

Chapter 2

From Descriptions to Duplicates to Data

Michael Lesk

Abstract Scholarly use of digital material moves from catalogues (locator services) to digital duplicates intended for human study to digital versions intended for computer analysis. We have been through this entire path for text, the easiest material to digitise, and we are now fairly far along with artistic imagery. More difficult content, such as costume and dance, will move through the same stages in the future. Perhaps the most important question is whether the nature of critical research changes as the tools change. Many early applications of computers were authorship studies, for example. More generally, does research based on computer analysis ask the same kind of questions as other research? Is it done on the same materials? So far, it would appear that the same materials are considered, and the same questions asked, but there are newer tools to apply. Algorithmic research can also study larger quantities of material, perhaps reducing the single-work focus of much cultural study.

Introduction

Two different forms of progress take place in digital cultural studies. First, we move from simpler to more complex media; text is easiest and is done first, followed by images and then video, sculpture, and specialised materials such as costumes. Second, we move from just listing the items available in catalogues, to providing substitute digital forms that may be suitable for human study, to doing the research automatically. This chapter compares the progress in both media and in study methods, dealing with previously existing objects, not “born-digital” items.

M. Lesk (✉)
Rutgers University, New Brunswick, NJ, USA
e-mail: lesk@rci.rutgers.edu

Digital collections may be larger than any traditional museum or library, and thus permit very wide-ranging comparisons and complete surveys. It is perhaps easier in the digital world to look at details rather than conceptual properties of works. It is easier to measure the size of a work than to say what it is about and still harder to say what emotions it will evoke in a person. Surprisingly, perhaps, it has been possible in textual studies to infer a surprising number of advanced properties, such as authorship or sentiment, from the statistical analysis of simple words. Such techniques are now appearing in research on images or sculptures as well as with text.

This chapter first looks at the problems of creating digital materials from historic objects, and then at their use. In each area, we tend to begin with very small amounts of material; in the 1960s, individuals would start by keying and studying one text, and by 1990 researchers would have large text libraries but only one video. As the amount of material increases, so does what we can learn.

Creating Digital Materials

The technology to digitise and analyze materials is easiest for text and cultural studies began there, but it has moved from text to images and then to video and sculpture, as displayed in this abbreviated chart (Fig. 2.1).

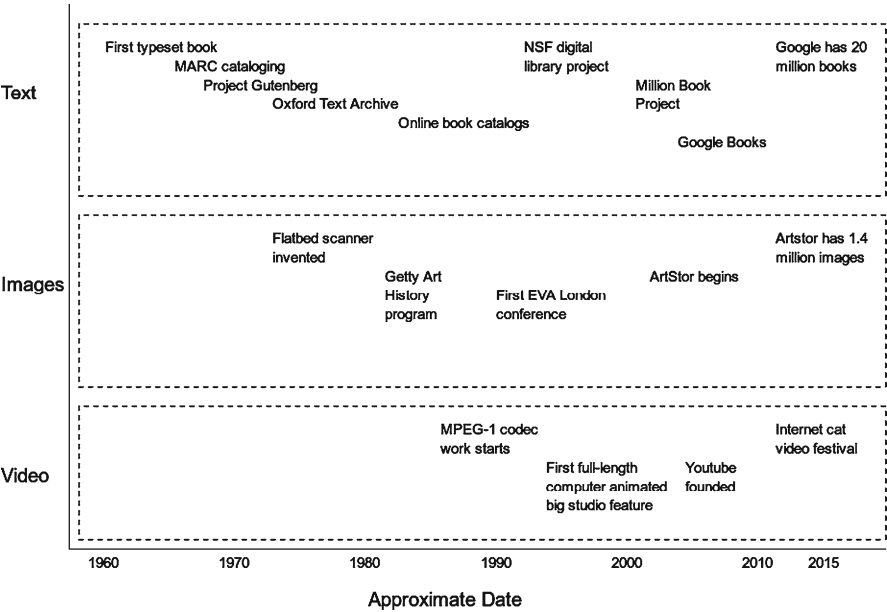


Fig. 2.1 Some milestones in computers and scholarship

In each area, we find that initially people have one item on which they work, and scholars deplore the inability to have full context. Efforts are made to have full descriptions and provide finding aids to let people know what is available where. Then, we start to have large enough collections with good enough reproductions that people start to do their actual work online. Finally, people start writing code to do studies of one or another level.

Books. The techniques to create digital materials have changed over the decades. In the 1950s and 1960s we converted text to machine-readable form by keystroking. Although a number of important works were completely converted, in the earliest days only some aspects of the text were done, for example the metrical patterns of poetry. Often a researcher would work on a single author or one book; I can recall from the 1960s conversions of Roman authors and of Icelandic sagas.

Libraries did the earliest broad conversion projects as they worked on catalogues. There was a long history of libraries maintaining locator services that reported which books were held in which libraries. For example, even 80 years ago Pollard and Redgrave [1] published their Short Title Catalog to locate pre-1640 books, and supplemented by Wing with the next 60 years [2]. Larger catalogues followed, including the *National Union Catalog*, whose hundreds of volumes were available in research libraries around the world.

