

LYNN ROTHER, MAX KOSS, AND FABIO MARIANI

The histories of artworks are linked. After a work is created, its path may be marked by stints in galleries, private homes, storage facilities, exhibitions, and museums. At each of these stops, its life intersects with the lives of other works. We can think of these intersections as events involving people buying, selling, inheriting, looting, or otherwise transferring objects in specific places, at specific moments in time. When artworks end up in museums, information about the object's journey is recorded in its provenance, but these intersections with the lives of other works usually are not. Moreover, until recently, these shared histories of museum objects were only made visible through scholarly studies or exhibitions focusing on themes such as the influence of a particular collector or art dealer.

With museums increasingly recording and publishing their collection information digitally, provenance data is calling out to be connected through the use of technology. Provenance records typically consist of information, with varying levels of detail, on individuals, organizations, locations, transfers, and time periods related to ownership and custody changes. Transforming such records into linked open data (LOD), a web standard that defines how to

publish resources online, would allow for large-scale analysis of the patterns, trends, and networks of the circulation and dislocation of objects relevant to disciplines such as art history, economic history, and sociology. Moreover, an LOD approach to provenance would help museums make better use of limited resources: identifying objects with the same owners or similar paths within and across museum collections would produce synergistic effects in the research of specific object histories within and across institutions. The distributed forms of knowledge production that an LOD-driven approach facilitates would also enable multiple stakeholders to share more efficiently the work of collecting and recording critical data on, for example, a specific art dealer or a historical event that influenced the circulation of many objects across many institutions.

As object records are a form of writing history, provenance data also offers museums an opportunity to fulfill their social responsibilities of transparency, accountability, and inclusivity in light of twenty-first-century efforts around restitution and decolonization. While the relationship between provenance records and social responsibility may seem insignificant, this is by no means the case. For instance, if museums applied LOD standards to provenance data, third parties could research, query, and link this data. Among other effects, this would provide an opportunity for the reparative extraction of data, especially for objects whose histories are of greatest concern. In our view, it is in the interest of these institutions' missions to facilitate such endeavors, as this would allow any researcher to learn not only about the objects but also about museums' entanglement with the aftermath of injustice, thereby contributing precisely the kinds of knowledge and perspectives that may be missing from mainstream museum data.

With Linked Art, moreover, the museum and cultural heritage community has developed a data model that explicitly addresses the needs of art museums in creating LOD, with a particular focus on provenance.<sup>[1]</sup> Additionally, the Getty Research Institute, with its Getty Provenance Index, has embarked on a multiyear project to transform information about art dealers and collectors compiled over more than thirty-five years into LOD.<sup>[2]</sup> With the imminent release of millions of auction and art dealer data, information from this unique repository can be linked to, and often enhance, provenance records of individual museum objects, providing new insights on both the micro and macro scales—particularly about the art market. These initiatives are, in turn, echoed in the fledgling field of digital provenance studies.<sup>[3]</sup> In light of these recent developments, we believe that this is the moment for museums to conceptualize the process of transforming analog museum records or their digitized offspring into machine-readable, searchable, and linkable provenance data.<sup>[4]</sup>

As we see it, this process of transformation must take into consideration four interdependent challenges. The first concerns how museums deal with the legacy of diverse and heterogeneous

provenance practices and information, which we address in the first section of this essay. Until recently, provenances were often recorded inconsistently both across and within museums, and so we face biases and divergent levels of detail in and across provenance records with regards to which events and parties are documented, and to what standards. As museums are being asked to take stock of their collections, the consistency and transparency of their provenance records is a pressing matter.

In the effort to make their records consistent and transparent, museums are faced with a second challenge: the absence of provenance information, given that much historical knowledge is still missing. Based on our own experience, we have identified four aspects of provenance records that result from gaps in historical knowledge and ultimately pose challenges for these records' representation in a digital format: incompleteness, vagueness, uncertainty, and subjectivity.<sup>[5]</sup> As museums consider moving from analog to digital records, they must base any complex data modeling on the data they actually have, not on the model they want.

Of course, transformation from analog to digital is labor intensive, which raises the question of how museums allocate their resources—the third challenge they face. Not only is provenance research itself demanding work, but so too is the process of mapping provenance records into linked data. Furthermore, funding to support basic research and infrastructure remains limited. In order to transform provenance records into linked open data, museums thus face a resource-intensive process that requires not only commitment but also specialist knowledge from data and provenance experts—at least while data literacy remains beyond the standard skillset of humanities scholars.<sup>[6]</sup>

Finally, if we zoom out and consider museums as part of wider society, we see that they constantly face external demands. At the moment, for example, calls for restitution and the decolonization of museums require both accountability and action. As museums reconsider their collections in light of these evolving exigencies, they must consider where provenance sits within their wider institutional goals, missions, and priorities.

In light of these challenges, it is clear that if museums are to remain relevant to contemporary audiences, they must engage critically and self-reflexively with their own collections and, ultimately, the provenances of these collections. In so doing, they must position themselves strategically and also set priorities to determine how to distribute their already limited resources. As the transformation of provenance records from analog into digital formats becomes a priority for museums, a crucial question arises: Which data should be included, and which should be excluded?

In this essay, we lay out a conceptual framework that may help museums make informed decisions when they create machine-readable data from existing provenance texts, or when they begin to build provenance data from scratch. The framework offers a limited, resource-conscious intervention tailored to our present moment, which is a transitional one toward a more fully digitally engaged museum. The framework can guide museums in developing strategies for what data to model (and to what level of detail) when transforming their provenance records into LOD. In its adaptability, our framework allows for depth of description, where needed, through a layered approach to building provenance—a thickness that is in itself networked and collaborative, and potentially inclusive of a multiplicity of voices, thus addressing the changing role of museums today and the diverse expectations they face. At its core, this conceptual framework allows for institution-specific strategies while arguing for the use of collaboratively developed resources.

## LEGACIES OF PROVENANCE

In its basic definition, provenance (derived from the Latin *provenire*, or “to originate”) refers to an origin, to where a thing comes from; in art history, a provenance is considered a record of ownership changes of a cultural artifact.<sup>[7]</sup> However, as Gail Feigenbaum and Inge Reist have noted, “Provenance, firmly entrenched though it may be as a standard part of art historical research today, is neither stable as a concept nor constant as an instrument.”<sup>[8]</sup> We build on this insight here to highlight three interrelated dimensions of legacy information that need ongoing critical engagement as the digitization of provenances continues apace: that this information has served specific and sometimes competing interests, that, by convention, it merely approximates historical complexity, and that it has been shaped by cultural biases, implicit and explicit.

The goals of the different practitioners of provenance have not necessarily aligned and may have even been opposed. Provenance has been and continues to be produced in varying contexts, such as museums, the academy, and the art market. Curators, art historians, art dealers—all have brought their agendas to the writing of provenance, whether driven by scholarly or commercial interests, or both. Before museums began to emerge as independent institutions in the eighteenth century, provenances had a practical use: to substantiate value. Namely, they served as tools to verify (or, in the hands of bad actors, fake) the authenticity and attribution of works.<sup>[9]</sup> (Provenance still performs this function today, of course.) In a similar vein, provenances containing illustrious names—such as those of European aristocracy—have been used to market works as having “pedigree,” turning ownership history into cultural capital (to reference Pierre Bourdieu), by prioritizing important names and omitting lesser-known ones

from the record.<sup>[10]</sup>

Both now and in the past, provenances are the outcome of historical research obtained from a variety of sources, including but not limited to documentary evidence (catalogues, contracts, deeds, photographs, or wills) and material information that can be gleaned from the object itself (inscriptions, labels, or stamps). Yet the richness of the historical complexity that may live in the documentary evidence or the material traces has often not been reflected in provenance records themselves. These are governed by recording conventions that have favored and continue to favor lists of names, creating a kind of provenance shorthand: a highly conventionalized style that was shaped by practical necessities and the demands of the art market. (This is in marked contrast to the space granted discussions of attribution and subject matter on, for example, museum labels and catalogue pages.) In other words, provenances, as conventionally written, have always constituted a reduction or simplification of the information available in the documentary or material evidence—with or without digital transformation.

Provenance records now in the files of institutions and collectors are palimpsests of prior uses and the concomitant historical entanglements and biases of those uses.<sup>[11]</sup> This means these legacy records may not align with today's expectations for historical accuracy and detail. One significant broad cultural and/or historical bias is gender discrimination. This has resulted in the under- and misrepresentation of women in provenances, still visible in old-fashioned naming conventions, such as “Mrs. John Doe.”<sup>[12]</sup> Similarly, usually only objects that could be attributed to a single maker were given provenances that start with this artist, expressing a cultural bias prioritizing individual authorship.<sup>[13]</sup> By contrast, provenances of ethnographic objects rarely included any information predating the object's extraction.<sup>[14]</sup> Sometimes, bias is expressed implicitly in the way transfers of ownership are described: an object that was looted during colonial terror, such as the 1897 sack of Benin City, may be described significantly more neutrally as having been “remov[ed] from the Royal Palace.”<sup>[15]</sup>

These interrelated dimensions of legacy information are still present in today's data but are being corrected by new approaches to the use of provenance records. Inspired by anthropologist Arjun Appadurai's concept of the social life of things, provenances have recently been used as a means of narrating biographies of works and have become an invaluable source for histories of taste, collecting, and art markets. Moreover, they have become the main source for identifying unlawfully appropriated artworks.<sup>[16]</sup> These newer political and legal demands regarding objects from contexts of injustice have put pressure on museums and changed their practices with regard to provenance, as we will discuss in the next section.

