Anne HELMREICH (moderation), Matthew LINCOLN,
and Charles van den HEUVEL

Data Ecosystems and Futures of Art History

Research data is a capacious concept for art history, encompassing the denotation of our primary subjects of analysis, typically manifested in digital surrogates, databases, and other tools used to organize this body of evidence, as well as results of investigations that can range from technical and scientific to socio-historical. This debate focuses on the digital or computational facet of research data, but, as the responses reveal, this focus does not ignore our primary objects of study. Data, as an item of information, are commonplace in art history, the product of the labor of scholars and collecting institutions such as museums, libraries, and archives. With the second meaning of data—"*Computing.* Quantities, characters, or symbols on which operations are performed by a computer, considered collectively"[1]—unfamiliarity creeps in. In short, when in the form of a database to be consulted, such as *Allgemeines Künstlerlexikon Online,* data have been easily absorbed into art history curricula and research practices, but when in the form of digitally-based methodologies and computational approaches, art history has proven to be a slow adopter.[2]

Yet, amongst the discipline's origins are modes of thinking that lend themselves easily to computational approaches, as in Aby Warburg's *Bilderatlas Mnemosyne.*[3] Museums, libraries, and archives are rapidly digitizing their collections while embracing open access, making the historical record of concern to art historians increasingly available as structured data for computational analysis.[4] As major research projects and infrastructures advance, the field is also producing new research data as part of the processes of analyzing and interpreting the past. As a discipline, we must acknowledge this moment and prepare for a future in which research data and its management or curation—to use a term more familiar to the art historian—play an increasingly prominent role and is recognized as a scholarly outcome.

Such efforts are already underway, for example, in the natural sciences and social sciences,[5] and within higher education and research libraries.[6] While such conversations about data management ecosystems may not yet have permeated deeply or broadly into art history as measured by pedagogy and curricula, the discipline is well-positioned to contribute significantly. The cultural heritage community has invested deeply in developing shared standards for information management, as in the Getty Vocabularies, Iconclass, and the CIDOC Conceptual Reference Model. Digitized cultural heritage assets are being made available to the field in accordance with FAIR Data Principles—Findability, Accessibility, Interoperability, and Reusability—through such platforms as *Europeana.* Scholars have advanced digital art history considerably as demonstrated by the *International Journal for Digital Art History* and such convenings as "Art History in Digital Dimensions" (University of Maryland, 2016), "Art Histories and Big Data" (Lorentz Center Leiden, 2018), and "Grand Challenges: Digital/Computational Methods and Social History of Art" (Research and Academic Program of the Clark Art Institute, April 2019). Best practices are emerging, as in *The Socio-Technical Sustainability Roadmap* (Visual Media Workshop, University of Pittsburgh). Major data-driven research endeavors are bearing fruit, such as *Project Cornelia,* examining 17th-century Flemish tapestry and painting production, *Golden Agents: Creative Industries and the making of the Dutch Golden Age*, studying the dynamics between producers and consumers of creative goods and the various branches of the creative industries in the Dutch Golden Age, and *Pharos*, an international consortium creating a digital research platform for the study of photographic archives.[7]

But the scale of effort in art history, when compared to other disciplines, has been arguably relatively modest, leaving room for growth, and revealing the need for dedicated training in the data life-cycle, including community-based standards, data enrichment, data re-use and sharing, data interoperability, and machine learning as well as best practices in determining the benefits, costs, and risks of data management (including sustainability) and assessing quality and viability of research data. Exciting, innovative opportunities abound, ranging from linked open data and its promise for data integration; data transformation as archival sources are rendered machine readable; and harnessing the power of artificial intelligence, more specifically computer vision and machine learning, to advance image recognition and analysis.[8] The field would also benefit from investing in foundational digital literacy, a need laid bare by work-from-home conditions in the wake of the global health pandemic and the resulting reliance on digitized resources and digital platforms.