Thanks to workers such as Henrietta Avram at the Library of Congress and Fred Kilgour at OCLC, in the 1960s we began to acquire machine-readable shared cataloguing [3]. Today OCLC WorldCat, a successor to both the original OCLC and the analogous Research Libraries Group's RLIN system, provides access to the largest catalogue of books that has ever existed.

As time went on, it became feasible to convert the actual books. At first conversion meant keystroking, whether on to punch cards or paper tape. The first effort completed was the *Thesaurus Linguae Graecae*, which did all the important works of classical literature, employing staff in Korea, China and the Philippines [4]. Project Gutenberg began in the mid 1970s, with a goal of 10,000 books.

Today books are often born-digital, but earlier works still needed to be converted. Several large conversion projects such as the Million Book Project, the Open Content Alliance, and Google Books, the largest of all, soon reached the point where the main barrier was copyright law. The average nineteenth century printed book has now been scanned multiple times. Although most of these projects have used manual page turning to present new pages to the cameras, Google has recently announced that they have been using a mechanical page turner. An explanation of how you could do this yourself using a vacuum cleaner even made it on to the web [5].

For some particularly valuable or difficult works to scan, unusual technologies can be used; these include IR and UV scanning, low-incidence light, and transmitted light illumination. For example, the British Library's manuscript of *Beowulf*, which was damaged in a fire in 1731, has been scanned so carefully [6] that the images are easier to read than the original. Cuneiform writing may now be scanned in 3-D to

help see what has been worn down over time; this is the modern equivalent of rubbing or tracing over the letters.

Flat artistic works. Once we had scanners, it became possible to scan some kinds of artwork such as prints, and digital cameras made it possible to scan paintings as well. Again, scholars had traditionally prepared directories to keep track of which works existed and where they were, such as Hind's 1912 [7] list of locations for all of Rembrandt's etchings.

The creation of online indexes to artistic works followed, of course. Howard Besser describes the Berkeley "imagequery" system planned in 1986–1990, to combine metadata access with images of artistic works [8]. As he writes, art materials are more difficult to catalogue than books; they do not come with a title page identifying the work and its creators, and may not even be signed. So art library catalogues have been a challenge to put online.

Obviously, once we had graphic display terminals, it became attractive to show the works themselves, rather than just their titles. The Andrew W. Mellon Foundation supported the ARTSTOR project which provided access to reproductions of art works. As of early 2013 it contains 1.4 million images [9]. In the United Kingdom the Public Catalogue Foundation has listed 145,000 of the estimated 200,000 oil paintings in the country, with images and locations for each artwork.

Digitising 3-D. There is also a history of cataloguing sculptures, but they are even less well organised than image cataloguing. Typically these are combined, since often the sculptural catalogue is illustrated with flat images, and many sculptors also did drawings or other flat works. The Public Catalogue Foundation, mentioned above, will add some 60,000 indoor sculptures to its files, while the National Recording Project of the Public Monuments and Sculpture Association will index outdoor objects.

The technology to scan 3-D objects is complex but includes "feeler" devices, photogrammetry from multiple cameras, structured light, and laser scanning. An early important project is the Digital Michelangelo work of Marc Levoy [10]. Using laser scanners, his team prepared 3-D images of many of Michelangelo's carvings. The "David" in Florence was scanned to an accuracy of 0.25 mm, letting scholars see individual chisel marks. Figure 2.2 shows a detail of a scan.

Still more difficult tasks. Many objects pose additional problems in digitisation. Costumes, for example, are not just 3-D objects, but they have insides and outsides. Some viewers wish to look at the whole garment, and some wish to examine details of fabric, stitching and decoration. For some scholars, it may be important to view the garment by transmitted light rather than direct light.

Similarly, recording 3-D motion for the study of choreography is still ambitious and difficult. Takeo Kanade [11] and others have shown how to capture multiple people in motion with multiple cameras, but it is far from a routine procedure. The movie industry is now using motion capture, as in the film *Happy Feet* where Savion Glover's dancing was captured and transferred to an animated penguin. In time, this technology should become affordable for dance groups.

