

Lynn Rother\*, Fabio Mariani and Max Koss

# Hidden Value: Provenance as a Source for Economic and Social History

## Stille Reserven: Provenienzen als Quelle für Wirtschafts- und Sozialgeschichte

<https://doi.org/10.1515/jbwg-2023-0005>

**Abstract:** Building on the extensive production of provenance data recently, this article explains how we can expand the purview of computational analysis in humanistic and social sciences by exploring how digital methods can be applied to provenances. Provenances document chains of events of ownership and socio-economic custody changes of artworks. They promise statistical and comparative insights into social and economic trends and networks. Such analyses, however, necessitate the transformation of provenances from their textual form into structured data. This article first explores some of the analytical avenues aggregate provenance data can offer for transdisciplinary historical research. It then explains in detail the use of deep learning to address natural language processing tasks for transforming provenance text into structured data, such as Sentence Boundary Detection and Span Categorization. To illustrate the potential of this pioneering approach, this article ends with two examples of preliminary analysis of structured provenance data.

**JEL-Codes:** N 01, Z 11

**Keywords:** art, art markets, artificial intelligence, deep learning, digital methods, gender, inheritance, museums, natural language processing, provenance, provenance data, value formation, wealth. Kunst, Kunstmärkte, Künstliche Intelligenz, Deep Learning, Digitale Methoden, Gender, Erbschaft, Museen, Natural Language Processing, Provenienz, Provenienzen, Provenienzen, Wertbildung, Reichtum

---

**\*Corresponding author: Lynn Rother** (Prof. Dr.), Institut für Philosophie und Kunstwissenschaft, Leuphana Universität Lüneburg, Universitätsallee 1, C5.411, D-21335 Lüneburg, E-mail: [lynn.rother@leuphana.de](mailto:lynn.rother@leuphana.de)

**Fabio Mariani** (MA), Institut für Philosophie und Kunstwissenschaft, Leuphana Universität Lüneburg, Universitätsallee 1, C5.418, D-21335 Lüneburg, E-mail: [fabio.mariani@leuphana.de](mailto:fabio.mariani@leuphana.de)

**Max Koss** (Dr.), Institut für Philosophie und Kunstwissenschaft, Leuphana Universität Lüneburg, Universitätsallee 1, C5.418, D-21335 Lüneburg, E-mail: [max.koss@leuphana.de](mailto:max.koss@leuphana.de)

# 1 Introduction

Since the end of the Cold War and the emergence of an international order focused on cooperation, cultural institutions have faced growing calls for transparency and accountability as to the origins of their collections. As a result, provenance research has benefited tremendously. And while the tradition of establishing ownership histories of artworks goes back centuries, born out of the wish to authenticate and establish the value of a work of art, the new context has brought with it new expectations. In the past, it may have sufficed to provide a simple list of names without proof or documentation. Today, the standards with which the quality of provenances is gauged are different. They are scientific.

The production of provenance is generally carried out on behalf of commercial actors, such as auction houses (to do their due diligence), artist estates (to establish catalogues raisonnés), and museums – the single most important type of institution engaged in producing, publishing, and maintaining provenances. With the responsibility of identifying objects in their collections that were unlawfully appropriated during the National Socialist period or in other contexts of injustice, museums are reconstructing and publishing the provenances of their collections at an unprecedented rate and level of detail. Indeed, to the level of complicated transactions and varying valuations.

This vast and ever-growing corpus of increasingly detailed information on the ownership histories of artworks presents an untapped source for applying data-driven approaches. While such quantitative and statistical analyses have only recently been taken seriously by art historians, they have long belonged to the methodological toolkit of economic and social historians.<sup>1</sup> Considering provenances from both an art historical perspective and a social and economic one, as our paper does, is thus a logical proposition. This article presents what such an interdisciplinary historical perspective on aggregate provenance data may look like and what technological steps are required for its quantitative and statistical analysis.

Section one of this article provides a conceptual sketch of research avenues in economic and social history that – intersected with perspectives from art history – aggregate provenance data can open up and explore. Sections two and

---

<sup>1</sup> As Diana Greenwald reminds us, some art historians looked at digital methods as early as the 1960s. See *D. Greenwald, Painting by Numbers. Data-Driven Histories Nineteen-Century Art*, Princeton 2021, pp. 1-2.

three describe how provenances, still essentially text-based sources of information, can be made machine-readable and enable historians to apply quantitative and statistical analysis methods. For this purpose, the article introduces deep learning models that address natural language processing tasks. The fourth and final section indicates how researchers can use structured provenance data for more complex analyses.

Why provenances? Provenances are ledgers containing the ownership and, to some extent, custody histories of artworks.<sup>2</sup> As Arjun Appadurai noted about the social lives of commodities:

It is only through the analysis of these trajectories that we can interpret the human transactions and calculations that enliven things. Thus, even though from a *theoretical* point of view human actors encode things with significance, from a *methodological* point of view it is the things-in-motion that illuminate their human and social context.<sup>3</sup>

The potential analysis of provenances on a large scale thus allows us to address issues related to art ownership that are *social*, from the construction and reproduction of value to the distribution and reproduction of wealth. Such an analysis is possible because provenances record discrete economic decisions concerning art ownership by individuals or institutions. According to Appadurai, decisions made by individuals regarding an object become meaningful only in the context of an object's overall history. That is, in relation to the other decisions recorded in its given provenance.