The contributors to this debate bring differing perspectives. Matthew Lincoln, in his position at Carnegie Mellon University Libraries, collaborates with scholars to plan and implement computational approaches to humanities research at the scale of individual projects. By contrast, Charles van den Heuvel, as the Project Leader for *Golden Agents*, financed by the Large Investments program of the Netherlands Organization of Scientific Research (NWO), is overseeing the development of infrastructures, ontologies, and interfaces for big data in art history, providing a framework for multiple research inquiries. These perspectives reflect the funding landscape of these contributors' respective geographies, with more opportunities for project-based funding in the United States in contrast to the large-scale infrastructure investment available in Europe, as demonstrated by DARIAH-EU, the pan-European infrastructure for arts and humanities scholars; NFDI-4Culture, the consortium for research data on material and immaterial cultural heritage; and the launch of the first task force for *Europeana*, the cultural heritage aggregator, focused on researchers' needs regarding digital tools and digitized cultural heritage.[9]

Recognition that research data are multivalent, shaped by differing needs and perspectives of collecting institutions and researchers, has framed this debate, which is also inflected by whether research is primarily driven by data, with researchers exploring pre-defined pools of data to surface observations, or particular research questions that delineate what data are gathered and then analyzed. Research data and their management are also faceted by various project or institutional roles, as well as long-term visions for data reusability and interoperability and attendant responsibilities of documenting provenance. This debate also surfaces how the concept of research data has been necessarily formed by art history's historiography. Implicit in this debate are questions regarding the future of the discipline, including whether we have an opportunity to shift research culture towards more consciously articulated hypotheses and more deliberate experimentation as research data and their management become core to our discipline's research practices.

Anne Helmreich: How would you define research data? Can research data be usefully distinguished from, for example, collections data, meaning the metadata that describes items held in an institutional repository (such as a museum collection, library collection, or archival collection)? What are new opportunities, challenges, and responsibilities that researchers have when working with institutional data, and vice versa?

Charles van den Heuvel: Data for me are inseparable from "doing research." Recently, I reviewed a paper in which the authors made the convincing case that selections when pre-processing text corpora for further machine-readable analysis (removing stop-words, conjugations of verbs or very frequent words with no essential meaning for the text) had an enormous impact on the quality of answers to certain research questions. They made clear that for some research questions it would be beneficial to remove numbers, dates, or words to minimize noise, but for other questions this came at a cost of important information loss. This difference in pre-processing is not only relevant for the statement that doing research cannot be separated from data handling, but it also means that research data differ from collection data in scope. Data of museums, libraries, and archives are often collected and described with either institutional (even national) purposes or the general, often unknown, public in mind. Frequently, additional selections and data processing are needed before data can be used for research.

Here, the distinction needs to be made between question-driven research and data-driven research. Data sets in the latter type of research stand closer to collection data although they are often composed from multiple "collections." Opportunities for working with large numbers of institutional data allow not only several hypotheses based on extrapolations of small data to be put to the test, but, more importantly, a greater critical reflection on doing research.

The biggest challenge for researchers is to develop mixed methodologies to handle partially integrated datasets and to analyze data of different quality and natures meaningfully. For cultural heritage institutions and creators of infrastructures, the responsibility is not to take decisions on data quality on behalf of the researchers, but to provide the provenance (i.e. the history of the origins of the data and of their enrichments) to let them to validate the data themselves.

Matthew Lincoln: Both academic art historians as well as researchers in art museums share the same broad mission to expand and share our understanding of the world's artistic heritage. We operate in the same scholarly ecosystem, citing each other as we work to advance our knowledge. To varying degrees, we all capture information about objects, describe their material and formal aspects, and structure information about the chain of events involving that object and its interaction with people, organizations, and places—from its creation, trade, and exhibition, to theft, destruction, and more. However, the responsibilities of the academic art historian differ radically from those of the collecting institution and its staff.

Museums' collections databases are shaped by their original (and, still, primary) function predating their current electronic forms: ensuring custodial accountability, to affirm the location and state of a given artwork at any given time.[10] Only as these cataloging systems were digitized and incorporated into other operational systems did they begin to store more and more historical data about objects, with ever more nuanced information about their creators, materials, provenance, bibliography, and exhibition history.

In contrast, academic art historians have far greater freedom when generating research data—whether those data are in spreadsheets, or just a carefully-kept notebook. They may define their own bounds of data collection based on whatever their current research project is, rather than constrain it to a single institution; they can include or exclude any attributes of artworks as their research questions dictate; they may develop data with far greater richness about a much smaller set of objects; their data may actually center on people, places, events, or themes, rather than on physical objects.