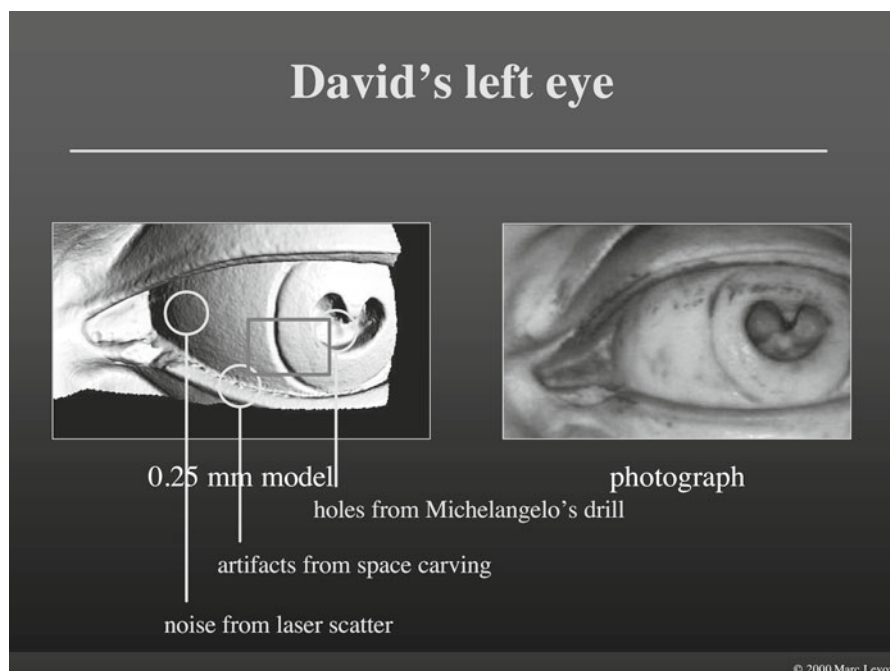


Fig. 2.2 Left eye of Michelangelo's *David*, Florence (Image by Marc Levoy, Stanford; used with permission)

Substitutes

Once digital material is available, the question is whether this substitute or “surrogate” is suitable for study. In the past, scholars demanded to see the original of anything being discussed. Microfilm, for example, was unpopular with generations of historians who much preferred to see the actual documents. Art historians have been even more insistent that they must do their work with originals.

Books. For books, we are well past the time when people insisted on paper copies. Robert Hayes reported in 1987 (quoted by Baker [12]) that half the readers he surveyed insisted on paper books, while today Amazon sells more e-books than paper books, and most scientists, certainly, do almost all their reading of research papers online. In 1990 chemists who were part of an early experiment on the use of digital journals [13] told us that they not only liked the arrangement of the articles in the journals they read but also liked the feel of the paper and even the smell of the brand-new issue (I offered to pour a bottle of PVA glue over the computer to provide the same experience). By 2009, the American Chemical Society had decided to publish only online in the future. Electronic texts are searchable,

enlargeable, and more flexible. Most of the physical properties of the paper book are not the work of the author, anyway; many more scholars study literature than study the way publishing houses used to design and compose books. Any instructor knows that students today resist all attempts to make them read paper.

Although the success of the Kindle seems sufficient to prove that people are happy reading on line, there are evaluations of both reading speed and comprehension [14–16]. Both speed and comprehension are so similar across devices that differences are within the standard error. Although the Harris paper noted that the Kindle might offer a slight increase in comprehension at the cost of a slight decrease in reading speed, the authors noted that the differences were not significant.

Other studies have shown differences in the way people read even if they take the same time and have the same performance. Wacholder [17] asked students to judge the utility of books for writing undergraduate papers, with some students skimming the books on paper and others on screens. They were both given the same time limit for the task. Although both groups of students performed comparably well at choosing the best books for a paper topic, the pages read as they examined the books were quite different. Readers who had the books on paper relied heavily on the table of contents and the index, and read only a small number of pages within the book. Those who had a PDF file, which they could search, looked at a larger number of pages in the book.

Do computer systems encourage the reading of “snippets?” Modern students are commonly criticised for superficiality, reading widely rather than deeply. Jidong Wang [18] noted that until about 100 years ago scholars of traditional Chinese literature were expected to memorise the books they were studying, and that the use of digital search systems had encouraged misunderstandings when people read without reading in context.

This complaint is not new. Plato in the *Phaedrus* had complained that writing was being used as an inadequate substitute for memory. Nor is griping about attention span limited to writing. As part of the 2012 election campaign there was an early debate in which each speech was limited to 30 s. The Lincoln-Douglas debate schedule was an opening speech of an hour, an opponent’s speech of 90 min, and a final 30 min reply by the first speaker. The use of photography instead of sketching also reduces the amount of time spent on each individual object, and presumably therefore the appreciation of its detail. However, the modern technologies permit the consideration of far more material, particularly when searching is available, and presumably if scholars did not like this, they would not use it.

Text processing has had the advantage of very generally applicable technology. We use the same formatting and searching algorithms on virtually all documents, whether they are letters, journal articles, or monographs. We can even share methods across multiple languages. By contrast, image processing methods are often quite specialised: we have software and even whole companies dealing with maps, CAD drawings, handwriting, and artworks [19]. The cataloging methods of libraries, museums and archives are converging, and as these groups continue to use digital methods and accumulate data, we may see more widely applicable algorithms.

Flat artistic works. Art historians have been more attached to the original image. Aby Warburg, immortalised by his library (now in London) seems to have pioneered the idea of using surrogate images for study [20]. He would lay out, in a large space, reproductions of works that he wished to compare or look at side-by-side. Even photographic reproductions were slow to be accepted in scholarship.

Quality is obviously a key consideration for images to be studied. Michael Ester [21] evaluated different image qualities and concluded that a 1 megapixel image (i.e. a $1,000 \times 1,000$ image) was an appropriate size. Now that cellphone cameras deliver 5 megapixel imagery, that seems pitiful, although megapixels are far from everything. Many visitors in front of Van Gogh's "Starry Night" today hold up a phone and take a picture of it, but no matter how many pixels in their cellphone, the low quality lenses, camera shake, viewing angle, and illumination result in an image inferior to what they could buy on a postcard in the MOMA store.

