Computational Perspectives in the History of Science

To the Memory of Peter Damerow

By Manfred D. Laubichler, Jane Maienschein,** and Jürgen Renn****

ABSTRACT

Computational methods and perspectives can transform the history of science by enabling the pursuit of novel types of questions, dramatically expanding the scale of analysis (geographically and temporally), and offering novel forms of publication that greatly enhance access and transparency. This essay presents a brief summary of a computational research system for the history of science, discussing its implications for research, education, and publication practices and its connections to the open-access movement and similar transformations in the natural and social sciences that emphasize big data. It also argues that computational approaches help to reconnect the history of science to individual scientific disciplines.

THE NATURAL SCIENCES, as well as the social sciences and increasingly also the humanities, are generally acknowledged to be in the midst of a "computational turn."[1] This transformation has profound implications that have already affected all aspects of the research enterprise, from the availability of data to methods of analysis, as well as broader research, publishing, and educational practices. The computational revolution is shaping the way younger generations conceptualize and approach problems. A recent report of the

* School of Life Sciences, Arizona State University, Tempe, Arizona 85287-4501; Marine Biological Laboratory, Woods Hole, Massachusetts 02543; and Santa Fe Institute, Santa Fe, New Mexico 87501.

** School of Life Sciences, Arizona State University, Tempe, Arizona 85287-4501; and Marine Biological Laboratory, Woods Hole, Massachusetts 02543.

*** Max Planck Institute for the History of Science, Boltzmannstrasse 22, 14195 Berlin, Germany.

Our friend and colleague Peter Damerow (1939–2012) was a pioneer in adopting new perspectives and methods within the history of science and one of the earliest advocates of computational approaches and open access. He has been an inspiration to all of us. See http://damerow.mpiwg.de/doku.php/obituary for more information about Peter's work and influence.

[1] David B. Searls, "The Roots of Bioinformatics," *PLoS Computational Biology*, 2010, *6*(6):e1000809, doi:10.1371/journal.pcbi.1000809. See also http://www.thecomputationalturn.com/ and the references therein; all web pages cited in this essay were accessed on 24 Aug. 2012.

Isis, 2013, 104:119–130

0021-1753/2013/10401-0008$10.00

National Research Council of the United States identifies the most transformative impacts of the "Information Age" as those connected to "computational thinking."[2]

This new mode of thinking is not only becoming the accepted way to produce knowledge, especially among younger generations, but it is seen as inevitable, necessary, and good. The standard working definition is: "Computational Thinking is the thought processes involved in formulating problems and their solutions so that the solutions are represented in a form that can be effectively carried out by an information-processing agent."[3] This definition is not very useful by itself, since it could include most anything we do. But the intention is clearly to shift the effort from the work of individuals to networks of researchers aided by machine learning and analysis. An information-processing agent can be a computer, a human, or any combination of the two. We urge historians of science to embrace this new approach, connecting the new with results of traditional scholarship. We agree that the changes are inevitable, necessary, and also good for the field. And we note that our training of scholars needs to change to be ready for the new computational history of science. Here we will develop a vision of what is possible with new computational methods. A lot of what we say is programmatic, but it is not just fantasy or wishful thinking. Together with a network of partners, we have already begun to develop many of the necessary elements of a computational infrastructure or research system for the history of science.[4]

A short version of our vision can be stated as follows: Computational history of science introduces big data–based approaches and computational analytical methods. These allow for new types of questions, such as *longue durée* assessments of the evolution of knowledge on a global scale, detailed analysis of the nature of scientific transformations and revolutions, and comparisons across wide temporal and spatial scales. Computational methods allow for automated data extraction, data and text mining, network and other types of visualization, statistical analysis, and causal modeling (such as agent-based models). Computational history of science is built on a foundation of openly accessible data (texts, images, quantitative and geographic information, but also the results of computational analyses) stored in a network of interoperable repositories. And it involves new venues for publication that are open access, connect data/sources with interpretations, and involve a wide array of digital formats, including exhibits and editions, short encyclopedic texts that act as guides to sources, and interactive formats that allow for commentary and discussion of sources and interpretations. Computational history of science adds more transparency to the field by providing easy and direct access to the data and sources of interpretations and introduces a more scientific culture of continuous and recursive testing and refining of hypotheses. It connects the history of science to multiple user communities, from educational contexts to scientific discussions. Finally, in the spirit of the pluralistic approaches to the history of science advocated in all the contributions to this Focus section: computational history of science is one approach among many, but it can also act as the glue connecting all these different perspectives in an integrative way.