## PROVENANCE AND THE MUSEUM

As the ultimate repositories for many artworks, museums have become, of all stakeholders, the most intimately engaged with the practice of provenance. Their role has become that of a clearinghouse for the diverse historical and historiographical tendencies that have shaped the production of provenance and the varying ways of studying, compiling, and recording information that underlie them. Today museums face a responsibility to account for the histories of the objects they own. When museums encounter provenance gaps—especially for historically problematic contexts, such as World War II or the colonial era—it falls to them to actively fill in the missing information, which often requires resource-intensive research in multiple archives spread over several countries. Regardless of what provenance information museums may receive, they carry the burden of ensuring that the provenances they publish and produce are in line with contemporary standards. This section describes the context in which, relatively recently, museums’ standards for recording provenance emerged. Any move toward provenance LOD will have to grapple with these standards while also recognizing the problematic character of different forms of legacy information.

The end of the Cold War and the unification of the two German states constitute a historical watershed for museums and their engagement with provenances. This development coincided with the rise of memory studies, especially concerning the legacies of the Holocaust. While the geopolitical changes meant that artworks and archives that had been difficult to access since the end of World War II were now available to study, a new awareness, especially in unified Germany, arose around the need to address the injustices related to National Socialism. The 1998 Washington Conference Principles on Nazi-Confiscated Art constituted a culmination of these developments and established a scientifically rigorous provenance practice, especially in museums. With their commitment to transparency, the so-called Washington Principles, endorsed by delegates from forty-four countries and non-governmental organizations, codified in a legally nonbinding way measures to make establishing provenance easier as a step toward restitution and historical justice.<sup>[17]</sup> More recently, provenances have also become essential in researching unlawfully appropriated objects in the context of European colonialism and the extractivist policies that exploited both natural and cultural resources.<sup>[18]</sup> Following the example set by the Washington Principles, Germany (to cite just one country with a colonial legacy) has established its own guidelines on the documentation and publication of collections from colonial contexts, the so-called 3-Road Strategy.<sup>[19]</sup>

In light of these developments, in recent years many museums have devoted resources to examine works possibly affected by National Socialism in particular, and this field has produced the best examples of detailed documentation of provenance.<sup>[20]</sup> With the available information,

we can describe in detail specific events, such as the removal of what the Nazis labeled “degenerate art” from museums in 1937, for example, or confiscations during World War II undertaken by the Einsatzstab Reichsleiter Rosenberg, or ERR (Reichsleiter Rosenberg Taskforce).[21] The provenances of affected objects may therefore be rich in information for a short period of the object’s life but thin for much of the rest. Moreover, as provenance research is sometimes funded by third parties (such as the Samuel H. Kress Foundation), and this research often targets specific types of objects or specific historical moments or geographies, there can exist biases within collections around which objects have more detailed provenances—especially at institutions that have not engaged in systematic provenance work before.[22]

In response to the Washington Principles and the aforementioned discrepancies in producing provenances, the American Association of Museums (AAM) published their guide to provenance research in 2001. The guide covers a host of issues relating to provenance research broadly, and it also offers parameters for how to record provenance information, aiming to establish a modicum of standardization across institutions.[23] The AAM format, which developed out of an analog, textual provenance practice, uses syntax to create limited structure in the provenance. It relies on punctuation to convey meaning: a period records a gap between events, whereas a semicolon between events signals that the second event was directly subsequent to the first; brackets set off the life dates of parties; and parentheses mark the type of party in an event, such as a dealer or agent. Information that is not immediately relevant to the actual transfer of ownership or title—for example, knowledge about location changes, consignment status, or illegal transfers, say in the context of National Socialist expropriation—may be kept in separate notes.

In line with the efforts initiated by the Washington Principles to make provenance information public, museums, particularly in the United States, have begun to transfer their provenance texts from the analog originals, usually held in object files in the registrar’s office or curatorial departments, to the digital domain. Some institutions, such as the Metropolitan Museum of Art in New York and the Art Institute of Chicago, have done so for tens, if not hundreds of thousands, of works. These provenances mostly live in free, unstructured text fields in the museums’ collection management databases, which is often the source material that populates the museums’ collection websites. Figure 1 shows one such example from the collection management system of the National Gallery of Art in Washington, DC: the provenance of a 1940 painting by Henri Matisse, *Woman Seated in an Armchair*. We will return to this example throughout this essay.[24]



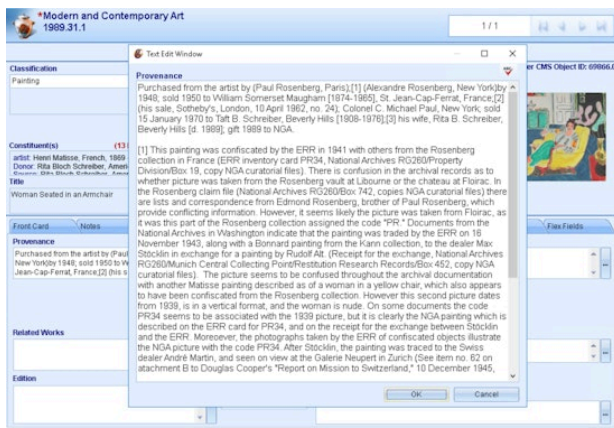


Fig. 1

The provenance of Henri Matisse, *Woman Seated in an Armchair*, as it appears in the free, unstructured text field of the collection management system of the National Gallery of Art, Washington, DC.

These provenances mainly conform to the limited best practice format laid out by the AAM guide. However, even within these parameters, there exists a fair amount of variety in provenances. And while variations do exist among museums (due to what we might call “house style” for recording provenance), they also exist across disciplines and curatorial departments, as well as within departments—even within single records, where it is sometimes possible to pinpoint provenance styles belonging to different individuals.<sup>[25]</sup> With digital provenance records now freely available on museum websites, the heterogeneity of provenance is more visible than ever before.

The increasing online publishing of provenance provides a basic level of transparency, useful especially for people with claims on specific artworks. We must remember that having a published provenance, even if only in a free text field, is better than not having *any* published provenance—which continues to be the case for the majority of works across thousands of institutions. While artworks can be found with relative ease through basic tombstone information (creator, title, date of creation, medium, dimensions), this is only possible when claimants know precisely what they are looking for and, furthermore, when the attribution, title, dimensions, and so on have not changed over time. As re-attributions and other changes are common for cultural objects, the presence of a specific family name or a gap within a published provenance on a museum website may be precisely what would allow claimants to find potentially looted objects. As cumbersome as that process may be, given that claimants would have to look up objects individually, object by object, at every museum because provenance criteria remain unsearchable on the majority of museum websites, claimants would at least have the possibility to search for some criteria. Lastly, if a provenance is published, even if only in the basic format laid out by the AAM, when new information appears, it can be kept up to date by editing the underlying information in the collection management database.

We must acknowledge the considerable effort it took museums and other cultural heritage institutions to arrive at this basic level of provenance transparency and recognize it as the game-



changing undertaking that it is and continues to be. Museum documentation is a herculean task with occasionally competing goals. Institutions have to balance researching, recording, and keeping records up to date with incorporating information from new sources, as well as responding to even more fundamental shifts such as the present moment's calls to decolonize the museum (and not only ethnographic museums).<sup>[26]</sup> At the same time—and this is the crucial part—given the ever-increasing amount of information available, research being done, and new archival sources accessible digitally, museums face legitimate questions as to whether their current provision of provenance information is adequate or not.

Considered together, the variation in how provenance texts are organized and written makes it clear that we are dealing with diverging concepts of what provenance is and what purpose it serves. Since we are concerned in this essay with the question of provenance *data*, we will need to keep these heterogeneous and sometimes ideologically unreconstructed provenance *texts* in mind as we consider the technical possibilities available for transforming existing texts into data.