As more museums continue to make their ever-richer collections databases accessible online for both humans and machines, and as more art historians begin to consciously develop data sets to power their research, the boundary between collections data and research data blurs. But both technical and intellectual frictions remain. Most collections' data systems were not originally designed to store complex historical data, much less publish that data to an external audience in a complete and usable way. It takes great effort to reconfigure these legacy systems to support public downloading of collections data in bulk, and to train researchers to access that data and put it to use. The larger intellectual challenge remains in bridging the gap between the broad, object-specific research priorities of collecting institutions and the heterogeneous priorities of individual academic researchers, for whom museum data can only ever tell a partial story. Both kinds of data producers must fully understand each other's contexts, and self-critically document their own perspectives when publishing collection and research data within our shared research ecosystem.

A. H.: The digital humanities increasingly recognize the value of reusability of data, open access, sharing of data, and reproducibility of methods and procedures in data management. From your project experience, what do you consider best practice or lessons to be learned in these areas? How might the meaning and importance of data reusability differ between digital art history and reusability in, for example, the social or physical sciences?

M. L.: When discussing differences between institutional research data versus data produced by an academic researcher, I stressed that the diverging histories and responsibilities of each producer would fundamentally inform the nature of their data. These differing contexts mean that collecting institutions and researchers will likely never arrive at a synoptic data model that covers all needs for all users.[11] This panoply of viewpoints is further multiplied by our disciplinary tradition. Like many humanistic fields, art history tends to favor scholarship that disputes or re-imagines interpretive frameworks, premises, and perspectives, rather than coming to a single consensus about what particular attributes "matter" in a work of art. For example, we often challenge underlying assumptions about attribution, redefining concepts of authorship, creative intent, and execution. These ontological debates are a strength and marker of an intellectually active field! But this tradition is difficult to square with data organization practices derived from life sciences or social sciences. Their longer experience in explicitly data-driven study has given them more time to approach consensus on how data should be captured, and greater agility as a community when defining new data schemas to accommodate a new object of research or a novel framework of analysis.

Writing as a trained art historian who now works as a technologist focused on challenges in representing cultural heritage as data, I stress that it is not digital technology that imposes limits here, but cognitive labor. It is *technologically possible* to create data systems that can accommodate extremely complex datasets, with multiple or even contradicting statements from different contributors. But every additional layer of data complexity requires ever more labor from technologists to build and maintain that system, more labor from project researchers to enter and edit data, and more labor from other users to understand the structure of that data and repurpose it for their own research. We must always keep in mind this labor of producing and consuming data, all the more so when working in a very small team or solo. Prioritize your particular research goals when constructing a data set, and explain those priorities in data documentation. This will produce far more useful data (and a finished project!) than trying to include more detail and complexity in your data than you truly need.[12]

While collecting institutions are currently investing energy in developing better shared data infrastructure[13] for their shared needs, there are straightforward steps researchers can take today to ensure their data is more reusable by scholars and institutions alike. The more that both institutions and researchers utilize shared vocabularies in order to reference the same objects, people, places, and concepts in their data, the easier it is to recombine those datasets in a useful way. For example, using an institution's permanent object identifiers makes referencing an artwork unambiguous, as does citing the Uniform

Resource Identifiers for specific individuals, geographic places, subject matter, materials, and iconographies that are particularly useful for reconciling different languages and variant spellings.[14] Focus on adding these identifiers to your data whenever possible, without worrying about achieving the broad data coverage with which institutions must concern themselves. This will go a long way towards increasing data reusability, while still allowing you to have the flexibility required for your particular research project.

C. v. d. H.: Funding for the building of large infrastructures—such as the Golden Agents project[15]—comes with the condition that the developed tools and data can be reused by many present and future users in the digital humanities and beyond. The development of an infrastructure with more generic and future end-users in mind preconditions the reusability of tools and data. It stands in the way of tailormade solutions for particular (sub-)disciplines, such as art history, or for answering specific research questions. Therefore, the application of FAIR principles is unfortunately often abstract and top-down. Linking or re-using data from other research or cultural heritage enterprises seems, at first, straightforward. The assumption is that data just can be converted or be reused in a common data format. However, it is our experience in the Golden Agents project that existing formats and standards still have to be "negotiated." And even if data is linked in the same format, enrichments in the infrastructure are not necessarily reused by the cultural heritage institutions that provided the data in the first place. The potential reusability of data depends heavily on the data selections of these institutions and, even when their data were used for research, on the modelling of their projects and their original questions.