Lindsay MacDonald [22] asked about the ultimate resolution required for art study. His measurements found that a brush with a single sable or mouse hair might be $20\text{ }\mu\text{m}$ across, and that the smallest artifacts in paintings appeared to be in the neighborhood of $50\text{ }\mu\text{m}$. Similarly, a person with good vision might be able to see features of about that size (but not smaller). He suggested that 1,200 dpi would be adequate to resolve features that are $40\text{ }\mu\text{m}$ wide and thus be "enough" for any practical problem. For a 20×30 in. object, that is almost a gigabyte of image, but today that's easily handled. Gigapixel images are routinely used now in panoramas.

Michael Ester [23] discusses a number of reasons why surrogates are deprecated. They do not give a sense of scale, for example; if everything is the size of your computer screen, you do not know what was originally large and what was originally a miniature. Another problem is context. Seeing Michelangelo's "David" in a museum in Florence, surrounded by other sculptures and by paintings, is different from a screen view or holding a small-size copy, although our context is not the context Michelangelo had. He thought originally that it was going to be on the cathedral roof and would have seen it placed in a public square; it was not moved to the museum until more than 200 years after his death.

Perhaps the most ambitious recent scanning project is the Google Art project, which now includes more than 30,000 images from more than 150 collections. The project also includes views of the galleries in which the artworks are presented, so that one can see the works in their actual context, and "walk through" the museum. Although not a 3-D model of the museum, it gives a similar feeling.

3-D. Perhaps the quickest adoption of 3-D digitising was in architecture. Architects use CAD models of buildings to generate the construction blueprints and help with structural engineering [24]. Often architects create a physical 3-D scale model of the building to show the clients, just as they once built models of cardboard or wood; the CAD drawings and 3-D printers build these more easily.

Models of historic buildings are also useful for studying structural as well as artistic issues. Beauvais, for example, has a cathedral that is in danger of collapse. It has been scanned in detail with the aim of doing careful modeling of the structure [25], remembering that many times in the past engineers who thought they

were helping to stabilise a building actually made things worse. Cathedral structures are very complex and computer modeling is helping in the structural analysis.

The virtual reconstruction of destroyed buildings is now an active area of research. For example, a number of historic synagogues destroyed in Germany have been modeled and it has become possible to “view” them in full 3-D, rather than just look at photographs [26, 27]. An excellent example of remote access is the International Dunhuang Project [28] using virtual reality to both (a) let people look at wall paintings in caves in the Chinese desert, and (b) to let people look at objects which are now scattered around the world.

Research

The step after relying on digital copies rather than originals is to have the research done by computer algorithms. At first this “research” concentrated on mechanical tasks, such as the preparation of concordances [29]. Now we see relatively sophisticated studies being done by algorithms, involving syntactic and semantic analysis. Technology to help scholars comes from natural language applications such as sentiment analysis or machine translation. Although we are not yet in a comparable state those areas of scientific research where data processing is essential, increasingly computation appears in cultural and artistic research.

Books. Early high-profile research done with digitised texts included authorship studies. The first work in this area was done by hand; it is the famous study by Mosteller [30] on authorship of the *Federalist* papers. Morton [31] and Wake [32] were other early authorship studies. Other early work was on poetry, such as Milic [33] and Sowa [34]. Some of these projects were done without full text; the metrical patterns alone sufficed.

Robert Harris [35] described simple ways of doing literary analysis with computers. For example, Mark Twain made fun of James Fenimore Cooper for his repeated use of a character stepping on a dry twig and thus disclosing his presence. Harris suggests that one can easily count the instances of *twig*, *stick*, or *branch* and check whether Twain’s ridicule was justified.

More advanced work has followed, especially now that large collections of texts are available. Even the *New York Times* [36] has recently published an article about “big data” in literary research, showing that Jane Austen and Sir Walter Scott were the most influential nineteenth century writers, comparing the number of times that other works shared words and themes with them. To do this, more than 3,500 novels in machine-readable form were analyzed.

Michel [37] introduced the term “culturomics” and reviewed many cultural changes in terms of searches over the Google Books corpus. He demonstrates one instance of censorship in Nazi Germany by noticing that references to Marc Chagall disappear from German literature between 1936 and 1944, and also covers topics ranging across medicine, science, religion and cooking using the book data.