## TOWARD PROVENANCE DATA

In this third section, we address the difference between unstructured data in text-based provenance records and structured provenance data. We do so to draw out the kinds of considerations that need to be kept in mind when data is being structured—labor performed by humans, aided or not by artificial intelligence—and to point out the potentials and pitfalls of structured and unstructured data, respectively.<sup>[27]</sup>

Let us begin by looking at the limitations of unstructured data, i.e., information whose internal logic is not explicit to computers and hence cannot be processed automatically. First, recording provenances in free text fields in collection management systems (as shown in figure 1 above) not only silos information but also duplicates (or multiplies) it within the database: for example, each mention of a particular collector's name does not reference a single, digitally available biography record of that person but rather constitutes a unique appearance in each provenance in which it is referenced, without any connection to its other appearances. This, in turn, multiplies the work involved in the upkeep of this information and the addition of new research. For example, if research uncovers new, essential documents about a nineteenth-century Parisian art dealer that affect the provenance records of multiple artworks in a collection, each record would need to be updated individually. Free-text provenances, even those produced and shared in line with the AAM format, also make it difficult to perform complex queries because they are not machine readable, which means that the structure required for a computer to understand the textual logic of the provenance is missing. This is highly relevant for people with

legal claims to specific works that may have been looted or otherwise unlawfully expropriated. Structured data can be queried by multiple parameters simultaneously, for example, by searching for objects that meet the criteria of being produced by Edo people and are known to have left the territory of Nigeria before its independence from the United Kingdom on October 1, 1960. Such a query would include only the objects meeting these criteria, and exclude those acquired legally after the end of colonial rule, for example. Without machine-readable data, claimants have to go through databases object record by object record, one museum at a time, to find objects with provenance gaps or information that might fit their search criteria, as indicated earlier.<sup>[28]</sup>

Siloing information also prevents analysis within or across institutions, and creates unnecessary obstacles in linking to data, such as archival materials or digitized auction catalogues, provided by external sources. Continuing with the example of Matisse, we would ideally like to be able to use provenance in databases to answer a question such as, “For how many and for which Matisse paintings sold between 1939 and 1945 in Paris is the purchaser’s name known?” But today, this is not a question we can answer through the available data, and thus we are unable to discern the links, patterns, and particular historical trends that would be made visible through the bigger picture—that is, with aggregate data.<sup>[29]</sup>

Finally, provenance records on museum websites are, currently, often not downloadable. This is a considerable impediment to networked research, which is a fundamental aspect of provenance research, in which researchers rely heavily on the work and findings of other researchers.<sup>[30]</sup> It is also exclusionary inasmuch as only a small number of individuals associated with the institution can access and edit the information. While it is understandable that museums feel a sense of responsibility around ensuring the accuracy of this information, it also may have an unintentional gatekeeping effect: it perpetuates professional, institutional, and disciplinary biases and hampers epistemological shifts toward a more inclusive, multi-perspectival approach.<sup>[31]</sup>

The shortcomings we have just described are particularly problematic in the context of looting, appropriation, and restitution, as well as in the context of accessibility. However, they are by no means limited to such cases. For example, authentication and identification of fakes could be much improved by cross-referencing and triangulating the data of objects, provenance records, and the records of collectors and dealers. Indeed, the history of collecting and art markets would benefit from identifying trends and patterns in large amounts of data, decentering consideration of individual objects in isolation and turning the focus of research to potentially unaddressed historical phenomena involving many objects.<sup>[32]</sup> The list of areas in which the analysis of provenance data could be useful is long: identifying tax fraud, money laundering, black market

movements, and the distribution (or lack thereof) of wealth and capital across time and geography, to name a few.

The limitations of unstructured provenance records notwithstanding, such records, like the one provided by the National Gallery for its Matisse painting (fig. 2), do currently allow for full-text searching: finding a particular string of alphanumeric characters (any combination of letters, numbers, and special characters such as exclamation marks) within their texts. The computer can, for example, retrieve all provenance records containing a particular string of characters such as “Paul Rosenberg.” With such a search, the computer will find all records containing “Paul Rosenberg,” opening up possibilities for analyzing the art dealer’s importance for the collection. However, as the example from the National Gallery shows, we could not reliably find this very record by performing a search for the string “Matisse” in the museum’s provenance fields. While Matisse was the first owner of this work, it cannot be found by searching for “Matisse” because he is recorded as “the artist” and not explicitly named. Instead, this information is only captured as so-called tombstone data and thus lives in a separate field in the database. Similarly, searching for “Paul Rosenberg” in unstructured provenance texts will return all entries containing “Paul Rosenberg” without any further specification. This means the search will bring up all objects in which his name appears with that exact spelling, regardless of his role in a particular provenance—whether he was buying or selling an object, or even just published a book referenced in the notes of the provenance text. For searches to return the objects that meet a set of criteria, such as a particular individual playing a particular role, we need structuring.

## Provenance

Purchased from the artist by (Paul Rosenberg, Paris);[1] (Alexandre Rosenberg, New York)by 1948; sold 1950 to William Somerset Maugham [1874-1965], St. Jean-Cap-Ferrat, France;[2] (his sale, Sotheby's, London, 10 April 1962, no. 24); Colonel C. Michael Paul, New York; sold 15 January 1970 to Taft B. Schreiber, Beverly Hills [1908-1976];[3] his wife, Rita B. Schreiber, Beverly Hills [d. 1989]; gift 1989 to NGA.

[1] This painting was confiscated by the ERR in 1941 with others from the Rosenberg collection in France (ERR inventory card PR34, National Archives RG260/Property Division/Box 19, copy NGA curatorial files). There is confusion in the archival records as to whether picture was taken from the Rosenberg vault at Libourne or the chateau at Floirac. In the Rosenberg claim file (National Archives RG260/Box 742, copies NGA curatorial files) there are lists and correspondence from Edmond Rosenberg, brother of Paul Rosenberg, which provide conflicting information. However, it seems likely the picture was taken from Floirac, as it was this part of the Rosenberg collection assigned the code "PR." Documents from the National Archives in Washington indicate that the painting was traded by the ERR on 16 November 1943, along with a Bonnard painting from the Kann collection, to the dealer Max Stöcklin in exchange for a painting by Rudolf Alt. (Receipt for the exchange, National Archives RG260/Munich Central Collecting Point/Restitution Research Records/Box 452, copy NGA curatorial files). The picture seems to be confused throughout the archival documentation with another Matisse painting described as of a woman in a yellow chair, which also appears to have been confiscated from the Rosenberg collection. However this second picture dates from 1939, is in a vertical format, and the woman is nude. On some documents the code PR34 seems to be associated with the 1939 picture, but it is clearly the NGA painting which is described on the ERR card for PR34, and on the receipt for the exchange between Stöcklin and the ERR. Moreover, the photographs taken by the ERR of confiscated objects illustrate the NGA picture with the code PR34. After Stöcklin, the painting was traced to the Swiss dealer André Martin, and seen on view at the Galerie Neupert in Zurich (See item no. 62 on attachment B to Douglas Cooper's "Report on Mission to Switzerland," 10 December 1945, National Archives RG239/Entry 73/Box 82, copy NGA curatorial files).

The NGA picture was returned to the Rosenbergs by 1948, according to the records of the gallery, which sold it to Somerset Maugham in 1950.

[2]According to Rosenberg gallery records.

[3]Correspondence between Paul and Schreiber in NGA curatorial files.

## Associated Names

ERR

Martin, André

Maugham, William Somerset

Paul, C. Michael

Rosenberg & Co. Paul P.

Schreiber, Rita Bloch

Schreiber, Taft B.

Sotheby's

Stöcklin, Max

Fig. 2

The provenance, with notes, of Henri Matisse's *Woman Seated in an Armchair*, as published on the website of the National Gallery of Art, Washington, DC, as it appeared on November 22, 2022.

Figure 3 shows an example of structured provenance data for *Woman Seated in an Armchair*, in which, for reasons of clarity, we have replaced "artist" with "Henri Matisse." Because we do not know the exact date when the object passed from Paul to Alexander Rosenberg, only that it happened before or in 1948, we have structured this information according to the Extended Date/Time Format (EDTF), created by the Library of Congress to address data fuzziness in date and time formats.[33] In this case, we chose a tabular data structure: each row of the table represents a provenance event, and each column represents an attribute associated with it, such as the parties involved (the sender, the receiver, and any intermediary agents), its location, its time, and the method by which the transfer was carried out.

Sender	Receiver	Agent	Location	Time (in EDTF)	Method of Transfer
Henri Matisse	Paul Rosenberg				purchase
Paul Rosenberg	Alexandre Rosenberg			..1948	
Alexandre Rosenberg	William Somerset Maugham			1950	sale
William Somerset Maugham	Colonel C. Michael Paul	Sotheby's	London	1962-04-10	auction
Colonel C. Michael Paul	Taft B. Schreiber			1970-01-15	sale
Taft B. Schreiber	Rita B. Schreiber				
Rita B. Schreiber	NGA			1989	gift

Fig. 3

The provenance of Henri Matisse, *Woman Seated in an Armchair*, structured in a table.

If we were to structure provenances in a standardized way on a large scale—say, all the provenance records from the National Gallery—it would be possible to query the data by analyzing the events (the rows of the tables) through the characteristics expressed by each

column. For example, it would be possible to formulate queries such as “Give me all the events in which Paul Rosenberg bought an object from Henri Matisse.” To answer this query, the machine can count all rows of all the tables in which “Paul Rosenberg” is the value in the “Receiver” column and “Henri Matisse” is the value in the “Sender” column. The value of structuring data thus lies in how it allows for more complex quantitative analysis. One can store a tabular data structure in proprietary formats such as Microsoft Excel or open formats such as CSV (Comma Separated Values).