Such problems in the reusability of data in large infrastructures are valid not only for research based on questions, but also for data-driven research projects. In particular, in the case of data trained in a machine-learning context with a very focused outcome in mind, the data are often shaped in such a way that the resulting models can hardly be reused. Therefore, for every project, each experiment should be well-documented. The physical and social sciences have a longer tradition with experiments than the digital humanities. This, in particular, holds for digital art history in which experimentation is still in its infancy with the exception, perhaps, in scientific computer vision and conservation-restoration projects. Digital art history can learn from the experimentation culture of the sciences which requires good data management of all stages of the research process. To set up such experiments, it is necessary to bring together experts in the humanities and art history with cultural heritage specialists and computer scientists to model the data, run tests, and validate results.[16]

A. H.: Digital humanities projects are often collaborative and iterative. Data managed within such projects can therefore be enriched (as well as interpreted) by multiple researchers as it is applied to different queries. How can these multiple processes be productively managed? Is it desirable or feasible to recognize and attribute these acts of enrichment and interpretation to individual scholars?

C. v. d. H.: Provenance is key; not only for the data but also for enrichments and annotations. Part of good data management is being able to recognize who is responsible for these enrichments. This is not only in the interest of individual scholars who deserve credit for their contributions, but also for all end-users of the databases and research infrastructures at large to enable assessment of data quality in relation to their research questions and successive queries. Institutions, when curating data and, in particular, transforming unstructured data into structured data, produce much data and annotations that need to be mapped to existing organized forms of knowledge (ontologies, vocabularies, standards) in a formal, machine-readable language that will facilitate the exchange of data to optimize interoperability and reusability. Hereto, temporary sets can be created of word lists, name variants, tags of objects, etc., that need to be harmonized and standardized before they can be included in existing vocabularies, thesauri, or classifications. This requires that not only the users but also the institutions have insight into the provenance of these data and enrichments.

M. L.: Data sets are nearly always the work of multiple people. This is clearly the case in large teams combining many researchers, assistants, and programmers, but also applies to solo researchers who draw on supplementary data from many sources. Different collaborators on a digital art history project may seek varying kinds of outcomes depending on which role(s) they take on in that project, and their professional position.[17] Collaborators from a data or computer science department, or a library, may need to publish different perspectives on the project in different venues than the art historian, while other partners such as programmers may find public-facing components, such as a user interface or public code, more valuable than a credit line in an academic journal. Researchers should also consider how to acknowledge data coming from institutions, particularly when they use it as a foundation to build their own research. Using unique object identifiers when possible, or just ensuring that you use a museum's preferred citation for published datasets to help institutions understand how their data are being reused, is critical for collections managers and technologists to argue for more investment in higher-quality and more accessible data.

A. H.: Computer vision and machine learning have emerged as leading trends and grand challenges in many disciplines, including art history and cultural heritage institutions. What possibilities and implications do these approaches have for both scholars and collecting institutions and for handling and even distinguishing the data produced by these processes from that created by humans?

M. L.: Cultural heritage institutions are increasingly using computer vision to create new descriptive data about their collections. As Ryan Cordell details in his recent review of machine learning use in libraries,[18] collecting institutions have used computer vision to augment many of their core management tasks: to categorize items based on algorithmically identified visual properties; to identify objects depicted in images; to cluster images sharing similar content, composition, or color palettes; and to power innovative browsing and searching interfaces that use images as search terms to return visually similar objects from collections databases. Such systems have been institutionalized: the Cleveland Museum of Art's ArtLensAI offers a playful search that retrieves the image in their collections most similar to a photograph that you upload, while the Williams College of Art features an interactive browsing visualization that displays all of their artworks clustered by visual similarity.[19]