Of course, we are well advised not simply to accept the call for particular innovations

[2] "Report of a Workshop on the Scope and Nature of Computational Thinking," National Research Council, 2010; see http://www.nap.edu/openbook.php?record_id=12840.

[3] For more information see the Center for Computational Thinking at Carnegie Mellon University: http://www.cs.cmu.edu/~CompThink/. See also Jeannette M. Wing, "Computational Thinking," *Communications of the ACM*, 2006, *49*(3):33–35 (from which the definition is taken); and Jan Cuny, Larry Snyder, and Wing, "Demystifying Computational Thinking for Non–Computer Scientists," work in progress, 2010.

[4] See http://www.digitalhps.org for a list of partners in this area.

without careful evaluation of what makes sense. While we do not have space here to examine fully all the implications of computational history of science, we explore some practical consequences and provide some examples. We introduce elements of a computationally based research system for the field and discuss how it can facilitate new perspectives and practices that will complement existing methods of scholarship and allow us to go beyond what we can do now and even to carry out scholarly work that we have only begun to imagine.

WHAT IS COMPUTATIONAL HISTORY OF SCIENCE?

For many scholars, computational approaches in the history of science most immediately offer ways to do familiar kinds of work more efficiently, more effectively, and on a broader scale. But as we noted in the short vision statement with which we opened this essay, they also enable novel and different kinds of work with additional kinds of data, asking different questions and achieving quite different kinds of insights that are presented in a genuinely open-access environment. And they further advance an ongoing change in the way history of science research is conducted, a move away from individual scholars' contributions to networks of multidisciplinary teams distributed across the globe.

As with similar transformations in the life sciences (another fundamentally historical field), the starting point of computational approaches is big data.[5] A limit on research capacity to date has been the individual scholar's ability to master many different sources and weave them into a well-crafted narrative. Collaborative projects are one way to transcend this limit. These are, however, greatly aided by computational approaches that facilitate working in real time with more and different kinds of data, compiled in ways that are interoperable and openly accessible.

Such a computationally based research system is by its very nature not limited by size or type of data, as long as data are in a computable form. The operative term here is "computable form."[6] For historical data, this means brought into standardized formats with shared metadata and ontologies for categorizing and connecting disparate historical materials. Or, at the least, we need clearly defined mapping rules for relating different standards when the data are structured and stored within different and dispersed databases and repositories. This requires a change in the way scholars work. Instead of collecting things for our own use (and referencing them in footnotes), we need to learn to share and make our data available to others through effective data management.

Such an approach, beyond adding to the breadth of what each study does, allows a broader comparison drawing on materials beyond what any one author or editor knows, engaging materials across a global scale in a way that locally constrained histories rarely do. Furthermore, it makes possible the exploration of different kinds of questions that are too big or too complex to address through individual studies. In the future, this approach should take us to questions about the evolution and the multiple contexts of knowledge that we can't now even imagine would be answerable.

A primary function of all database-based projects is aggregation of different types of

[5] On big data and "big science" see http://www.nsf.gov/news/news_summ.jsp?cntn_id=124398; and http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504767. A broad overview can be found in the McKinsey Global Institute's "Big Data: The Next Frontier for Innovation, Competition, and Productivity": http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation.

[6] See, e.g., Stephen Wolfram on this subject: http://www.stephenwolfram.com/publications/recent/datasummit2010/.

data, revealing networks of connections and enabling comparison across far more types of data than would otherwise be possible. In the life sciences, the various "-omics" projects linked with detailed annotations are a good illustration of the possibilities of this approach. In the history of science, these networks can be of people, materials, texts, and intertextual comparisons of published or archival materials. They can come from different cultures and different time periods and (eventually, with more tools facilitating translation) different languages. These networks can be extended to allow integration of data from the history of science and technology with data from military, economic, environmental, art, gender, medical, demographic, or any other kind of database in order to reveal patterns of interactions and causal relationships.