The information’s semantics is not explicit in a table, however: the machine does not have a semantic understanding of any of the columns. It knows that “London,” for example, is in the “Location” column but is oblivious to the fact that it represents a location in space with administrative or geographic value (e.g., a city) that is part of a larger area (e.g., a region or a country). For this reason, even though the machine can perform a search with a given query, it cannot use the implicit semantic value of the table and therefore cannot combine it with external knowledge (e.g., data from other museums or repositories) to infer alternative knowledge. For example, the provenance of *Woman Seated in an Armchair* involves a 1962 Sotheby’s auction held in London, yet the machine has no understanding that London is in England. This information is not in the text, so it probably would not be included during the data structuring process. Nevertheless, associating the London entity in the provenance to the respective entity in the Getty Thesaurus of Geographical Names would add the notion that London is in England and is its current capital. With this additional data and the proper provenance knowledge modeling (which we will address in the next section), the machine could logically infer that the auction was held in England, more precisely in the capital, thereby increasing possibilities for analysis and research (e.g., on the art market in capital cities and the art market in England).

The museum could instead add this geographical information to its database once the provenance is structured. However, even for elementary notions such as “London is the capital of England,” such effort would require extra and unnecessary work to structure knowledge. Given that such information is already findable, accessible, interoperable, and reusable on the web in the form of LOD, it is best to simply make use of it.

## PROVENANCE LINKED OPEN DATA

As noted in the previous section, structured data can help provide museum professionals and researchers with quantitative insights into specific collections and collecting histories. However, when an individual museum structures its data in an idiosyncratic way that is not compatible with how other organizations have done it, its usefulness will be limited only to queries about its

collection. To overcome this limitation, museums will have to structure their data according to a set of shared principles that makes the data findable, accessible, interoperable, and reusable—that is, according to a set of principles known as the FAIR principles, which the scientific community established in 2016.[34]

However, for historical research (and provenance research is just that), the FAIR principles alone are insufficient for an inclusive, multi-perspectival approach, as these principles do not deal with open data's ethical and moral implications. While *accessible* data can be potentially *open*, it is not necessarily so.[35] Open data by definition “can be freely used, modified, and shared by anyone for any purpose,” and to produce open provenance data, therefore, we must apply a data standard that respects both the FAIR principles and the open principle.[36] That standard is LOD, which relies on structured data and can link data in a way that allows for complex and potentially valuable queries, especially across institutions.[37]

Publishing LOD, built on web standards, means publishing resources online and identifying them through URIs (universal resource identifiers, i.e., a unique name for a given resource), such as the URI of the Getty's Union List of Artist Names referencing Henri Matisse; <http://vocab.getty.edu/ulan/500017300>. In addition, the URIs used to identify LOD resources are HTTP URIs, that is, URIs associated with the hypertext transfer protocol (HTTP). This type of URI makes every LOD resource findable. The curation and preservation of URIs are two of the core responsibilities of linked open-data producers: indeed, the stability of URIs and their maintenance are a prerequisite for their long-term usefulness. The LOD community established another standard, RDF (resource description framework), to describe relationships between resources identified by URIs.[38] This standard relies on the fact that every entity and every relationship between entities in a given dataset can be identified by a discrete URI. In the context of provenance records, such entities can be people, organizations, objects, or events, and the relationship is what binds two such entities together, usually expressed by a verb (i.e., “Paul Rosenberg has French nationality”). As we have seen already, such descriptions come in the form of syntactical statements (i.e., sentences) based on a triple structure: subject—predicate—object. Thus, in our example, Paul Rosenberg is the subject (<http://vocab.getty.edu/ulan/500372940>) with the predicate of having a nationality (<http://schema.org/nationality>) that is French, which is specified as the object (<http://vocab.getty.edu/aat/300111188>).

The advantage of a shared standard based on these so-called triples is that they are structured and hence machine-readable, thus allowing for queries. Because these triples are constructed with stable URIs, we can analyze this data quantitatively not only within one museum collection but also in the context of the entire world wide web, where other museums can also publish



their information following the same standard. Furthermore, this syntax also allows us to make descriptions using URIs managed by multiple stakeholders, which means that the labor involved in producing provenance LOD is distributed across institutions because all producers can rely on the interoperability and reuse of each other's linked open data.<sup>[39]</sup>

To guarantee interoperability between one's own provenance LOD and that of other stakeholders, it is essential to build data according to not only LOD standards such as RDF but also to a shared community standard so that triples are built from a common set of URIs. This shared standard is CIDOC CRM (Conceptual Reference Model of the International Committee for Documentation). It is an ISO standard (ISO 21127) developed by museum professionals under the auspices of the International Council of Museums (ICOM) since 1996.<sup>[40]</sup> As a standardized ontology (a data model that defines what kinds of things make up a domain, and what kinds of relationships exist between them), it currently provides URIs for 160 properties (such as "P74 has current or former residence," useful to, e.g., describe that Paul Rosenberg has residence in Paris) and 81 classes, categories to which an entity can belong (e.g., Paul Rosenberg belongs to the class: "E21 Person").<sup>[41]</sup>

One of the most critical aspects of CIDOC CRM is its event-oriented modeling. Whereas the AAM format introduced earlier proposes a list of owners—an object-centric method—from which a chronology of events can be deduced, but does not build the provenance on events, provenance LOD (built on the CIDOC CRM standard) does. In other words, building LOD provenance data from AAM-formatted provenance records involves transposing an object-centered structure to an event-oriented structure that links people or organizations to events that may involve one or more objects. With this potential for knowledge production in mind, the Linked Art community is currently actively developing a data model built on CIDOC CRM that will cater to the specific needs of art museums. With a pared-down version of CIDOC CRM, Linked Art aims to encourage LOD implementation for museums that want their collection data to "be part of the Web, and not just on the Web."<sup>[42]</sup>

Despite its potential and ease of use, the Linked Art Data Model has not yet been applied across institutions in a real-world scenario.<sup>[43]</sup> In fact, only two projects that structure and link provenance data have been developed and published so far. In 2000, the ethnographic collection at the Museum of Cultural History in Oslo pioneered provenance data modeling when it implemented an event-oriented database for curation and research purposes using CIDOC CRM.<sup>[44]</sup> Even though LOD standards were far from defined, the Museum of Cultural History modeled 50,000 object records with 2 million events and 3.6 million relationships between events, objects, parties, locations, and time. The second project is Art Tracks, developed at the Carnegie Museum of Art in Pittsburgh between 2014 and 2017.<sup>[45]</sup> The

project's goal was to reconstruct the history of Old Master paintings from one collection, the Northbrook Collection of the Baring family, and then to visualize the collection's growth and later dispersal on a digital map and timeline for the benefit of gallery visitors and online users. Art Tracks was at the forefront of applying technologies of structuring and linking to the problem of generating well-formatted provenance data from free-text information based on the AAM format.<sup>[46]</sup> As of the time of writing, this extension of the AAM format proposed by Art Tracks is the only existing provenance text standard that anticipates machine readability.<sup>[47]</sup>

Both the CIDOC CRM and Linked Art communities and the Oslo and Art Tracks projects have begun to address challenges specific to provenance data, such as how to model gaps and uncertainties and how to record subjective assertions. Because of their importance for documentation and potential further research, these aspects have been discussed in the recent literature on digital provenance.<sup>[48]</sup> However, while some extant modeling techniques are waiting to be applied and tested on a large set of real-world provenance data, others still have to be conceptualized and defined.

Although LOD theoretically allows for modeling all historical information that can be found in and around provenance records, such as in notes or supplemental documents, it has not been determined what kind of provenance information should be selected for structuring and linking—and which should *not* since it is already available on the internet and would thus not constitute a smart use of resources. As mentioned above, massed data projects can be resource intensive; to quote computer scientist Ian Foster, “the creation, curation, maintenance, and delivery of digital information are all expensive and time-consuming activities.”<sup>[49]</sup> This is true as well for museum documentation and provenance data, within which complexity, biases, and subjectivity require human intervention when structuring and linking. This requisite human element makes a resource-conscious approach to producing provenance LOD even more crucial.

A strictly economical approach would limit the structuring and linking to data directly related to ownership changes, in a sense replicating the traditional concept of provenance as a list of different owners. On this approach, any additional information would not be part of the museum's linked provenance data. Such information includes custody changes of an object (as opposed to ownership changes) as well as biographical information about the people involved, including their birth and death dates, alternative spellings of their names, and the relationships among them. However, in the spirit of the LOD principles laid out above, such additional information could be sourced from external LOD repositories that depend either on crowdsourced knowledge, such as Wikidata, or on more scientifically reliable terminologies (or vocabularies) for the cultural heritage domain, especially those provided and edited by the

Getty Research Institute such as ULAN (for parties), AAT (for methods of transfer), and TGN (for historical locations).[50] Such crowdsourced platforms and authoritative institutions produce and maintain data, which museums can link to and which machines would be able to find and analyze in conjunction with provenance data from across museum collections. Using LOD from external sources in this manner addresses one of the significant pitfalls of digitizing provenance records in a structured way: the potential loss of information that is part of the provenance record but has nowhere to “live” in a structured environment—for example, biographical information that is not directly relevant to the event’s description. However, for the (art) historian, such “extra” information can be useful, and therefore museums might want to preserve it.

