that of "objectivity" that run throughout. And Gillispie adds his own interpretation, saying that modern science "is impersonal and objective. It takes its starting points outside the mind in nature and winnows observations of events which is gathers under concepts, to be expressed mathematically if possible and tested experientially by their success in predicting new events and suggesting new concepts. Modern science has not abandoned rationality, but it is first of all metrical and experiential." (p. 10)

Further, Gillispie recognized that science is not something automatic running along on its own, but rather that it involves interpretation by human actors to develop the science. And these actors work in social contexts for particular purposes, and therefore exercise judgment about the use of science in society. That is, "the influence of science is not simply comfortable. For neither in public nor in private life can science establish an ethic. It tells us what we can do, never what we should." (p. 154) That comes from society, and the result changes in different societies.

Capturing such complex interpretations and understanding of the *contexts* digitally is harder, but not impossible. New scholarly tools include annotation approaches to include abstract *concepts* and *interpretations*, linked to the authors, and such tools are providing considerable promise for complex interpretive linking that yield new scholarly findings and make possible new scholarly interpretations because of the new ability to visualize connections.

Gillispie's own *Edge* does not include illustrations, but a 21st century project easily could, if a collaboratively collected database of such images existed in a stable repository with shared protocols and standards for documentation and archiving of metadata to facilitate shared use. The objects in question would each need to have adequate and appropriate metadata identifiers, but standards already exist for such work.

And what if a shared repository included biographies and descriptions of the people, places, institutions, contexts, ideas, concepts, and interpretations. Plus published work included in the references and mentioned in the discussion. If all these objects were available in the same repository, what glorious fun it would be to read the wide-sweeping scholarly interpretation. Instead of being an ill-informed undergraduate student struggling to connect the many pieces and not knowing much about each, the reader could look up additional information, find pictures, see the original texts mentioned, and so on. And this imagined

research methodology does not require that all the scholars and all the owners of scholarly materials put them in the same physical space. Certainly not. The repository can be distributed in as many places as desired. What is necessary is shared vocabularies so that different word usages are linked, and shared metadata and descriptors so that difference objects are found.

The nucleus of such a distributed digital repository is already emerging as individual projects within HPS have begun to collaborate, standardize their practices and develop the necessary tools together. These projects—including the Newton projects at Indiana and Sussex, the Embryo Project at Arizona State and the Marine Biological Laboratory (MBL), the Einstein Papers project at Caltech, the History of Quantum Mechanics Project and the many digital projects connected to the Max Planck Institute for the History of Science in Berlin as well as others—have formed a consortium that coordinates these activities and also lobbies for this approach. So far the Marine Biological Laboratory in Woods Hole and the Max Planck Institute for the History of Science in Berlin have stepped up to become the kind of partner that Charles Scribner was to Gillispie's *DSB* project. These two institutions act as hosts for repositories, clearing houses for tools, incubators for projects and, as of Spring 2011 also as an educational and training center. The MBL, with it long tradition of advanced summer courses, is beginning to offer a course in computational methods for digital HPS.

This is being done, and we can and should do much, much more to share resources. If we took all the money that all the scholars have used to visit collections dispersed around the world, and used them to make materials available to everybody, that would be a fine start.

2. New Publications can take traditional forms, of course, and appear in proprietary journals. But we can hope that scholars will increasingly come to feel able to contribute to open access publications, where a work can include links to all the found objects used to generate the results. This is starting to happen. And a new publishing project at the Max Planck Institute for the History of Science in Berlin (MPI)—The Max Planck Research Library for the History and Development of Knowledge—is taking these possibilities in new directions. The goal here is to publish annotated editions of sources and translations as well as edited collections and workshop volumes in an open access environment while maintaining all the standards of peer review. However, historians of science who appreciate the tactile experience of a book need not worry. While all the works are available in an open access digital format, and thus also accessible to scholars from less prosperous countries and universities, they can also be acquired as well produced bound volumes. Print-on-demand technology has developed to such a degree that this is now possible—these volumes can even be ordered on amazon.com.

Digital publications can also take the form of short entries that help to link across a wide range of other contributions. This is an approach we are developing with the Embryo Project Encyclopedia (http://www.embryo.asu.edu),

not unlike that of Wikipedia and other similar projects that take the same basic approach (see Laubichler et al., 2009; Maienschein and Laubichler, 2010). The idea here is to provide the complement to Gillispie's breadth and interpretative scope, his "profound reflection." The focus of these projects is a further development of the encyclopedic vision; to take big ideas and complex developments and "atomize" them. Break the big into component pieces, and provide short descriptive articles. What is different from the traditional multi-volume encyclopedia is that this approach is open-ended, that the links and cross-references are easy to follow and can grow dynamically and that interpretative texts can directly link to sources. Such a structure becomes increasingly necessary as the amount of available information begins to threaten its use. Such (short) interpretative essays then serve as the glue that provides links among objects that might not be linked otherwise.

3. Newly discovered findings also can result from the application of new computational tools and methods. For example, such tools enable us to establish relationships among items not known by any one scholar to be linked. It is the small world paradigm of network theory applied to the understanding of the history of science or for Princeton insiders, it is a very complex application of multiple Erdös numbers. To begin with, annotation tools can be used to search texts and other documents and harvest information that is then coded in the form of relationships. Of particular interest are so-called RDF triples, which allow connections such as "Newton was a critic of Hooke" or "Harvey studied at Padua." While some naïve researchers assume that they can find such relationships with tools like Google, just try it yourself and you will immediately see why that cannot work. Only when the exact words appear together will they be found. But sophisticated scholarly annotation tools can let the researcher discover connections. Then these can be captured and stored as new scholarly findings for other scholars to use. To be even more useful, the triples can also be given contextualizing information as well, though tools for such "quadruples" are still in development.