Of course, the ability to carry out such analyses of larger-scale patterns depends on the size of available databases and repositories. With computational approaches, the history of science can take advantage of economies of scale, since there are far more large-scale projects already operating in relevant areas of the social sciences and the humanities (such as demographic, military, climate, or local history) than there are as yet in the digital history of science. Right now, the balance of up-front investment of effort in technology development and data generation to resulting novel insights is still high for the history of science, but this balance is changing as more and more data from different projects are entered in shared or interconnected repositories. As a relative late adopter of computational approaches, the history of science can benefit enormously from existing methods and concepts that can be adapted for the specific needs of the field. The most promising link here is to computational biology—not just on the methodological side, but also conceptually, as biologists have learned how to analyze complex, historically evolving systems. Computational methods thus also contribute to methodological discussions within the history of knowledge that connect it to perspectives of cultural evolution and complex systems.[7]

A central question for data-driven computational approaches concerns what counts as data. The history of science uses a wide range of sources, from texts to artifacts, from many languages. All of these have to be represented by standardized data models. We need to change the way we work to capture our traditional information as shared data, just as biologists had to learn to work with GenBank and related databases to make their data useful to more than just one lab at a time. One new type of data with far-reaching computational implications for historical studies is the contextualized relational statement, or the contextualized RDF triple (RDF stands for Resource Description Framework) or quadruple.[8] This type of data should be intuitive to all historians, since it formalizes some of the most basic analytical methods in the analysis of textual sources.

A triple is a proposition that connects two objects with a relationship term, such as "A was a student of B," and a quadruple might add a date or any other metadata information. At present there is no straightforward way to do a Google-type search to ask "Who was a student of James Watson?" or "Who has written on the Rutherford atom?" If all scholars add the relationships "was student of" or "has written on" as triples to a shared database, then its collective value grows tremendously and allows us to discover new information that we did not know. And in the (very near) future, machine-learning tools will discover

[7] For two representative accounts see R. Boyd and P. J. Richerson, *Culture and the Evolutionary Process* (Chicago: Univ. Chicago Press, 1985); and A. Mesoudi, *Cultural Evolution: How Darwinian Theory Can Explain Human Culture and Synthesize the Social Sciences* (Chicago: Univ. Chicago Press, 2011).

[8] See http://www.w3.org/TR/rdf-concepts/; and http://linkeddata.org/.

and add such entries so that they do not have to be entered by hand— once scholars have indicated that a vocabulary that includes "was student of" is something they want.[9]

A CASE: HISTORY OF MALARIA RESEARCH

It should by now be clear that we see computational history of science not just as a more efficient way of conducting business as usual but, rather, as a qualitatively different approach to the study of the history of knowledge. Here we will demonstrate some of the promise of this approach by first showing how computational methods can rediscover the insights of well-established scholarship and then pointing to some of the additional possibilities it affords. Space constraints require us to be impressionistic, but we are confident that we can convey a sense of the new perspectives that can be gained by these methods. One perennial goal in the history of science is to identify points of inflection that led to substantial change in knowledge about the world. Can computational approaches help to recognize the moments of change and identify causal factors in ways that go beyond what we can do with the usual historical tools?

To make this less abstract, take a familiar example where the data and interpretations are well known: the case of understanding malaria—what characterizes the disease, its causes and treatments. A number of detailed historical accounts have investigated how research in tropical medicine and protozoology transformed knowledge about the causes and etiology of malaria. Over decades, individual scholars have uncovered details, which have led to several aggregated historical overviews.[10] In short, we know a great deal about localized details of the history of science, medicine, and society as they relate to malaria. One interesting pattern that emerged in these studies is that a series of important discoveries about the causative agent (a protozoan parasite), pertaining to the life cycle of the parasite and its carriers and the distribution of different types of parasites and disease forms, led to dramatic changes in the understanding of the etiology and biology of malaria between 1880 and 1910. At the same time, treatments and prevention efforts continued to be informed largely by common sense and traditional approaches.