We now aim to show how digitized provenance can be structured based on LOD principles, once again using the provenance of Matisse’s *Woman Seated in an Armchair*. In its current form, this provenance is recorded in the syntactical AAM format, which tells us (focusing for a moment only on the first provenance event listed) that Paul Rosenberg purchased the painting from Henri Matisse. However, the provenance specifies neither the location nor the time of this event. On the other hand, it gives biographical information for both parties involved. Matisse is called “the artist,” and Rosenberg’s location is given as Paris, while his occupation is listed as “art dealer.” The museum gives that information in an institution-specific style using parentheses.[51]

Both parties (Matisse and Rosenberg) and the method of transfer (purchase) are already recorded in ULAN or the AAT. Happily, machine-learning methods can assist humans in structuring and linking provenance events and their core entities. Already, semi-automated data services, such as the Getty Vocabularies OpenRefine reconciliation tool, can help link the event’s entities to ULAN and the ATT.[52] With such a resource-conscious approach, the structuring effort would be relatively small but deliver considerable results. In the case of this example, perhaps even more important than the cost-effectiveness would be the fact that Paul Rosenberg, art dealer, would be indexed by a stable URI that would not only tie the object to the “right” Paul Rosenberg but immediately put it in relation to all other objects tied to this Paul Rosenberg. While not error-proof (authority files are not 100% accurate all the time), if multiple museums would take even a limited LOD approach, it would, with little effort but high reliability, enable researchers to find objects purchased by Paul Rosenberg directly from Matisse (or more generally from artists), across these museum collections.

In light of these considerations, we can model the information of the first provenance event described by the National Gallery of Art—that is, Paul Rosenberg’s purchase of the painting from Henri Matisse—and show it in a diagram (fig. 4).[53] Using this limited example for demonstration purposes, the information is structured using the Linked Art Data Model based

on CIDOC CRM; we have reused and linked the ULAN entities for Henri Matisse and Paul Rosenberg; and the acquisition method, “purchased,” comes from ATT. The resulting structured data is thus provenance LOD.

```
graph TD
classDef object
stroke:black,fill:#E1BA9C,rx:20px,ry:20px; classDef actor
stroke:black,fill:#FFBDCA,rx:20px,ry:20px; classDef type
stroke:red,fill:#FAB565,rx:20px,ry:20px; classDef name
stroke:orange,fill:#FEF3BA,rx:20px,ry:20px; classDef dims
stroke:black,fill:#c6c6c6,rx:20px,ry:20px; classDef infoobj
stroke:#907010,fill:#fffa40,rx:20px,ry:20px; classDef timespan
stroke:blue,fill:#ddfffe,rx:20px,ry:20px; classDef place
stroke:#3a7a3a,fill:#aff090,rx:20px,ry:20px; classDef event
stroke:#1010ff,fill:#96e0f6,rx:20px,ry:20px; classDef literal
```

Fig. 4

The purchase of *Woman Seated in an Armchair* by Paul Rosenberg from Henri Matisse, described in a diagram using the Linked Art data model.

The more provenance research scholars conduct, the more complicated the object histories become, which we can observe in the data. For example, figure 2 shows the complexity of Paul Rosenberg’s ownership, recorded in a note longer than the entire provenance record. From the note, we learn that the painting was confiscated by the ERR in 1941, together with other paintings from two of Paul Rosenberg’s storage locations, either in Libourne or at a chateau at Floirac. It was then traded together with another painting by another artist from another looted collection in France to a dealer and was later seen in Switzerland before eventually being returned to Paul Rosenberg. In a strict understanding of provenance as a sequence of mere ownership changes and by applying US legal standards according to which looted paintings do not change title (as is the provenance practice at the National Gallery of Art), this information would be omitted because Paul Rosenberg was the painting’s only owner through all the cruel twists and turns of World War II. However, from a holistic perspective, secondary provenance such as the painting’s changes in location provide vital insights. Just imagine, for example, the possibilities for provenance research related to National Socialism and the potential queries one could bring to the data: the objects that have been identified as being looted from one of Paul Rosenberg’s storage spaces; the objects that were exchanged with and against each other by the ERR; the objects that have (or have not) been restituted to the Paul Rosenberg family.

Discarding well-documented and researched provenance information when moving or remodeling provenances in line with LOD principles would not constitute good scientific practice and would go against the AAM recommendations for publishing provenance information.<sup>[54]</sup> On the other hand, structuring these additional events—with their parties, transfers, locations, and times, plus the additional detail, some of it incomplete, vague, uncertain, or subjective—increases the overall effort needed to transition to structured provenance data and cannot be done without significant expert intervention. These hurdles notwithstanding (including those that have yet to be fully addressed from a modeling perspective, such as uncertainty), we believe that this extra effort is not only desirable but indeed necessary for an expanded notion of provenance that counters the risk of reduction

implicit in the AAM format.

Provenance LOD is nowhere to be found today, as museums are still by and large merely digitizing provenance records without introducing *any* kind of machine-readable structure. Yet the kind of modeling that LOD allows for is, as we have just seen, not without its own pitfalls: it requires, above all, significant investment of resources in labor and data infrastructure, and it needs to account for the kinds of omissions, biases, and reductions (as well as complexities) that legacy provenance records contain.

## STRATEGIZING PROVENANCE DATA

To help facilitate museums' transitions to provenance LOD (PLOD), we will now propose a conceptual framework for what data to model, and to what level of detail, within a museum dataset. We think of this PLOD conceptual framework not as a strict roadmap but rather a blueprint for formulating a conscious, responsive, and sensitive data strategy, and, because it is conceptual in nature, it can be applied regardless of the data model. We designed it with the application profile of the Linked Art Data Model in mind, as we consider it the current benchmark. The approach that we detail below begins, necessarily, with a base layer of information, capturing the provenance's core entities. A system of layers built from descriptive bricks complements the base layer to allow for a thicker description of the data, thus improving its quality and usefulness. This modularity addresses the need for a compromise between the resources to be invested in digitization and the problem of losing or flattening data. We again emphasize that the possibility of reusing LOD resources from external authorities such as ULAN can obviate the need to describe an entity from scratch.

A provenance record should, in theory, be understood as a sequence of provenance events in chronological order, each comprising one or more transfers, which occur between parties at a given location at a given time. Thus, five entities are required for the base layer: provenance event, parties, transfers, location, and time. The *provenance event* is a meta entity to which the other information is gradually associated. *Parties* are people, organizations, or any constituents involved in the provenance event, acting alone or in groups either as a sender (the one who loses ownership or custody), a receiver (the one who obtains ownership or custody), or as a mediating agent for one or the other. *Transfers* convey either ownership or custody changes by such methods as inheritance, gift, purchase, exchange of objects, looting, or restitution. A provenance event can have one transfer from a sender to a receiver, or it can have multiple transfers because multiple objects changed hands at the same location and the same time, such as an auction like the one held by Sotheby's in London in 1962. The *location* of this provenance event is London. In practice, such a location is usually only recorded and relevant for

provenance events with multiple transfers (e.g., an auction). Locations needed for tracing an object's spatial movement tend to be recorded in provenances through the respective parties involved (e.g., through their places of residence). *Time* indicates when the provenance event took place, recorded in the internationally used, unambiguous calendar-and-clock format (ISO 8601). Building the base layer is easy enough for museum practitioners, but they could also apply crowdsourcing or machine-learning methods to existing provenance texts, although that would require a more extensive—but entirely realistic—implementation effort.<sup>[55]</sup>

In line with their ideals of inclusivity and in the spirit of decentering the individual object, museums should consider including additional information that has often been harvested through time-consuming provenance research and can be crucial for object documentation and data analysis. In our framework, four types of bricks are available for structuring advanced information and enhancing the data: biographical, economic, geographic, and contextual (see fig. 5).

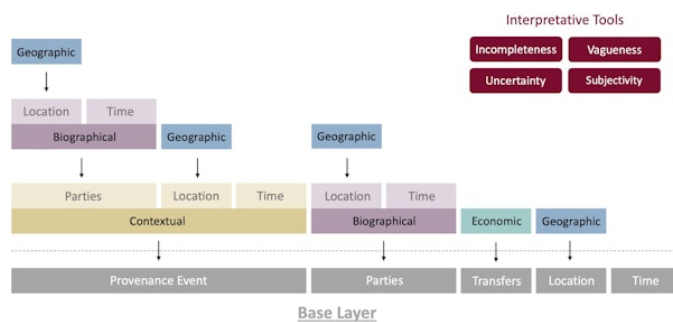


Fig. 5

The PLOD conceptual framework with the various options for descriptive bricks and interpretative tools.