To illustrate how computational methods and approaches can change the practice and methodological foundations of scholarship, we briefly describe the change from traditional interpretive history to a new form of scientific history. What are the methods that allow for detailed historical analysis with data driven computational approaches, what we are calling scientific history? It requires starting with *defining basic informational units* for complex historical processes—what we call a functional relational unit (see below) and *storing* these "atoms of information" in a *centralized repository*, as well as *developing tools that enable queries that can retrieve more information and links than an individual scholar has access to* in his university or research institute. Resulting search outcomes can then reveal aspects of historical developments that were not visible initially: not just more information, but new kinds of information.

The structure of functional-relational units as statements linking two objects creates a highly connected network of knowledge, which, in turn, leads to multiple kinds of synthetic knowledge units that would not have been obvious without the linkages enabled by computational annotation tools and storing information units within a centralized repository.

This means that we have to define historical statements as data so that they can be analyzed with computational methods, and that also requires social cooperation as scholars see the value of this new approach and contribute data. Traditionally, historians have shared their interpretations of data but not the data themselves, which can be found in their notebooks and private annotations of sources. This has made it very difficult and time-consuming to reanalyze most historical works. But, as changed practices requiring that data be shared together with published analyses have transformed the sciences and made them more effective, we argue that scientific history will have a similar transformative effect on understanding of the history and nature of science.

Focusing on textual information we need to extract basic informational statements as well as the relevant contextual information. Our "atom of information" is a functional relational unit that connects two objects through a relevant functional relation; for example "Charles Darwin travelled on *HMS Beagle*." Without additional contextual information such a statement has limited use. We therefore need to be able to connect this simple statement with more layers of contextual information, such as metadata of the text from it was extracted, information about who created the statement (a trusted or less trusted source, for instance, or a machine), and additional layers of information that represent the larger historical context (such as 19th century Britain, or history of biology, the history of exploration). The structure of the statement as well as the layers of contextual information allow us to reconstruct complex networks of events, such as where Darwin's journey took him, who else was on board the HMS Beagle, what publication refer to the *Beagle*, additional information about places Darwin visited, animals he collected and what we know about them today, who he wrote to on his trip, and what the British government felt they got out of their investment in the *Beagle*, and so on.

Now, these are all different sorts of information that require looking in many different places and relying on a certain amount of chance to find and link the "atoms" of knowledge. Historians of science have perfected this approach as part of their craft. However, no one can know it all. Furthermore, the situation is much harder if we are looking at a complex large-scale 20th century science project rather than somebody famous and well-documented like Darwin. For these cases, a data driven social-computational and collaborative approach that is the core of the third dimension of digital HPS is essential.

4.2 Conclusion

Scholarship is changing, and this is good. Gillispie wrote at the end of *Edge* that "History is made by men, not by causes or forces, and I have tried to write with due attention to the intellectual personalities who have borne the battle, and not without sympathy for its casualties. And though I have written as closely to the texts as my competence permits, I want the tale to move unencumbered by the barnacles of scholarly apparatus." (p. 521). In his *Isis* review, Hall reflected that he would have wished for a few more such barnacles to guide his reading. Yet we have appreciated reading without all the footnotes and side trails considered necessary by historians such as Hall himself.

Nonetheless, the new digital HPS can make everybody happy. By including all the barnacles, which are anchored in their own carefully documented archives or libraries and storage facilities but also freed through digital linkages, we have both. Narratives with their own personalities and diversity of directions, yes, and also all the scholarly apparatus. This apparatus can reassure other scholars, but how much more effective that reassurance is when we can all read the same documents and study them, complete with annotations and links to other materials and sources.

Diderot argued that the three means of acquiring knowledge are rarely combined, but if they are we are confronted with a creative genius. Few individual have reached this level. But as cultural evolution progresses, the locus for a creative genius increasingly shifts towards a well connected network of individuals each making their own contributions, but doing so as parts of a whole that is most certainly bigger than the sum of its parts.

Acknowledgment Thanks to the National Science Foundation for support through multiple grants.

References

Darnton, Robert. 2008. "The Library in the New Age." *New York Review of Books*. June 12, 2008.

Darnton, Robert. 2010. "Can we create a national digital library?" *New York Review of Books*. October 28, 2010.

Dictionary of Scientific Biography. 1970–1980. *Supplement*. 1981, edited by Charles Coulston Gillispie; Gillispie and F. L. Holmes et al. New series 2007, edited by Noretta Koertge. New York: Scribner.

Diderot, Denis. 1753. Quoted from *On the Interpretation of Nature, no. 15* in Lester G. Crocker, editor and Derek Coltman translator. 1966. *Diderot's Selected Writings*. New York: Macmillan.

Gillispie, Charles Coulston. 1960. *The Edge of Objectivity. An Essay in the History of Scientific Ideas*. Princeton: Princeton University Press.

Gillispie, Charles Coulston. 1980. *Science and Polity in France Science and Polity in France. The End of the Old Regime*. Princeton: Princeton University Press; 2004. *Science and Polity in France: The Revolutionary and Napoleonic Years*. Princeton: Princeton University Press.

Hall, A. Rupert. 1960. Review of *Edge*. *Isis* 51: 344–347.

Laubichler, Manfred, Jane Maienschein, and Grant Yamashita. 2009. "The Embryo Project and the Emergence of a Digital Infrastructure for History and Philosophy of Science." *Annals of the History and Philosophy of Biology* 12: 79–96.

Maienschein, Jane and Manfred Laubichler. 2010. "The Embryo Project: An Integrated Approach to History, Practices, and Social Contexts of Embryo Research" *Journal of the History of Biology* 43: 1–16.