Given these known facts, our first question to test the value of computational approaches was, Can we reveal the same insights through computational analysis of sources? We were able to test the power of computational methods in the case of the history of malaria because many of the relevant historical sources have been digitized and are available in searchable text files as part of the Biodiversity Heritage Library.[11] Applying our computational tools and extracting terms and relations from multiple sources, we have visualized networks of those terms connected to the concept of malaria at different periods over the last two hundred years. We have extracted functional relationships (triples) between the concept of malaria and a number of categories representing knowledge or ideas about the causes and etiology of the disease, its treatments, modes of prevention, geographic distribution, and so forth. We organized these into networks, produced a set of

[9] Machine-learning tools already work well for highly structured texts, and there is tremendous progress in this area of computer science.

[10] Two recent overviews of the history of malaria are James L. A. Webb, Jr., *Humanity's Burden: A Global History of Malaria* (Cambridge: Cambridge Univ. Press, 2008); and Randall M. Packard, *The Making of a Tropical Disease: A Short History of Malaria* (Baltimore: Johns Hopkins Univ. Press, 2011).

[11] Many sources can be found at http://www.biodiversitylibrary.org. The Biodiversity Heritage Library, part of efforts to develop the Encyclopedia of Life (http://www.eol.org), intends to digitize and make available historical information about biodiversity.

graphs to visualize these network patterns (see Figure 1 for a sample from a small section of such a graph), and compared the resulting graphs both between and within different periods. Our observations largely confirmed what we already knew from the substantial amount of scholarship in this particular field, thus answering our first question about computational approaches affirmatively.

It might not seem especially exciting that we have confirmed things we already knew, but we are showing that the approach works—and, by implication (after more testing and calibration), that it can be applied to cases for which we do not have the results. The approach can be used for uncovering patterns that help show where to dig more deeply, for example. Analyzing this case computationally, we find that the networks of key terms used to discuss the disease of malaria differ substantially before and after important discoveries, such as the vector/mosquito-borne transmission of a protozoan parasite. This allows us to identify interesting patterns of correlation and also to carry out experimental approaches to test concrete causal models about drivers of scientific change.

So are we doing more than just rediscovering what good historical scholarship has uncovered before? Yes, because it was not necessary or predetermined that the results of this approach would yield the same patterns as those established over time by standard studies. Our approach confirmed what we predicted—and what had taken dozens of scholars literally decades to generate. Our exercise allows us to calibrate our research approach using known cases, in the same way as a new instrument is calibrated to ensure that its measures conform to well-established results or standards. If we had come out with different results, we would know that the system is not helpful or needs to be refined. But as it did work, the real power of computational history of science comes in the next step: applying these same methods to a large number of cases, many of which have not been analyzed historically, and then developing and testing specific causal models as explanations of scientific change and innovation that are less based on singular cases and more representative of the totality of scientific activity, both within any given period or research domain as well as across a wide range of fields and historical and geographical areas.

For example: as noted above, intensive studies in tropical medicine and protozoology greatly changed understandings of the causes and etiology of malaria. Yet for a long time after those discoveries were made discussions of treatment and prevention did not change much, if at all. This raises the question as to whether a similar pattern occurs in other tropical diseases or with infectious diseases more generally. While this would seem likely, we simply don't know for sure, as no one scholar—or even a team of scholars—has the time and resources to carry out studies as extensive as the traditional historians' studies of malaria. Even the whole community cannot study a large number of tropical diseases with the same degree of detail. Rather, the comparison and search for patterns requires analysis of a large amount of literature that can be done exhaustively only with the data mining and analysis of computational approaches.

Another question raised by our exploratory study concerns the ways in which knowledge spreads. What is the relation between the spread of theoretical and of practical knowledge? Which institutions, media, and forms of representation shape the diffusion of knowledge? How does the diffusion of knowledge compare with other diffusion processes and how can it be modeled? In what sense can one speak of a globalization of knowledge in history—that is, how was the world connected through knowledge in the past?[12]

[12] See the recently published volume on globalization as part of the Edition Open Access at the MPIWG: http://www.edition-open-access.de/studies/1/index.html.