Provenance records often contain biographical information about the parties that may fall by the wayside during structuring. For example, if birth and death dates have served to disambiguate a person's identity in the provenance text, this data becomes unnecessary in PLOD. If these dates indicate the time of a transfer like an inheritance, the biographical data should be directly linked to the base layer party, either in the museum's dataset or from a trustworthy vocabulary like ULAN. Similar decisions must be made for other biographical aspects of individuals as well as groups and organizations, such as detailed onomastic descriptions, gender, nationality, religion, or life events and relationships between parties, which can, in turn, be modeled with a location and time. Such biographical data allows provenance aspects, like the buying of Matisse paintings, to be brought into dialogue with demographic analysis, which can be especially meaningful for the histories of collecting and taste.

The biographical brick can be used to correct the records of historically marginalized groups that are inadequately represented in LOD. Museums can and should actively contribute this data, becoming a de facto authority for a given digital resource. The Museum of Modern Art in

New York, as the holder of the gallery archives of Paul Rosenberg, for example, might want to create its own “Paul Rosenberg” entity with its own URI linked to the one already in ULAN but described in more detail, thereby making this information available to the community.

All transfers of cultural objects have an economic dimension, whether prices or values are expressed explicitly or not. Some museums tend not to share prices publicly. The AAM, however, recommends recording this information for objects transferred in Europe during National Socialism, as this may be relevant for assessing, for example, whether an artwork was a forced sale or not (particularly if the information includes details about currencies, discounts, and taxes).<sup>[56]</sup> Especially in the art market, transfers often go hand in hand with complex economic features such as joint ownership or bidding at auctions. Considering these details when structuring data lays the groundwork for financial analyses of provenances that would be meaningful for studies of the art market and economic trends, and potentially for identifying illegal acts (tax evasion, money laundering, black market deals).<sup>[57]</sup>

*Paris, Texas* is a film directed by Wim Wenders and, in a geographical sense, a city in the state of Texas named like the (current) capital of France. In a PLOD context, this presents a disambiguation challenge not unlike the one we encountered with parties. While this is often solvable through linking to shared vocabularies in the base layer, there are biases in the documentation of provenance research on this topic that must be addressed, for example with respect to under-recorded African towns and locations with complicated histories, such as Lviv in present-day Ukraine (but previously part of other countries and empires).<sup>[58]</sup> Geographic bricks can help carry that weight. Adding geospatial data is crucial for mapping and analysis, and data on administrative hierarchies and demographic aspects such as census and density can be meaningful for studying regional trends and patterns of collecting and taste.

Historical events such as Europe’s colonial occupation of Africa, the Great Depression, and World War II, and the related German occupation of France affected transfers of ownership of artworks and are meaningful contextual information relevant to provenance records. At present, objects on museum websites are manually flagged as having changed ownership during National Socialism. By modeling the various territories occupied by the National Socialists between 1939 and 1945, the machine could precisely identify which objects changed ownership in this context and are therefore potentially relevant for potential claimants to this day. Other political, cultural, social, economic, and environmental situations of historical importance that have or might have influenced provenance events could also be meaningful to model for more extensive analysis—although, as we addressed earlier, modeling such historical events with hundreds if not thousands of subevents with parties, locations, and dates, though possible in this brick and its related layers, should ideally live in crowdsourced or authoritative vocabularies and be



maintained by a larger community than simply cultural heritage institutions (see fig. 6).

Brick	Scope	Descriptive Range
Biographical	Parties	<p>Onomastic (e.g., title, first name, family name, maiden name)</p> <p>Gender</p> <p>Nationality</p> <p>Religion</p> <p>Biographical events (e.g., people's birth and death, group's formation and dissolution, change of location)</p> <p>Relationships between people (e.g., family ties), between groups (e.g., a department of a museum), or between people and groups (e.g., a person's membership in a group)</p>

Brick	Scope	Descriptive Range
Economic	Transfers	<p>Financial features (e.g., prices, currencies, sales discounts, profit margin, commission, sale tax)</p> <p>Financial activities (e.g., payments, loans)</p> <p>Auction activities (e.g., bidding, withdrawing)</p> <p>Inventorying activities (e.g., insurance values, inventory numbers, depreciation)</p> <p>Joint-ownerships</p>
Geographic	Location, Provenance Event, Parties	<p>Geospatial features (e.g., geographical coordinates or shapes)</p> <p>Administrative hierarchy (e.g., the country, region, province, and/or city of the location)</p> <p>Historical names</p> <p>Demographic aspects (e.g., census, density)</p>

Brick	Scope	Descriptive Range
Contextual	Provenance Event	Historical events of: Social dimension (e.g., Olympic Games, World Fairs) Political dimension (e.g., regimes, occupation) Cultural dimension (e.g., epochs, fashions, ideology) Economic dimension (e.g., depression, boom) Environmental dimension (e.g., floods, fires, pandemics)

Fig. 6

Overview of the biographical, economic, geographic, and contextual bricks available in the PLOD conceptual framework.

History is rarely straightforward, and neither are our records of it—which, from a data perspective, is unfortunate. Historical knowledge in provenance is often incomplete (“unknown buyer,” “private collection”), vague (“around,” “circa”), uncertain to varying degrees (“probably,” “possibly”), and subjective (“according to”). This dimension of knowledge is crucial for object documentation, interpretation, querying, and analysis of provenance. Hence, this incompleteness, vagueness, uncertainty, and subjectivity should be captured in PLOD, so it is available not only for humans but in a machine-readable way for use within and ideally across museum collections. Interpretative tools, as envisioned in our framework, should be applied at whatever scale they may prove useful in a brick-modeled provenance.

On the other hand, we understand the four bricks as options that museums can use based on their priorities according to their (provenance) mission and their (provenance) data. For example, a museum might begin the publication of PLOD by focusing on the core entities of the provenance and additional biographical information (base layer and biographical brick), leaving the description of economic details (economic brick) for a later stage—especially given that the Getty Provenance Index, currently transforming millions of auction and art dealer data to LOD, might soon provide these details in a structured and linkable format, ready to be employed by

museums.

Indeed, our framework's modularity allows museums to address the inconsistency of their data. A move to PLOD will not fix this common issue, but it offers a chance to reframe it, and provenance along with it. In a field that has never been stable or constant, our framework aims to assist museums in making resource-conscious decisions so that perhaps their data strategies remain in line with their evolving missions.

## CONCLUSION

Museums and their many internal and external stakeholders are faced today with innumerable and often competing demands, and they cannot address all such demands with the same care that they may warrant. Museums are also political actors because they tell stories about the past. These political pressures extend to museum data policy, including, but certainly not limited to, the question of provenance data: its production, long-term care, and accessibility. Through its web structure and commitment to openness, LOD can help begin to address concerns around these issues. The inevitable adoption of linked open-data standards in provenance is thus both a challenge and an opportunity. Its benefits outweigh the costs, for when done carefully and with a well-conceived data strategy, the move to PLOD can help museums pursue their goals of transparency, accountability, and inclusivity. It can also help them address epistemic shifts and allow for a multi-perspectival but standardized and structured data practice. Finally, it is paramount that museums not only acknowledge but fully embrace the fact that recording and publishing provenance is a form of writing history. Whether museums rise to the challenge remains to be seen, but those that do will indeed write history.

---

Banner image: Detail of fig. 4.

---

## NOTES

1. Chaired by Robert Sanderson of Yale University and Emmanuelle Delmas-Glass of the Yale Center for British Art, the Linked Art community is a consortium of people working with cultural heritage data. The community is currently comprised of twenty-four institutions based primarily in North America and Europe. Lynn Rother has served on the editorial board of Linked Art since January 2019. See <https://linked.art/>.

The authors want to thank the editorial team, the anonymous peer reviewer, and Duane Degler for generously sharing their experience and valuable feedback. Margaret Doyle from the National Gallery of Art has graciously helped with procuring a crucial illustration. Thanks also go to Amy R. Peltz for editing this essay and to Liza Weber for her editing.