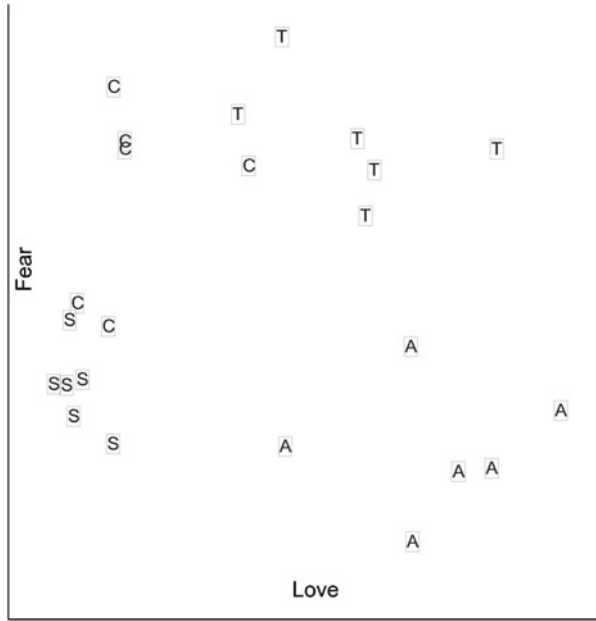


Fig. 2.3 Plot of thesaurus category occurrences. Books are as follows. (A) Jane Austen: Emma, Mansfield Park, Northanger Abbey, Persuasion, Pride and Prejudice, Sense and Sensibility. (C) Willkie Collins: After Dark, Armadale, The Evil Genius, The Moonstone, No Name, The Woman in White. (S) Sir Walter Scott: Guy Mannering, The Heart Of Mid-Lothian, Ivanhoe, Quentin Durward, Rob Roy, Waverley. (T) Anthony Trollope: Can You Forgive Her, The Duke’s Children, The Eustace Diamonds, Phineas Finn, Phineas Redux, The Prime Minister

As a very simple example of grouping literary works by genre, Fig. 2.3 is a plot of the relative occurrences of words in the Roget categories “Fear” and “Love” in six works by each of four authors: Jane Austen, Willkie Collins, Sir Walter Scott and Anthony Trollope. This is actually a visualisation and extension of an exercise in a first programming course. For both categories, the words listed in the category in a 1911 Roget’s Thesaurus were counted in each novel. The results are plotted; to nobody’s surprise, Jane Austen scores high on “love” and low on “fear” while Willkie Collins is high on “fear” and low on “love”. The Jane Austen novel that scores highest on “fear” is of course *Northanger Abbey*.

Flat artistic works. Research into paintings has been slower to develop, but James Z. Wang has done several provocative studies. One of these [38] looked at the ability of algorithms to recognise esthetically pleasing images. Features extracted from the images which had been rated for aesthetic appeal were used to fit the ratings; important features for this purpose included color saturation, round shapes, and image similarity (whether the image had components similar to components of other images in the data set).

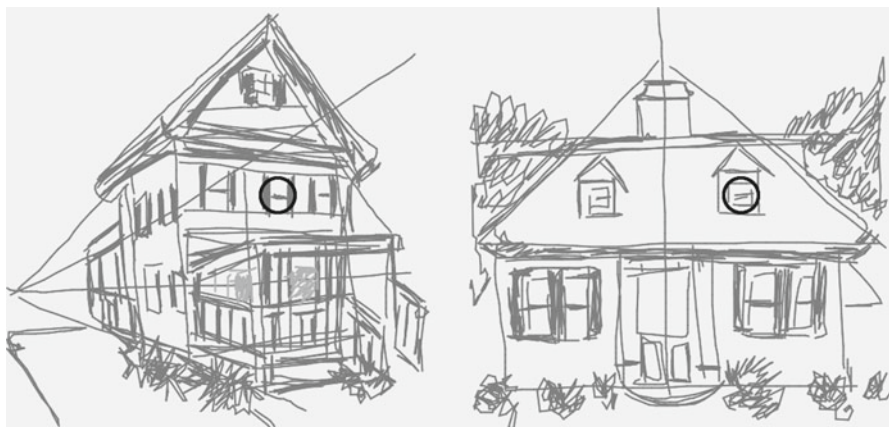


Fig. 2.4 Image matches across sketches (Drawings by Annamarie Klose, used with permission)

In another study [39] the brushstrokes in Van Gogh paintings were extracted automatically and his paintings were found to have longer and more regular brushstrokes than those of his contemporaries. This confirmed earlier art historical opinion but provided numerical measurements for the features. The work was not successful, however, at detecting a known forgery. Just as authorship studies have been of great significance in literary research, forgery detection matters in art history. Polatkan [40] describes promising work on forgery detection using wavelets.

Similarly Li [41] was able to extract brushstrokes from traditional Chinese painting. Graham [42] tried to extract stylistic features by spatial frequency analysis and analogised the problem to both the study of human perception and to the kinds of literary analysis described in the previous section. The paper attempted to distinguish naturalistic representations from more diagrammatic imagery, and talked about the new field of “visual stylometry.”

Figure 2.4 shows some very preliminary work on finding common elements in sketches. Drawings for which we know the stroke sequence were segmented by both time and space and then visually similar portions were found automatically.

Just as genres of novels can be distinguished by elements, Shamir [43] has been able to plot relations between artistic works. In his tree of painters, Rembrandt and Rubens are close as are Renoir and Monet, but those two pairings are far apart.

Other uses of image technology include Crandall [44], who has an ambitious discussion of tracking human motion and finding landmarks in general imagery. Manovich [45] also writes about applications of “big data” in cultural contexts. He discusses the role of computation for exploring massive data, to be followed by human analysis, and reviews the wide variety of data now available, ranging over social media, photographic sites, and commercially produced materials.

3-D Sculptures. Less has been done on sculptures, of course. As one example, Rodriguez-Echevarria [46] is building a 3-D inventory of sculpture with detailed

tags on the elements of the sculpture, but the tags are assigned manually. Flaherty [47] explains how 3-D printing techniques are being used to repair sculptures; but in this work, the decisions are made by scholars.