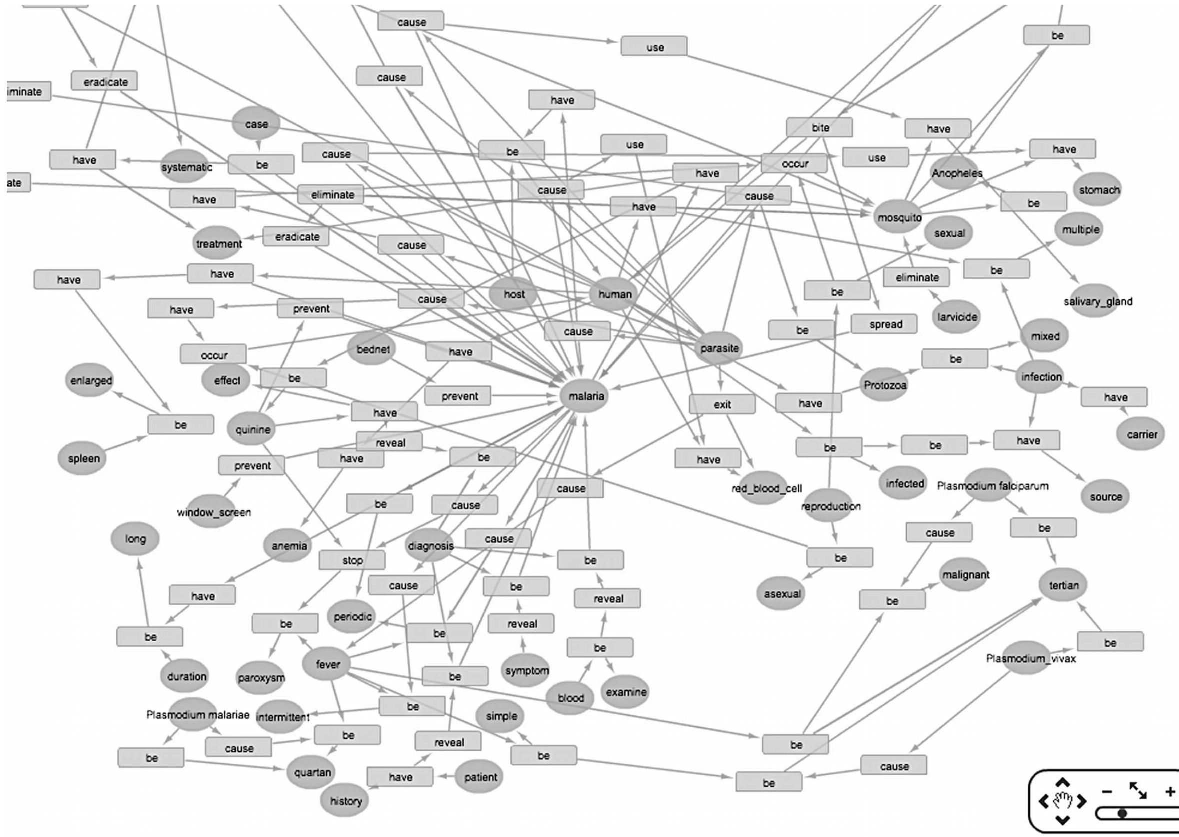

Figure 1. *A small set of extracted functional relationships organized as a network representing ideas about causes, symptoms, and treatments of malaria. Comparing this graph, extracted from S. P. James,* Malaria at Home and Abroad *(London: Bale & Danielsson, 1920), to other graphs (synchronically and diachronically) highlights the continuities and discontinuities in the development of knowledge about malaria.*

Again, large-scale data mining and computational analysis should be able to give us a larger and richer data set and allow more generalizable answers. It also introduces a more explicit and recursive form of hypothesis testing into the practice of the history of science, enabling us to test or model multiple scenarios with an ever-growing data set that also includes the results of earlier interpretations.

Getting to a point where the computational approaches are fully operational will, of course, take time and patience. Not all efforts will pay off fully or quickly. The pioneers who do the work now are setting up the research system for the future, helping to reveal what kinds of new tools we need and what kinds of new results we can anticipate. At the Marine Biological Laboratory we are developing the HPS Repository precisely to collect what we know now and also to push for new discoveries and tools. Similarly, at the Max Planck Institute for the History of Science in Berlin we are working on an innovative knowledge infrastructure constituting an interactive and integrative environment for multidisciplinary research and dissemination that will support the initiation of overarching research questions. It will, in particular, support real-time resource sharing and the managed accumulation of shared knowledge—for instance, by providing the capability of dynamically overlaying different data sets. By allowing for the creation and representation of complex dynamic knowledge structures, it will overcome the conventional distinction between research and publication, while continuing to ensure quality control through such means as scholarly apparatus and peer review. The realizability of knowledge integration through the implementation of dynamic structures in the electronic medium is illustrated by the successful transfer of traditional knowledge representation models such as encyclopedias and maps to the Web. Although the potential of such transfers is far from being exhausted (as is evident from the promise offered by electronic glossaries, virtual exhibitions, and electronic tools for persistent, flexible, and shared annotation), we also conceive of radically new knowledge representation models that can adequately capture the outcome of multidisciplinary research projects engaged in the study of complex processes such as the globalization of knowledge. The adequate representation of long-term globalization processes requires, for instance, the representation of layered time developments such as the changing relationships between technological, linguistic, and genetic diffusion processes in human populations.