2. See Kelly Davis, “Old Metadata in a New World: Standardizing the Getty Provenance Index for Linked Data,” *Art Libraries Journal* 44, no. 4 (October 2019): 162–66, <https://doi.org/10.1017/alj.2019.24>; and Maximilian Schich et al., “Network Dimensions in the Getty Provenance Index,” arXiv.org, June 8, 2017, <http://arxiv.org/abs/1706.02804>.
3. Daniel Grana-Behrens, “Digitalbasierte Ethnologische Provenienzforschung: Chancen und Herausforderungen am Beispiel WissKI der Bonner Amerikas-Sammlung (BASA-Museum),” in *Digitalisierung Ethnologischer Sammlungen*, ed. Hans Peter Hahn, Oliver Lueb, Katja Müller, and Karoline Noack (Bielefeld, Germany: Transcript Verlag, 2021), 215–38, <https://doi.org/10.1515/9783839457900-013>; Babette Claassen et al., “Linked Art Provenance,” in *Proceedings of the Network Institute Academy Assistants Programme 2018–2019*, August 27, 2020, <https://doi.org/10.5281/zenodo.4003499>; Christian Huemer, “The Provenance of Provenances,” in *Collecting and Provenance: A Multidisciplinary Approach*, ed. Jane Milosch and Nick Pearce (Lanham, MD: Rowman and Littlefield, 2020), 2–15; Matthew Lincoln and Sandra van Ginhoven, “Modeling a Fragmented Archive: A Missing Data Case Study from Provenance Research,” *Digital Humanities 2018: “Puentes/Bridges,”* June 21, 2018, <https://dh2018.adho.org/en/modeling-the-fragmented-archive-a-missing-data-case-study-from-provenance-research/>; Jodi Cranston, “Mapping Paintings, or How to Breathe Life Into Provenance,” in *The Routledge Companion to Digital Humanities and Art History*, ed. Kathryn Brown (London: Routledge, 2020), 109–19; Anne Luther, “Digital Provenance, Open Access, and Data-Driven Art History,” in Browne, *Routledge Companion*, 448–58; David Newbury and Louise Lippincott, “Provenance in 2050,” in Milosch and Pearce, eds., *Collecting and Provenance*, 101–9; Jeffrey Smith, “Toward ‘Big Data’ in Museum Provenance,” in *Big Data in the Arts and Humanities: Theory and Practice*, ed. Giovanni Schiuma and Daniela Carlucci (New York: Auerbach, 2018), 41–50; and Steven Kuhnen et al., “Structuring Cultural Heritage PROVenance: The Rijksmuseum Use Case” (presentation, DHBenelux, Amsterdam, Netherlands, June 6–8, 2018).
4. Machine readability refers to information’s ability to be read and analyzed by a computer automatically. Non-digital material, such as printouts or handwritten letters, are not machine-readable. But digital material, such as a JPG image file showing text, can also be non-machine-readable. To the computer, such digital material constitutes an image that it cannot automatically read and process as a text. If the text is stored differently, e.g. in word-processing software, it is machine readable. While all machine-readable texts have some structure,

structured data refers to data where the relation between textual elements is explicit in the way the data is stored. This means that the logic that is embedded in texts and understandable by humans, is made explicit to the computer by creating a structured and machine-readable representation, e.g., a table, of the text's logic.

5. Fabio Mariani, "'Probably Sold to Paalen, Possibly by Exchange': Vagueness, Incompleteness, Subjectivity and Uncertainty in Digital Art Provenance" (paper delivered at the Workshop on Computational Methods in the Humanities, June 10, 2022), available at [https://wp.unil.ch/lilist/files/2022/06/COMHUM\\_2022\\_paper\\_5.pdf](https://wp.unil.ch/lilist/files/2022/06/COMHUM_2022_paper_5.pdf).
6. On data literacy, see Harald Klinke, "The Digital Transformation of Art History," in Browne, *Routledge Companion*, 32–42.
7. Nancy H. Yeide, Konstantin Akinsha, and Amy L. Walsh, *The AAM Guide to Provenance Research* (Washington, DC: American Association of Museums, 2001).
8. Gail Feigenbaum and Inge Reist, introduction to *Provenance: An Alternate History of Art* (Los Angeles: Getty Research Institute, 2013), 1.
9. See Huemer, "The Provenance of Provenances."
10. See Elizabeth A. Pergam, "Provenance as Pedigree. The Marketing of British Portraits in Gilded Age America," in Feigenbaum and Reist, *Provenance*, 104–22; and Pierre Bourdieu, *Distinction: A Social Critique of the Judgement of Taste*, trans. Richard Nice (Cambridge, MA: Harvard University Press, 1984).
11. Feigenbaum and Reist, "Introduction," 1–4.
12. Sonja Niederacher, *Eigentum und Geschlecht: jüdische Unternehmerfamilien in Wien (1900–1960)* (Vienna: Böhlau, 2012). In 2021 Stanford University's Archeological Collections published an online exhibition titled *Women in Provenance*, available at <https://storymaps.arcgis.com/stories/5bf914cc05164cd2a7758457567f7c33#ref-n-4YiOcR>, accessed December 12, 2022.
13. Susan Elizabeth Gagliardi, "Mapping Senufo: Mapping as a Method to Transcend Colonial Assumptions," in Browne, *Routledge Companion*, 135–54. Gagliardi highlights the problematic character of attributing authorship of an African bronze statue to an entire ethnic group given that a European bronze statue is often easily attributed to a single creator such as Picasso.
14. Anne Higonnet, "Afterword: The Social Life of Provenance," in Feigenbaum and Reist, *Provenance*, 197.
15. See, for example, the provenance for the Edo Queen Mother Pendant Mask at the Metropolitan Museum of Art, 1978.412.323, <https://www.metmuseum.org/art/collection/search/318622>, accessed on Oct 14, 2022.

16. See Leila Amineddoleh, "The Role of Provenance in Resolving Art-World Disputes," in *Provenance Research Today: Principles, Practice, Problems*, ed. Arthur Tompkins (London: Lund Humphries, 2020), 25–38; and Lynn Rother and Iris Schmeisser, "Provenance Research in Museums: The Long Run," in Tompkins, *Provenance Research Today*, 106–16.
17. US Department of State, Office of the Special Envoy for Holocaust Issues, "Washington Conference Principles on Nazi-Confiscated Art," December 3, 1998, <https://www.state.gov/washington-conference-principles-on-nazi-confiscated-art/>.
18. See Gesa Grimme, "Systemizing Provenance Research on Objects from Colonial Contexts," *Museum and Society* 18, no. 1 (March 2020): 52–65; and Christoph Zuschlag, "Vom Iconic Turn zum Provenancial Turn? Ein Beitrag zur Methodendiskussion in der Kunstwissenschaft," in *Von analogen und digitalen Zugängen zur Kunst: Festschrift für Hubertus Kohle zum 60. Geburtstag*, ed. Maria Effinger, Stephan Hoppe, Harald Klinke, and Bernd Krysmanski (Heidelberg, Germany: Arthistoricum.net, 2019), 409–17, <https://doi.org/10.11588/arthistoricum.493.c6573>.
19. "3-Road Strategy on the Documentation and Digital Publication of Collections from Colonial Contexts Held in Germany, German Contact Point for Collections from Colonial Contexts, <https://www.cp3c.org/3-road-strategy/>.
20. Especially in the German-speaking countries, a vast amount of provenance scholarship related to National Socialism has been produced over the last twenty years. For a critical take on this development, see Christian Fuhrmeister and Meike Hopp, "Rethinking Provenance Research," *Getty Research Journal* 11 (2019): 213–31; and Marc Masurovsky, "The Current State of Nazi-Era Provenance Research, and Access to Nazi-Era Research Resources and Archives," in Tompkins, *Provenance Research Today*, 136–49. Specifically for the US context, see "Provenance Research in American Institutions," ed. Jane C. Milosch, Megan M. Fontanella, and Lynn H. Nicholas, special issue, *Collections: A Journal for Museum and Archives Professionals* 10, no. 3 (2014).
21. See, for example, the confiscation inventory database of "Entartete Kunst" (degenerate art), Freie Universität Berlin (<https://www.geschkult.fu-berlin.de/e/khi/ressourcen/diathek/beschlagnahmeinventar/index.html>) and "Cultural Plunder: Database of Art Objects at the Jeu de Paume" by the Einsatzstab Reichsleiter Rosenberg (<https://www.errproject.org/jeudepaume/>). See also Uwe Fleckner, "Dubious Business: Trade in Modern Art under the 'Third Reich,'" in Bresciani and Hansen, *Looters, Smugglers, and Collectors*; Stephanie Barron, ed., "Degenerate Art": *The Fate of the Avant-Garde in Nazi Germany* (Los Angeles: Los Angeles County Museum of Art, 1991); Lynn H. Nicholas, *The Rape of Europe: The Fate of Europe's Treasures in the Third Reich and the Second World War* (New York: Vintage, 1995).
22. For example, the Presidential Advisory Commission on Holocaust Assets, the American



Alliance of Museums (former American Association of Museums), and the Association of Art Museum Directors established guidelines regarding objects misappropriated during the National Socialist Era in 1999 that recommended an initial focus on European paintings and Judaica. See <https://www.aam-us.org/programs/ethics-standards-and-professional-practices/unlawful-appropriation-of-objects-during-the-nazi-era/>, accessed Oct 14, 2022.