A fact elided when using an interface like ArtLensAI is that no computer vision solution is universal. Computer vision models are created by training with large sets of images that humans have already marked with the correct classifications for a given task. Some generic computer vision models, trained to detect everyday objects such as cars, trees, or chairs based on large sets of tagged internet imagery, are widely used across many domains from e-commerce to social media. While these generic models can be surprisingly effective for some artworks, they center conventions of modern, representational photography. This makes them only a rough fit for the wide spectrum of the global history of art. There is a strong need for multiple specialized models that could focus on different kinds of visual similarities (color palette versus overall composition versus figural pose,[20] for example) or ones fine-tuned for works on paper, photographs of sculpture, etc.[21]

For all their shortcomings, the most successful computer vision implementations in visual cultural heritage tend to work as technological aids for human experts, rather than supplanting them. A recent initiative with the Carnegie Mellon University Archives paired generic visual similarity algorithms alongside archival organization information to allow

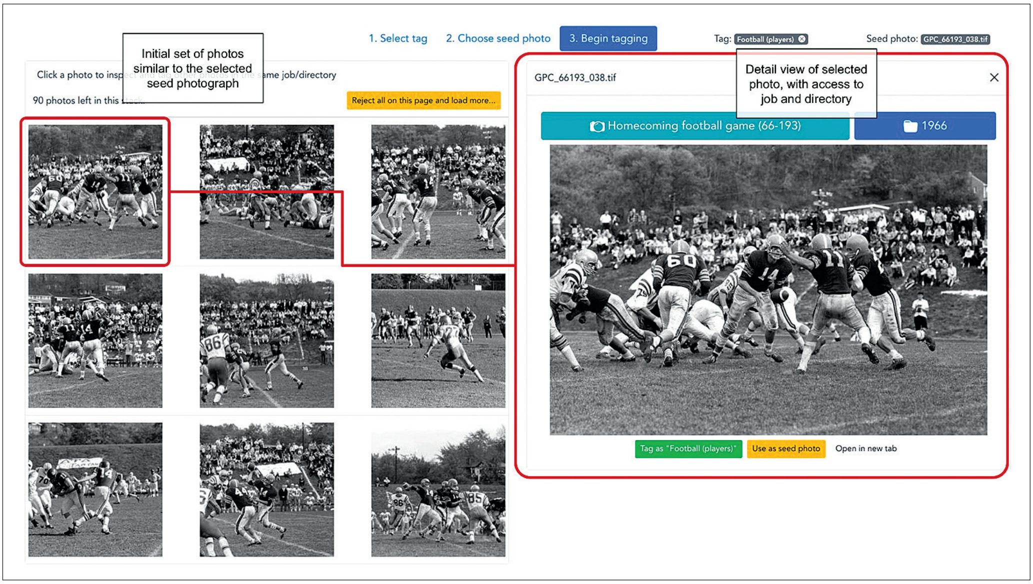

Fig. 1. CMU Computer-Aided Metadata Generation for Photo-archives Initiative (CAMPI) software, showing visually-similar images retrieved by a computer vision algorithm and a rapid description interface that linked back to the original location of the photograph in the original collection organization. Screenshot, 2021.

archivists to efficiently process a very large historical photographic collection (**fig. 1**).[22] Instead of trying to fully-automate the tagging of images, the program helped archivists to retrieve potentially relevant, but unlabeled, photographs for further description. All decisions remained under human control; the computer vision mechanism merely smoothed the otherwise-arduous process of sifting through tens of thousands of photographs so that humans could focus on making decisions that the computer could not.

C. v. d. H.: Peter Bell, a scholar in computer vision with a background in art history, believes that computer vision in digital art history/humanities and GLAM (Galleries, Libraries, Archives, and Museums) can have a "symbiotic relationship."[23] Computer vision needs big data to train algorithms. In return, it can make historic images machine-readable and provide additional information about metadata of images (that often describe the image's context, rather than the image itself) and similarities hereof, making GLAM data more accessible. Those similarities allow for recognizing and analyzing patterns, for instance, in motives and iconographical themes in cultural heritage and art historical studies. Yet, promising automated possibilities have serious implications for research practices and data-handling for both scholars and collecting institutions. For scholars, it implies working in teams of specialists in art history and cultural heritage, computer science, and data-analysis. For experts in art history, it often results in changing research roles by providing input to (pre-)classifying and indexing clusters produced by machine learning or training computer vision algorithms. For cultural heritage institutions, collecting research data requires finding ways to incorporate and to sustainably manage data that often temporarily (for the duration of a project) are stored on the larger servers of the science departments of universities. This entails that collecting cultural heritage institutions understand data-analytical research practices. For instance, while setting up experiments, training sets need to be kept separate from ground truth for validation. Moreover, instead of describing data as well-defined boundary objects, they