Together with our network of partners, we consider this effort worthwhile because evidence suggests that it has been productive in related fields (to wit, the revolutionary transformations in the life sciences and increasingly also in the social sciences) and because it is already paying off in some of our own projects.

Once large parts of these computational processes are automated with machine-learning tools, the approach could perhaps even be used to monitor science and scholarship in other fields, including the history of science, in real time to help identify contributions that introduce something genuinely novel. In principle, we should be able to fine-tune computational methods in the context of well-understood historical cases and apply these methods to both additional historical cases and instances of current science. Such possibilities could extend the impact of and help make the case for additional grant support for history of science research from a broader set of communities.

Does everybody have to become a computational historian? No, there is room for traditional historical work by individuals who rely on small collections of data and store the results on their home computers or in file cabinets (or even in piles on the floor). Yet the decision not to embrace new methods does involve a choice—and one that has consequences. We argue that there are obligations, in particular, to share data. Everybody should have a data management plan (and one that goes beyond references to archives in footnotes), as the National Science Foundation demands in the United States. When scholars receive public funding for their work, and when they work at private institutions and therefore receive private support, they have responsibilities to their patrons. They collect data in the form of materials and interpretations. We firmly believe that all scholars must share their data by adding metadata and contributing to shared databases—most of which will be associated with libraries. Computational history of science calls for extending this practice to all data, and it is providing an infrastructure that will make it easy for individual researchers to contribute.

A RESEARCH SYSTEM FOR COMPUTATIONAL HISTORY OF SCIENCE

As mentioned, all these approaches are based on highly structured data stored in databases and repositories. Every database is built around what can be called atoms of information. For many biomedical questions the gene or the DNA sequence fulfills this purpose; other databases consist of key terms. What then, would an atom of information look like for the history of science? At the minimum, it would offer relations or propositions of the sort "A was a student of B" or "Result C was obtained by experimental method D." These relations are expressed in formalized triples, each stored in the repository alongside the information about A, B, C, and D. We are also interested in the source, as well as the generator/interpreter, of each of these triples, so we also add several layers of context through metadata.

Metadata include bibliographic and other information about the text/source and a profile of the researcher who added this relationship. More complex statements can be represented as nested triples. With such triples as the basic atoms of information, it is then possible to generate complex networks or graphs that capture what is of interest with regard to a specific question. As soon as basic data are stored in repositories they need to be analyzed, modeled, and, finally, molded into complex historical interpretations.

Traditionally, this interpretive process has been referred to as the "art or craft of the trained historian." It has led to some of the most outstanding products of literary expression. But the approach has never quite shaken off a degree of subjectivity. We are all aware how historiographic fashion changes and how some of the most convincing interpretations of earlier periods suddenly become outdated. Of course computational approaches are also governed by assumptions. But these are more openly accessible, and more adjustable, as computational history of science is based on explicit modeling of testable causal hypotheses, and, as the results of specific models become part of the data, it is a much more open-ended and dynamic process of constant reevaluation and recalibration. Historical interpretation itself then becomes a specific dynamic process governed by its own set of internal and external evaluation and selection criteria. Computational history of science is thus its very own complex adaptive and evolving system.