Today we can create 3-D maps of cities based on aerial photography, and 3-D databases of buildings. A remarkable recent study is Agrawal’s “Building Rome in a Day” [48], which used vast numbers of amateur photographs for photogrammetric modeling. Once data is available, we can expect algorithms that identify features characteristic of different sculptors and different sculptural subjects.

Has Research Changed?

Inclusion. A colleague of mine, Marc Donner, suggested that the most important aspect of digitisation would be the large number of items rescued from obscurity. Tens of thousands of nineteenth century books have heretofore been available only in the largest libraries, and the typical art museum can only display a small fraction of its paintings. Wider availability of little-known material might redirect research.

I skimmed the titles of articles in *English Literary History* for the years 1936, 1961, and 2010–2011. The subjects of study remain the traditional corpus: Spenser, Shakespeare, Chaucer and Milton continue to dominate. I saw only one author who had been obscure and but had been the subject of a recent article (James Hogg). I then checked the most recent issues of the journal *Literary and Linguistic Computing*, which of course contains a great many articles about tools rather than about authors. Its list of authors studied is more inclusive, since it includes non-English writers and more modern authors (including Agatha Christie and P. D. James) but it still contains Bacon, Bentham, Darwin, and Shakespeare. There are of course, some authors whose popularity rises and falls; see Fig. 2.5.

It may be too soon to make conclusions here. It has only been a few years since we have had millions of books online, and the barriers to commercial content still make it difficult for many scholars to do very wide-ranging studies, so that, for example,



Fig. 2.5 Google “ngram” showing references to Anthony Hope, Rex Beach and Pearl Buck

the work on literary influence described by Lohr [36] was done on nineteenth century books that are out of copyright.

A subject that is particularly popular in computer literary studies is still authorship, even beyond text. Dobrzynski [49] discusses attribution of Native American artworks, often collected a century ago with no attempt to record the name of the carver or other creator. For example, the Denver Art Museum identified someone known previously as the Master of the Chicago Settee (a wooden object in the Field Museum). Widespread access to multiple images, just as with multiple texts, is critical to such studies.

Questioning. Sometimes authors using computers address new styles of research problems. Consider three kinds of questions:

1. Those only suitable for computers (e.g., counting the number of superlatives in Dickens and Smollett).
2. Those that could be done either with traditional or with algorithmic methods, such as tracing influences by Jane Austen in the work of other writers.
3. Those that it is still hard to address algorithmically, such as discussing the role of religion in Keats' poetry or the importance of gold in the research of Michael Faraday.

The perceived danger is that research which is easy to do will supplant work that is more important. Looking at literary research, where the most progress has been made, we see traditional questions such as authorship, influence, and literary style still dominating. In artistic research, we see work on brushstrokes and composition. The academic community seems to maintain its focus, perhaps following Lord Balfour's aphorism "History does not repeat itself. Historians repeat each other." In science, "big data" is changing the way we study protein chemistry or flu epidemics.

The humanities are also benefitting from data analysis, including techniques such as quantitative historical methods, or Schilit's paper [50] on quotations from one book to another in a million book collection. Perhaps the most remarkable success has been statistical machine translation. If Google Translate can succeed based on statistical methods, so can stylistic or thematic analysis.

Conclusions

What sequence of research projects should we expect? There are techniques in commercial and government applications, such as sentiment analysis and social network analysis, now being introduced to cultural studies. The wide scope of cultural materials available cannot yet be easily exploited for practical reasons: Google may have scanned 20 million books, more than all but a few libraries, but copyright bottlenecks prevent access for researchers. Academic inertia also impedes new studies, including a perception that there are greater rewards for inventing a new theory than for accumulating or analyzing data, a problem that also exists in the sciences.

Image and sculpture study will follow behind literary studies, and will in turn be followed by film and dance. We can anticipate authorship studies based on comparisons of details, and similar stylometric analysis. We can expect emotional analysis to follow, just as sentiment analysis is now widespread with text. The kinds of image analysis done for searching and scanning – face recognition, feature extraction, analysis of differences – will be applied in cultural studies, just as statistical methods moved into historical research. This is the model of the intelligence agencies, which try to integrate information over millions and billions of messages and images.

The most relevant aphorism may be “more data beats better algorithms” [51]. The style of textual research that relies on huge collections and answers questions by statistical methods rather than individual item analysis can be applied to other media as well. Once we can recognise the features of cultural objects, and collect these features over millions of items, we can hope to reach some level of conceptual and intellectual understanding.