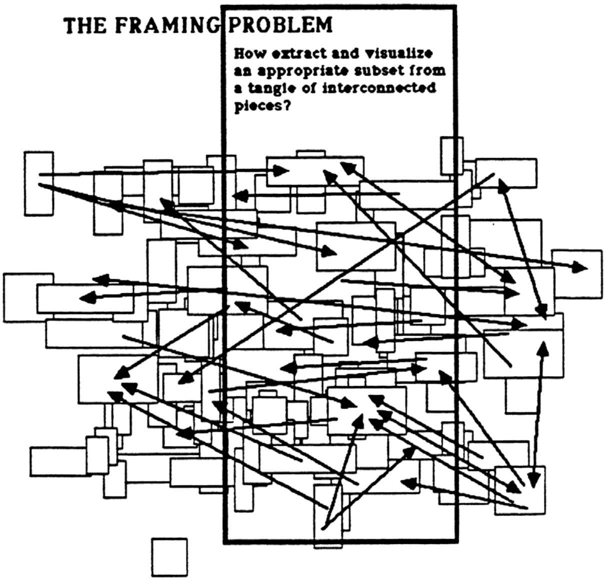

Fig. 2. Ted Nelson, "The Framing Problem," in *Literary Machines*, 1993, cited n. 24.

need to be open to new ways of determining meaningful patterns and clusters for what Ted Nelson called the framing problem (**fig. 2**): i.e. "how to extract, visualize (and I add preserve) an appropriate subset from a tangle of interconnected pieces of text and image."[24] Information specialists of GLAM institutions with longstanding expertise in metadata description can play a crucial role in the development of standardized ways of describing these new, mixed, rather loosely-defined collections and annotations hereof and to preserve them for future reuse in other projects.

A. H.: In your experience, when and how does managing research data appear in art historical training or curricula (if it does at all)? Looking ahead, and recognizing the increasing imperative of many cultural heritage institutions to be "digital first," how might we prepare future generations of art historians in best practices in curating research data? What broader implications for training might engaging with data hold for methods of hypothesis formation and argumentation in art history?

C. v. d. H.: Art historians, similar to all historians, could be taught basic skills in curating research data by learning how to read errors and translating research questions into spreadsheets. This, together with elementary training in network analysis, geographic information systems (GIS), etc., could be incorporated into the bachelor or undergraduate program and gradually extended to more complex methods and techniques underlying computer vision and artificial intelligence. The more complex methods of the latter could be taught in multidisciplinary research lab settings or in cultural heritage internships allowing art historians and others to work in teams. Engaging with data in a critical way can be beneficial to testing hypotheses that are very often based on extrapolations of small data sets and in falsifying or underpinning argumentation in art history. In particular, art historical research based on large, heterogeneous and distributed datasets requires a good understanding of managing data. Finally, it is crucial to teach art historians that the recognition of patterns that interest computer scientists often (implicitly) already make up part of their basic training (patterns in stylistic features such as composition, motives, iconological themes, etc.). Making these patterns explicit is crucial for data selection and for methods of hypothesis formation and argumentation in art history.

M. L.: At a recent colloquium on the state of art historical graduate education in the digital turn hosted by The Pennsylvania State University, art historians of all career stages took up the question of how to prepare the next generation of art historians to more closely engage with questions of data.[25] We all agreed that some core level of data and technological literacy were crucial for all art historians and should not be reserved solely for those who want to thoroughly integrate computational approaches into their work.