The kind of computational history of science proposed here requires a complex computational research system infrastructure, several layers of associated tools, and substantial transformations in the social organization and reward structures of the academic discipline. While these are daunting challenges, we already see encouraging signs of progress in all areas. A sizable community within the history of science, connected to informaticians in the sciences and digital humanities, forms an active network that is currently developing this research system in the context of concrete research projects. Here we sketch the overall design.[13]

One element we are building now, and for which the field needs more instances to compare, is a centralized repository and research system (the models here are centralized data repositories and associated tool layers in the life sciences, such as GenBank or the Human Cancer Genome Atlas). This will offer support and tools to individuals and smaller research projects in the history of science and allow them to benefit from the work of the larger community, giving them access to data and tools in our repository and its connected nodes. The system is being built at the Marine Biological Laboratory in collaboration with the Max Planck Institute for the History of Science and together with a network of partners. It consists of integrated layers to store and manage a wide array of digital objects from different projects in the history of science and related endeavors. Individual projects interact with the core repository through an ingest layer to add data, an access layer to display data and the results of analyses (customizable for a whole range of different users and audiences), an Application Programming Interface (API) layer that facilitates integration of components of the research system with other projects, and a tool layer that contains a whole suite of analytical, annotation, and visualization tools. Figure 2 shows the schematic architecture.

NOVEL FORMS OF PUBLICATION AND REWARD STRUCTURES

Computational and digital history of science brings new modes of publication. The digital humanities have moved in that direction, but many of the current digital publications and platforms still fall short of their possibilities. The Max Planck Institute for the History of Science, together with its partners, has recently launched a new open-access book series that combines the advantages of printed books with the potential of the electronic medium, for instance, to represent, search, and annotate sources in high-quality reproductions. The vision is to build a complete workflow from digital libraries such as the ECHO environment to scholarly publications in the form of a "Digital Scrapbook," supporting every step of the research process. This vision is being shared by some editors at major university presses, who are preparing to launch a novel publication venue following these principles.[14]

What we advocate here draws on our conception of computational history of science as a data-driven endeavor. Data—sources but also various types of extracted and aggregated data such as triples—are creative objects of publication, as are collections of such data. Recognizing this involves substantial changes to the values and practices of historical scholarship, and open-access needs add new challenges. Computational history of science is closely linked to the open-access movement that advocates free access to both the results and the sources of scholarship, and the research system calls for us to embrace new kinds of publication as deserving credit equal to what accrues from publication in traditional print journals.[15]

Publishing data and sources together with scholarly analyses is one way to add

[13] More about this community can be found at its website: http://www.digitalhps.org.

[14] http://www.edition-open-access.de (open-access book series); and http://echo.mpiwg-berlin.mpg.de (MPI vision). Among the university presses and editors involved are Harvard University Press (Michael Fischer) and MIT Press (Marguerite Avery).

[15] http://oa.mpg.de/lang/en-uk/berlin-prozess/berliner-erklarung/.