References

1. Pollard, A. W., & Redgrave, G. R. (1926). *A short-title catalog of books printed in England, Scotland, and Ireland, and of English books printed abroad, 1475–1640*. London: The Bibliographical Society.
2. Wing, D. *Short-title catalogue of books printed in England, Scotland, Ireland, Wales and British America and of English books printed in other countries, 1641–1700*, 3 vols. New York: The Index Society, 1945–51.
3. Borgman, C. L. (1997). From acting locally to thinking globally: A brief history of library automation. *The Library Quarterly*, 67(3), 215–249.
4. Brunner, T. F. (1991). The thesaurus Linguae Graecae: Classics and the computer. *Library Hi Tech*, 9(1), 61–67.
5. Quay, A. (2012). Google engineer builds automatic ‘page-turning scanner’, uses vacuum cleaner. Taxi, 13. <http://designtaxi.com/news/354337/Google-Engineer-Builds-Automatic-Page-Turning-Scanner-Uses-Vacuum-Cleaner/>. Accessed 28 Jan 2013.
6. Kiernan, K. S. (1991). Digital image processing and the Beowulf manuscript. *Literary and Linguist Computing*, 6(1), 20–27.
7. Hind, A. M. (1912). *A catalogue of Rembrandt’s etchings*. New York: Fred Stokes.
8. Besser, H. (1990). Visual access to visual images: The UC Berkeley image database project. *Library Trends*, 38(4), 787–798.
9. ARTstor Blog. *About ARTstor*. <http://artstor.wordpress.com/about-2/>. Accessed 28 Jan 2013.
10. Levoy, M., Pulli, K., Curless, B., Rusinkiewicz, S., Koller, D., Pereira, L., Ginzton, M., Anderson, S., Davis, J., Ginsberg, J., Shade, J., & Fulk, D. (2000). The digital Michelangelo project: 3D scanning of large statues.” In *Proceedings of the 27th annual conference on computer graphics and interactive techniques (SIGGRAPH’00)*, (pp. 131–144). New York: ACM Press/Addison-Wesley.
11. Kanade, T., Rander, P., Vedula, S., & Saito, H. (1999). Virtualised reality: Digitizing a 3D time-varying event as is and in real time. In Y. Ohta & H. Tamura (Eds.), *Mixed reality, merging real and virtual worlds* (pp. 41–57). Heidelberg, Germany: Springer.
12. Baker, N. (2001). *Double fold: Libraries and the assault on paper*. New York: Random House.
13. Egan, D. E., Lesk, M. E., Ketchum, R. D., Lochbaum, C. C., Remde, J. R., Littman, M., & Landauer, T. K. Hypertext for the electronic library? CORE sample results. In *Proceedings Hypertext ’91* (pp. 299–312). San Antonio, 15–18 December 1991.

14. Nielsen, J. iPad and Kindle reading speed. Jakob Nielsen's Alertbox, 2 July 2010. <http://www.nngroup.com/articles/ipad-and-kindle-reading-speeds/>. Accessed 29 Jan 2013.
15. Siegenthaler, E., Wurtz, P., Bergamin, P., & Groner, R. (2011). Comparing reading processes on e-ink displays and print. *Displays*, 32(5), 268–273.
16. Harris, P. (2012). Reading speed variations on paper vs. computer vs. Kindle. In *6th annual conference, vision performance institute*. Pacific University College of Optometry, OR. <http://commons.pacificu.edu/cgi/viewcontent.cgi?article=1006&context=vpir6>. Accessed 29 Jan 2013.
17. Wacholder, N., Liu, L., & Liu, Y-H. (2006). User behavior during the book selection process. *Proceedings American Society for Information Science and Technology*, 43:1–16.
18. Wang, J. (2006). Approaching pre-modern China through the computer: The impact of full-text databases on Sinological research. Private communication.
19. Lesk, M. (1985). Graphical information resources: Maps and beyond. In *Proceedings of 8th annual ACM SIGIR conference on Research and Development in Information Retrieval (SIGIR'85)*, (pp. 2–8). New York: ACM.
20. Humphries, T., & Thompson, S. (2012). *A Hippocratic Intuition for Balance in Warburg's Mnemosyne Atlas*. Transtechnology Research Reader, Plymouth University. <http://trans-techresearch.net/wp-content/uploads/2010/11/Humphries-and-Thompson.pdf>. Accessed 23 May 2013.
21. Ester, M. (1990). Image quality and viewer perception. Digital Image Digital Cinema. Supplemental Issue. *Leonardo*, 1(23), 51–63.
22. Macdonald, L. The limits of resolution. In A. Seal, J. P. Bowen, & K. Ng (Eds.), *EVA London 2010 conference proceedings*. Electronic Workshops in Computing (eWiC), British Computer Society, 2010. <http://ewic.bcs.org/content/ConWebDoc/36085>. Accessed 23 May 2013.
23. Ester, M. (1993). Image use in art: Historical practice. Gateways, gatekeepers and roles in the information omniverse, Association of Research Libraries, Washington DC. <http://web.archive.org/web/20110602124423/www.arl.org/resources/pubs/symp3/ester.shtml>. Accessed 23 May 2013.
24. Boland, R. J., Lyytinen, K., & Yoo, Y. (2007). Wakes of innovation in project networks: The case of digital 3-D representations in architecture, engineering, and construction. *Organization Science*, 18(4), 631–647.
25. Allen, P. K., Troccoli, A., Smith, B., Stamos, I., & Murray, S. (2003). The Beauvais Cathedral Project. In *Computer Vision and Pattern Recognition Workshop* (p. 10), Madison. Los Alamitos: IEEE Computer Society.
26. Krebs, F., & Brück, E. (2002). Historical buildings in 3D. In *EVA London 2002. Electronic Visualisation and the Arts*, London. http://www.memo38.de/root_IE/main/d/d_01/3d-visualisation_of_lost_jewish_heritage.pdf. Accessed 23 May 2013.
27. McNamara, M. Destroyed German synagogues reconstructed on web page. *The New York Times*, 24 September 1998.
28. Lutz, B., & Weintke, M. (1999). Virtual Dunhuang art cave: A cave within a CAVE. *Computer Graphics Forum*, 18(3):257–264. See also the International Dunhuang Project, idp.bl.uk.
29. Packard, D. W. (1968). *A concordance to Livy* (Vol. 4). Cambridge, MA: Harvard University Press.
30. Mosteller, F., & Wallace, D. (1963). Inference in an authorship problem. *Journal of the American Statistical Association*, 58(302), 275–309.
31. Morton, A. Q. (1965). The authorship of Greek prose. *Journal of the Royal Statistical Society, Series A*, 128(2), 169–233.
32. Wake, W. C. (1957). Sentence length distributions of Greek authors. *Journal of the Royal Statistical Society*, 120, 331–346.
33. Milic, L. T. Winged words: Carities of computer applications to literature. In *Proceedings Fall Joint Computer conference* (pp. 321–326). New York: ACM, 14–16 November 1967.
34. Sowa, C. A., & Sowa, J. F. (1974). Thought clusters in early Greek oral poetry. *Computers and the Humanities*, 8, 131–146.