Our bibliographic and image databases, our museum collections, and even our work in word processors and managing our research images on a filesystem are all mediated through complex information systems with varying degrees of human and algorithmic intervention. In addition to the standard "methods" course that covers the historiography of our field, could art history graduate programs adopt a true "research methods" course? Such a course could provide both practical experience and critical reflection on a wide array of art historical information practices: (1) How libraries collect and make searchable our secondary literature. (2) The nuances of how curatorial records from a museum collection make their way (or not) into online databases and how those practices affect what objects are findable and how they are organized.[26] (3) How to use archival collections and reflect on their unique intellectual history, structure, and digitization practices compared to that in art museums. (4) How diverging practices in organizing knowledge across libraries, archives, and museums are being reexamined as they increasingly publish collections data, and how that data, in turn, can power new research and integrate with data collected or created by researchers.[27]

This approach could twine together the often unfairly-separated strands of "digital" versus "traditional" art history. It would underline how deeply knowledge organization practices underpin all types of art historical research, hopefully making the possibilities of data-driven research more accessible to *all* students, while also giving a crucial historical and critical grounding for budding digital art historians looking to use and create new datasets in their work.

Anne Helmreich is Associate Director at the Getty Foundation (former Associate Director, Digital Initiatives, Getty Research Institute). She and Pamela Fletcher recently coauthored "Digital Methods and the Study of the Art Market," in Kathryn Brown (ed.), *The Routledge Companion to Digital Humanities and Art History* (Routledge, 2020); in 2019, she coedited, with Murtha Baca and Melissa Gill, a special issue of *Visual Resources* dedicated to digital art history.

Charles van den Heuvel, art historian, is Head of Research in History of Knowledge at Huygens Institute for the History of the Netherlands of the Royal Netherlands Academy of Arts and Sciences and Professor of Digital Methods in Historical Disciplines at the University of Amsterdam. He is the principal investigator of *Golden Agents: Creative Industries and the Making of the Dutch Golden Age* and *Virtual Interiors as Interfaces for Big Historical Data Research: Spatially enhanced publications of the creative industries of the Dutch Golden Age.*

Matthew Lincoln is the Collections Information Architect at Carnegie Mellon University Libraries, where he focuses on computational approaches to the study of history and culture, and on making library and archives collections tractable for data-driven research. Recent publications include "Tangled Metaphors: Network Thinking and Network Analysis in the History of Art," in Kathryn Brown (ed.), *The Routledge Companion to Digital Humanities and Art History* (Routledge, 2020.)

NOTES

1. "Data, n.," *OED Online*, September 2020 [URL: oed.com/view/Entry/296948?result=1&rskey=hsBOmz&].

2. Diane Zorich, "Transitioning to a Digital World: Art History, Its Research Centers, and Digital Scholarship," *Kress Foundation*, June 2012 [URL: kressfoundation.org/Resources/Sponsored-Research/Research-Items/Transitioning-to-a-Digital-World].

3. Leonardo Impett and Franco Moretti, "Totentanz. Operationalizing Aby Warburg's Pathosformeln," *Literary Lab Pamphlet*, 16, November 2017 [URL: litlab.stanford.edu/LiteraryLabPamphlet16.pdf].

4. For an exploration of art-related data available on the internet, see Kaylee P. Alexander, Fiene Leunissen and Hans J. van Miegroet, "Imperfect Data, Art Markets and Internet Research," *Arts,* 8, 2019 [DOI: 10.3390/arts8030076].

5. "Research Data Framework," *National Institute of Standards and Technology*, 2019- [URL: nist.gov/programs-projects/research-data-framework-rdaf].

6. "Implementing Effective Data Practices Toolkit," *Association of Research Libraries*, 2020 [URL: arl.org/implementing-effective-data-practices]. Jane Radecki and Rebecca Springer, *Research Data Services in US Higher Education, A Web-Based Inventory*, New York, Ithaka S+R, 2020.

7. *Project Cornelia* [URL: projectcornelia.be]; *Pharos: The International Consortium of Photo Archives*, 2016- [URL: pharosartresearch.org].

8. For example, see the test application of Artificial Intelligence to Iconclass: "Iconclass AI Test Set," *Brill*, February 21, 2020 [URL: labs.brill.com/ictestset].

9. "Research Requirements," *Europeana*, 2019-2020 [URL: pro.europeana.eu/project/research-requirements].

10. Jennifer Trant, "Curating Collections Knowledge: Museums on the Cyberinfrastructure," in Paul F. Marty and Katherine Burton Jones (ed.), *Museum Informatics: People, Information, and Technology in Museums*, New York, Routledge, 2008, pp. 277-278.