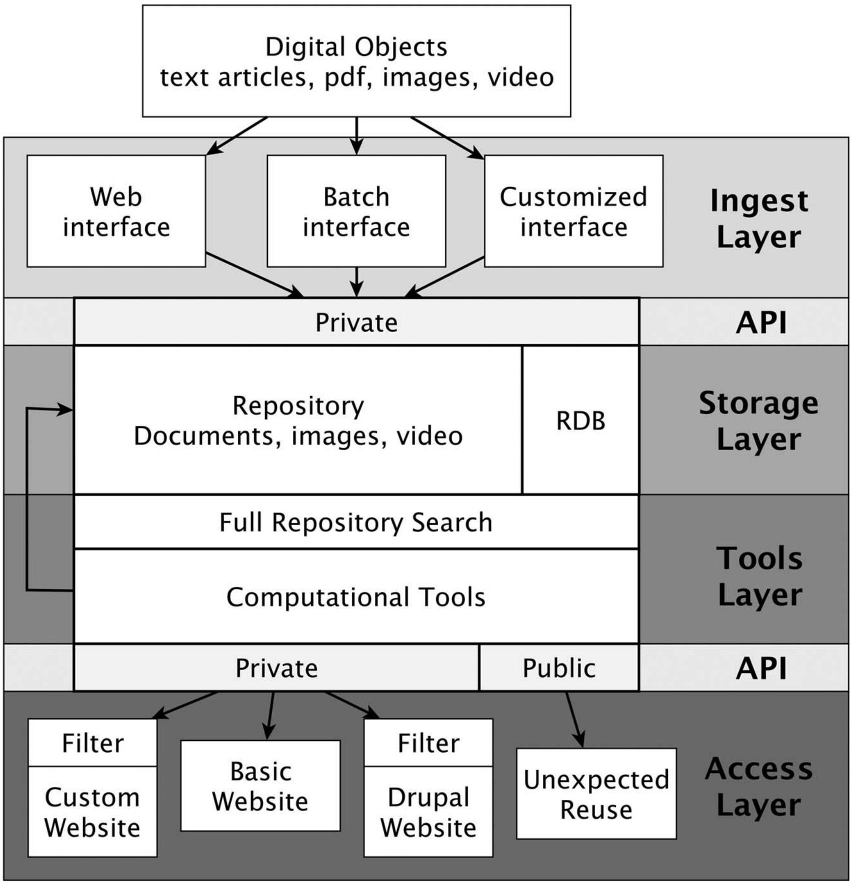

Figure 2. *Basic design of the research system in computational history of science, showing the connections between different layers of computational tools.*

transparency to historical scholarship. It allows for direct comparison of different interpretations of the same materials, which makes scholarly exchanges and debates faster and more lively. It also opens up new possibilities for crowd or communal projects. Even as traditional interpretive products of historical scholarship (monographs and articles) will continue to appear, these will be aided by shorter descriptive pieces that facilitate access to rich sources by leading readers to a vast array of different data objects in dispersed databases.[16] Virtual exhibits will bring together different types of objects, including original photographs or digitized archival materials alongside written narratives, to provide additional kinds of stories. Other resources will be added as well—for example, the results of computational approaches described earlier, such as interactive networks or quantitative analyses.

[16] http://www.edition-open-access.de/; http://embryo.asu.edu to the Embryo Project); and http://www.mpiwg-berlin.mpg.de/en/resources/index.html.

MULTIPLE LINKS WITH THE SCIENCES

Computational history of science broadens the audience for and the reach of the field by reinvigorating ties between the sciences and the history of science. It has long been recognized that the history of a science is, in principle, an important source of knowledge for most current scientific endeavors (although in practice this insight is often sacrificed in the rush to generate new data and results and is mainly tolerated in the context of pedagogy). One reason for this disconnect between science and its history and between intellectual insight and scientific practice is the divergence between the professional practices of sciences and of history of science. It is difficult to connect rich historical narrative to current scientific debate and demand for new data. The same is not true, however, for the more data-based structures, models, and networks that are the core of computational history of science. What we are already finding in some experimental studies is that when scientists can find the historical materials and interpretations easily, because they are linked through keywords and data, they read the history and find it useful. Computational research systems can make the history more useful and also more used.

CONCLUSION: EDUCATION AND BROADER IMPACTS

Our final concern is to ask why all this matters beyond its value for our own research. Computational approaches can contribute to improved education in several ways. First, as has long been recognized, history of science has value for education in K–12 venues and in promoting public understanding of science. Here it is not the connection of data so much as its presentation that can make the historical stories visible and accessible beyond our traditional scholarly presentations. The repository serves scholars, and web interfaces make the accumulating knowledge available to different user groups.

Furthermore, it is not just that the knowledge we historians of science have thought to add is now accessible; rather, a well-constructed system will allow users to explore and find new connections that we did not even envision. For the MBL History Project, we are developing exhibit tools to allow users to add their own stories, in a way similar to projects like Mass.Memories.[17] There is growing evidence that such projects attract teachers, students, parents, and the public to the work being generated.

In addition, the way we train our graduate students needs to change. But it is already changing, even as some traditionalists are either resisting or not seeing the changes. In fact, it is likely that every graduate program is learning that its students are well ahead of the faculty when it comes to finding, understanding, using, adapting, and developing computational tools for making historical scholarship better. We need to offer more training opportunities for these students—and also opportunities for them to teach the older generations.

Computational research systems for the history of science will be part of libraries and training programs. They will let researchers draw on more materials in new and effective ways. It is not the case that every historian of science must become a computational historian or must learn new tools or approaches, but it is the case that many will—and they will become the leaders. And it is certainly the case that we all have a responsibility to share our data and our findings.

[17] http://history.mbl.edu (MBL History Project); and http://www.massmemories.net/ (Mass. Memories).