35. Harris, R. The personal computer as a tool for student literary analysis. VirtualSalt, 30 December 1994. <http://www.virtualsalt.com/comptool.htm>. Accessed 29 Jan 2013.
36. Lohr, S. Dickens, Austen and Twain, through a digital lens. *The New York Times*, p. BU3, 27 January 2013.
37. Michel, J., Shen, Y., Aiden, A., Veres, A., Gray, M., Pickett, J., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M., & Aiden, E. (2011). Quantitative analysis of culture using millions of digitised books. *Science*, 331(6014), 176–182.
38. Datta, R., Joshi, D., Li, J., & Wang, J. Z. (2006). Studying aesthetics in photographic images using a computational approach. In *Proceedings of European conference on computer vision*, Part III (pp. 288–301), Graz, Austria, May 2006. Springer, Lecture Notes in Computer Science, Vol. 3953. Heidelberg, Germany: Springer.
39. Li, J., Yao, L., Hendriks, E., & Wang, J. Z. (2012). Rhythmic brushstrokes distinguish van Gogh from his contemporaries: Findings via automated brushstroke extraction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(6), 1159–1176.
40. Polatkan, G., Jafarpour, S., Brasoveanu, A., Hughes, S., & Daubechies, I. Detection of forgery in paintings using supervised learning. In *16th IEEE international conference on image processing (ICIP)* (pp. 2921–2914), Cairo, 7–10 November 2009. Piscataway: IEEE.
41. Li, J., & Wang, J. Z. (2003). Studying digital imagery of ancient paintings by mixtures of stochastic models. *IEEE Transactions on Image Processing*, 13(3), 340–353.
42. Graham, D. J., Hughes, J. M., Leder, H., & Rockmore, D. N. (2012). Statistics, vision and the analysis of artistic style. *WIREs Computational Statistics*, 4, 115–123.
43. Shamir, L., & Tarakhovsky, J. (2012). Computer analysis of art. *Journal on Computing and Cultural Heritage*, 5(2), 1–11, article 7.
44. Crandall, D., & Snavely, N. (2012). Modeling people and places with internet photo collections. *ACM Queue*, 10(5), 30–44.
45. Manovich, L. Trending: The promises and the challenges of big social data. 28 April 2011. http://www.manovich.net/DOCS/Manovich_trending_paper.pdf. Accessed 5 Apr 2013.
46. Rodriguez-Echavarria, K., Morris, D., & Arnold, D. (2009). Web based presentation of semantically tagged 3D content for public sculptures and monuments in the UK. In S. N. Spencer (Ed.), *Proceedings of 14th international conference on 3D Web technology (Web3D'09)* (pp. 119–126). New York: ACM.
47. Flaherty, J. Harvard's 3D-printing archaeologists fix ancient artifacts. *Wired*, 10 December 2012. <http://www.wired.com/design/2012/12/harvard-3d-printing-archaeology/>. Accessed 29 Jan 2013.
48. Agarwal, S., Yurukawa, Y., Snavely, N., Simon, I., Curless, B., Seitz, S. M., & Szeliski, R. (2011). Building Rome in a day. *Communications of the ACM*, 54(10), 105–112.
49. Dobrzynski, J. Honoring art, honoring artists. *The New York Times*, p. AR1, 6 February 2011.
50. Schilit, B. N., & Kolak, O. (2008). Exploring a digital library through key ideas. In *Proceedings of 8th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL'08)* (pp. 177–186), Pittsburgh. New York: ACM.
51. Rajaraman, A. More data usually beats better algorithms. *Datawocky*, 24 March 2008. <http://anand.typepad.com/datawocky/2008/03/more-data-usual.html>. Accessed 4 Apr 2013.

Electronic Visualisation in Arts and Culture

Bowen, J.P.; Keene, S.; Ng, K. (Eds.)

2013, XII, 291 p. 130 illus., 92 illus. in color.,

ISBN: 978-1-4471-5406-8