11. Erzsébet Tóth-Czifra, "The Risk of Losing Thick Description: Data Management Challenges Arts and Humanities Face in the Evolving FAIR Data Ecosystem," April 30, 2019 [URL: halshs.archives-ouvertes.fr/halshs-02115505].

12. For a cogent explanation of how historians can prioritize the design of their data, see Mark Merry, *Designing Databases for Historical Research*, London, Institute of Historical Research, 2015 [URL : port.sas.ac.uk/mod/book/view.php?id=75], section C3.

13. *Linked Art* [URL: linked.art] is a rich and growing consensus model for art museums to share data online.

14. See, for example, the Getty Vocabularies, the Virtual International Authority File, the World Historical Gazetteer, Library of Congress Subject Headings, and Iconclass.

15. The Golden Agents project is a large investment project (3.6 million euros) funded by the Dutch Research Council; *Golden Agents* [URL: goldenagents.org]. For data management, see Chiara Latronico, Jirsi Reinders *et al.*, "Unlocking the Archives: A pipeline for scanning, transcribing, and modelling entities of archival documents into Linked Open Data," June 1, 2020 [URL : zenodo.org/record/3862817#.X50n2S9Q1E8].

16. The Dutch Research Council has stimulated collaborations between humanists, computer scientists, and cultural heritage experts for over fifteen years. See programs "Continuous Access to Cultural Heritage" (CATCH/CATCH+) and the "Common Lab Research Infrastructure for the Arts and Humanities" (CLARIAH and CLARIAH+).

17. Alison Langmead, Tracey Berg-Fulton *et al.*, "A Role-Based Model for Successful Collaboration in Digital Art History," *International Journal for Digital Art History*, 3, July 27, 2018 [DOI: 10.11588/dah.2018.3.34297].

18. For a recent review of GLAM initiatives doing this work, see Ryan Cordell, "Machine Learning + Libraries," *LC Labs Reports*, July 22, 2020 [URL: labs.loc.gov/static/labs/work/reports/Cordell-LOC-ML-report.pdf].

19. "ArtLens AI," *Cleveland Art Museum* [URL: clevelandart.org/artlens-ai]; "Collection Explorer," *Williams College Museum of Art* [URL: wcma-explorer.williams.edu].

20. For example, Peter Bell and Leonardo Impett, "Ikonographie und Interaktion. Computergestützte Analyse von Posen in Bildern der Heilsgeschichte," *Das Mittelalter*, 24/1, July 11, 2019, pp. 31–53 [DOI: 10.1515/mial-2019-0004].

21. Matthew D. Lincoln, "Yet Another Digital Surrogate? Computer Vision and the Future of Collections Management Systems," conference, *DHNord 2020: The Measurement of Images* (Lille, MESHS Lille Nord de France, November 18, 2020) [DOI: 10.1184/R1/13203917].

22. Julia Corrin, Emily Davis *et al.*, "CAMPI: Computer-Aided Metadata Generation for Photo Archives Initiative," conference (Pittsburgh, Carnegie Mellon University Libraries, October 8, 2020) [DOI: 10.1184/R1/12791807].

23. The author is indebted to Peter Bell for his information exchanged by email. For the computer vision project of Peter Bell, see Fabian Offert, *imgs.ai* [URL: imgs.ai].

24. Ted Nelson, "Literary 3/9 Machines," in *Literary Machines 93.1*, Sausalito, Mindful Press, 1993, with reference to his *Computer Lib* of 1974 in which he discussed the concept of "intertwingularity" and the need for alternative strategies for understanding and structuring complex data.

25. Center for Humanities and Information Summer Workshop: "Digital Art History and Graduate Education," convened by Elizabeth C. Mansfield and John Russell, Penn State University, 2020.

26. On the impacts of full text search technologies on the archival practices of historians, see Lara Putnam, "The Transnational and the Text-Searchable: Digitized Sources and the Shadows They Cast," *The American Historical Review*, 121/2, April 2016, pp. 377–402 [DOI: 10.1093/ahr/121.2.377].

27. Laurie Allen, Thomas Padilla *et al.*, "Final Report—Always Already Computational: Collections as Data," May 22, 2019 [DOI: 10.5281/zenodo.3152935].