23. Yeide, Akinsha, and Walsh, *The AAM Guide to Provenance Research*.
24. The full tombstone information for this work, as given on the National Gallery website, is as follows: Henri Matisse, *Woman Seated in an Armchair*, 1940. Oil on canvas. National Gallery of Art, Washington DC, Given in loving memory of her husband, Taft Schreiber, by Rita Schreiber, 1989.31.1, <https://www.nga.gov/collection/art-object-page.71071.html>. The National Gallery is, in fact, one of the few museums that enters the names of former owners in not only its free text provenance field but also the “Constituent Assistant,” provided by the collection management software The Museum System (by Gallery Systems), allowing searchability for names.
25. For example, at the Museum of Modern Art in New York, there exist different provenance formats for the same artist, some making use of the AAM format and some not. This is the product of different authors working at different times. Compare, for instance, the provenance for two works by Giorgio de Chirico: <https://www.moma.org/collection/works/78738> and <https://www.moma.org/collection/works/80588>, accessed October 14, 2022.
26. Efforts at decolonizing the museum are not entirely new developments and can be traced back to the post–World War II era. See Claire Wintle, “Decolonising the Museum: The Case of the Imperial and Commonwealth Institutes,” *Museum and Society* 11, no. 2 (July 2013): 185–201.
27. The development of NLP (Natural Language Processing) techniques over the past fifty years allows for extracting meaning from unstructured data. In the context of the Provenance Lab at Leuphana University Lüneburg, the authors of this essay are developing statistical models to apply NLP techniques on unstructured provenance text to extract meaning and facilitate structuring. See Lynn Rother, Fabio Mariani, and Max Koss, “Hidden Value: Provenance as a Source for Economic and Social History,” *Economic History Yearbook, Special Issue on Digital History* (forthcoming May 2023).
28. It is noteworthy that in 2003 the AAM launched the Nazi-Era Provenance Internet Portal to facilitate the search for objects potentially affected by National Socialist looting across US museum collections. However, while the portal provides object information and links to the museum websites, it does not provide provenance information and museums have not updated their registered objects. See <https://nepip.org>. In 2022 the German government launched a database for holdings from colonial contexts in German museums, recording 6,636

objects. See <https://ccc.deutsche-digitale-bibliothek.de/?lang=en>, accessed October 14, 2022.

29. With regard to exhibition data and how the computational analysis can provide new insights to art history, see Diana Seave Greenwald, *Painting by Numbers: Data-Driven Histories of Nineteenth-Century Art* (Princeton, NJ: Princeton University Press, 2021); and Béatrice Joyeux-Prunel and Olivier Marcel, “Exhibition Catalogues in the Globalization of Art: A Source for Social and Spatial Art History,” *Artl@s Bulletin* 4, no. 2 (Fall 2015): 26.
30. Kristin Weber-Sinn and Paola Ivanov, “‘Collaborative’ Provenance Research: About the (Im)Possibility of Smashing Colonial Frameworks,” *Museum and Society* 18, no. 1 (March 2020): 66–81, <https://doi.org/10.29311/mas.v18i1.3295>; and Fuhrmeister and Hopp, “Rethinking Provenance Research,” 213–31.
31. See Joshua A. Bell, Kimberly Christen, and Mark Turin, “After the Return: Digital Repatriation and the Circulation of Indigenous Knowledge Workshop Report,” *Museum Worlds: Advances in Research* 1, no. 1 (July 2013): 195–203, <https://doi.org/10.3167/armw.2013.010112>.
32. See, e.g. Rother, Mariani, and Koss, “Hidden Value.”
33. See <https://www.loc.gov/standards/datetime/>.
34. Mark D. Wilkinson et al., “The FAIR Guiding Principles for Scientific Data Management and Stewardship,” *Scientific Data* 3, no. 1 (March 2016): 160018, <https://doi.org/10.1038/sdata.2016.18>.
35. Barend Mons et al., “Cloudy, Increasingly FAIR: Revisiting the FAIR Data Guiding Principles for the European Open Science Cloud,” *Information Services & Use* 37, no. 1 (March 2017): 49–56, <https://doi.org/10.3233/ISU-170824>.
36. “Open Data,” Open Definition 2.1, Open Knowledge Foundation, <https://opendefinition.org/>.
37. The potential usefulness of LOD is perhaps best summarized by Tim Berners-Lee, the co-inventor of the world wide web: “With linked data, when you have some of it, you can find other, related data.” Tim Berners-Lee, “Linked Data,” W3.org, <https://www.w3.org/DesignIssues/LinkedData.html>.
38. “Resource Description Framework (RDF): Concepts and Abstract Syntax,” W3.org, February 10, 2004, <https://www.w3.org/TR/rdf-concepts/>.
39. *Interoperability* is the technical integration of data (the structure of other data is compatible with one’s own), whereas *reusability* refers to the legal integration of data, i.e. the rights related to that data (what can and cannot be done with that data).

40. Chryssoula Bekiari et al., eds, *Definition of the CIDOC Conceptual Reference Model*, version 7.1.1, International Committee for Documentation (CIDOC), April 2021, [http://www.cidoc-crm.org/sites/default/files/cidoc\\_crm\\_v.7.1.1\\_0.pdf](http://www.cidoc-crm.org/sites/default/files/cidoc_crm_v.7.1.1_0.pdf).
41. For the example of a property such as “P74 has current or former residence,” see [http://cidoc-crm.org/cidoc-crm/7.1.1/P74\\_has\\_current\\_or\\_former\\_residence](http://cidoc-crm.org/cidoc-crm/7.1.1/P74_has_current_or_former_residence) and for an example of a class such as “E21 Person”, see [http://cidoc-crm.org/cidoc-crm/7.1.1/E21\\_Person](http://cidoc-crm.org/cidoc-crm/7.1.1/E21_Person).
42. “Linked Art Profile of CIDOC-CRM” Linked Art, <https://linked.art/model/profile/>.
43. In 2019 the Georgia O’Keeffe Museum in Santa Fe became the first museum to apply the Linked Art Data Model across its collections under the auspices of Liz Neely from the Georgia O’Keeffe Museum and Duane Degler from Design for Context. See their Linked Data documentation here: <http://gokm-docs.okeeffemuseum.org/>. We must note that while provenance information was included, it was not structured.
44. Ellen Jordal, Espen Uleberg, and Brit Hauge, “Was It Worth It? Experiences with a CIDOC CRMbased Database,” in *Revive the Past: Computer Applications and Quantitative Methods in Archaeology (CAA), Proceedings of the 39th International Conference, Beijing, April 12–16, 2011*, ed. Mingquan Zhou et al. (Amsterdam, Netherlands: 2012), 255–60.
45. David Newbury, “Art Tracks: Using Linked Open Data for Object Provenance in Museums” (presentation, Museums and the Web 2017, Cleveland, Ohio, April 19–22, 2017), <https://mw17.mwconf.org/paper/art-tracks-using-linked-open-data-for-object-provenance-in-museums/>.
46. Art Tracks operates with a format on their own, which they call a superset of the AAM format.
47. See “The CMOA Digital Provenance Standard,” draft version 0.2, Art Tracks: A Project of the Carnegie Museum of Art, October 14, 2016, <http://www.museumprovenance.org/reference/standard/>.
48. See especially Lincoln and van Ginhoven, “Modeling a Fragmented Archive.”
49. Ian Foster, “How Computation Changes Research,” in *Switching Codes: Thinking through Digital Technology in the Humanities and the Arts*, ed. Thomas Bartscherer and Roderick Coover (Chicago: University of Chicago Press, 2011), 32.
50. ULAN is the Getty Union List of Artist Names, AAT is the Getty Art & Architecture Thesaurus, and the TGN is the Getty Thesaurus of Geographical Names. See <https://www.getty.edu/research/tools/vocabularies/>.
51. “Reading Collection Information: Provenance Texts,” National Gallery of Art, <https://www.nga.gov/collection/collection-information.html>.

52. OpenRefine is an open-source tool for data clean up and transformation. See <https://www.getty.edu/research/tools/vocabularies/obtain/openrefine/>.
  53. We created the diagram using the Mermaid JS library (<https://mermaid-js.github.io/>), adapting Linked Art's layout and style (<https://linked.art/model/intro/>).
  54. In fact, the AAM recommends including this data when museums provide information to the public about objects transferred in Europe during National Socialism. American Alliance of Museums, "Recommended Procedures for Providing Information to the Public about Objects Transferred in Europe during the Nazi Era," <https://www.aam-us.org/wp-content/uploads/2018/01/nepip-recommended-procedures.pdf>.
  55. Johan Oomen and Lora Aroyo, "Crowdsourcing in the Cultural Heritage Domain: Opportunities and Challenges," in *C&T 2011: Proceedings of the 5th International Conference on Communities and Technologies* (Brisbane, Australia: Association for Computing Machinery, 2011), 138–49, <https://doi.org/10.1145/2103354.2103373>. See also Newbury, "Art Tracks," which proposes a similar approach for the base layer.
  56. American Alliance of Museums, "Recommended Procedures for Providing Information."
  57. For an account of how provenance data can be made useful for social and economic historical analysis, see Rother, Mariani, and Koss, "Hidden Value."
  58. See Mark Graham and Stefano De Sabbata, "Mapping Information Wealth and Poverty: The Geography of Gazetteers," *Environment and Planning A* 47, no. 6 (June 2015): 1254–64, <https://doi.org/10.1177/0308518X15594899>.
- 

## HOW TO CITE

Lynn Rother, Max Koss, and Fabio Mariani, "Taking Care of History: Toward a Politics of Provenance Linked Open Data in Museums," in *Perspectives on Data*, ed. Emily Lew Fry and Erin Canning (Art Institute of Chicago, 2022).

This contribution has been peer reviewed through a double-anonymized process.

© 2022 by The Art Institute of Chicago. This work is licensed under a CC BY-NC 4.0 license: <https://creativecommons.org/licenses/by-nc/4.0/>

<https://doi.org/10.53269/9780865593152/06>