In their most common form, provenances are lists of names, locations, and dates, identifying the consecutive owners or custodians of artworks, where and when they obtained them, and how they were transferred. To give an example, consider the provenance of Pablo Picasso's painting *Head of Young Boy* (1944), as published on the website of the Art Institute of Chicago, which owns the painting today:

Sold by the artist to Paul Rosenberg & Co., New York, 1950 [invoice]; sold to Samuel Marx (1885–1964), Chicago, November 21, 1951 [invoice]; by descent to his wife, Florene May

---

<sup>2</sup> Provenances are usually defined as records of ownership changes. With auction houses prominently featured, for example, information on custody has always been present in provenances. Yet, such custody information remains partial and does not systematically include exhibition histories or other types of information on the geographical whereabouts of artworks.

<sup>3</sup> A. Appadurai, Introduction: Commodities and the Politics of Value, in: *Idem* (Ed.), *The Social Life of Things. Commodities in Cultural Perspective*, Cambridge 1986, p. 5.

Schoenborn (1903–1995), New York and Chicago, from 1964; bequeathed by Florene M. Schoenborn to the Art Institute, 1997.<sup>4</sup>

We can see from this example that, beyond a mere index of ownership, provenances are matrixes of textual, historical information. By documenting a multiplicity of human activities relating to a specific class of goods, provenances are representations of complex social and economic relations and contexts.

The first line of provenance, culminating in a semicolon, indicates a single, direct purchase by the dealer Paul Rosenberg from the artist Pablo Picasso at a specific time in their careers and within a particular market. Artist-dealer relationships can take many forms, changing over the course of their careers and shifting according to wider trends in the circulation of art. Galleries successfully promoting living artists can influence market value and supply by enabling the artists to increase their output. The importance of such relationships for the marketing and valuation of certain artists, styles, media, and genres has been underlined in monographic studies.<sup>5</sup>

The above provenance also provides a sense of the geography traversed, not merely identifying but also locating the individuals involved. A large enough dataset of provenances can thus deliver insights into both the local and trans-regional, if not trans-continental, circulation of art, helping us to map economic phenomena in relation to cultural ones.<sup>6</sup> In this case, the life of the painting, and with it its provenance, began in the artist's studio in Paris, the world's cultural capital in the first half of the 20<sup>th</sup> century. The ownership of the painting then passed to Paul Rosenberg's gallery in New York, which took over Paris in importance in the second half of the 20<sup>th</sup> century. Subsequently, the work belonged to a collector in Chicago, underscoring the role of cultural accumulation in the industrial heartland of the United States.

The final line of provenance concerns a bequest made to a museum in 1997 by the artwork's last private owner, a widowed woman. When analysed against

---

4 *Pablo Picasso*, *Head of Young Boy*, 1944, brush and black ink and grey wash, with scraping and touches of pen and black ink, on ivory wove paper, 500 × 288 mm, The Art Institute of Chicago, Chicago, Bequest of Florene May Schoenborn, 2012.569, <https://www.artic.edu/artworks/158479/head-of-young-boy>, 16.07.2022. The information contained in brackets in this provenance points to the existence of archival documents supporting the respective statement. As such, they are an idiosyncrasy of the museum's provenance style.

5 See for example, *M.C. FitzGerald*, *Making Modernism*. Picasso and the Creation for the Market of Twentieth Century Art, New York 1995.

6 For the potential of spatialisation of artistic phenomena based on numerical and statistical data and its analysis, see *B. Joyeux-Prunel/C. Dossin/S.A. Matei*, *Spatial (Digital) History*. A Total Art History? The Artl@s Project, in: *Visual Resources* 29, 2013, pp. 47–58.

a large provenance dataset, such an event is telling for several reasons. Firstly, it raises the question of gender and the circulation of art – an issue highlighted in our paper. Secondly, it throws into sharp relief the role of donations and gifts, a particular form of exchanging value that can be investigated in and of itself. Lastly, and perhaps most importantly, it highlights not only the role of museums in the circulation of art but also in the creation and maintenance of the reputation of artists – another aspect addressed in our article. As it turns out, donating a work by Picasso to the Art Institute of Chicago in 1997 is much less consequential than it would have been decades earlier. This is because the artist's stature as a giant of modern art and the Art Institute of Chicago's global reputation as a major museum of significant modern art holdings were already firmly entrenched by 1997.

From a single provenance, then, we can already intuit how individual pieces of information can open up and are tied in with geo-temporal social networks and their historical dynamics and cycles. Aggregate provenance data and its analysis thus allow us, for the first time, to not only account for them quantitatively but also make them visible.

## 2 Provenance as a Source for Economic and Social History

Based on the literature engaging with socio-economic questions related to art market mechanisms and behaviour, we can make out at least two areas of inquiry for which provenances provide unexplored source material: the construction of value and the dynamics of wealth (as embodied in cultural objects). When taking our cue from Appadurai and considering the commodity status of artworks, we are immediately confronted with a set of complications. In economic terms, i.e. in terms of market behaviour, artworks function according to a unique and nontransparent set of rules.<sup>7</sup> The art market and its construction of value are opaque. Despite an active research field tracing and studying the prices of artworks, explanations for how artworks obtain their economic value remain limited due to a lack of sufficiently detailed data on a large scale.

---

<sup>7</sup> See various chapters in *R. Towse/T. Navarrete Hernández (Eds.), Handbook of Cultural Economics*, Third edition, Cheltenham 2020.

Indeed, purchasing art, a luxury commodity, is a form of conspicuous consumption and can be understood as a proxy for wealth.<sup>8</sup> Yet, it remains difficult to measure due to the relative discretion of the actors involved. The historical interest in the dynamics of wealth indicated by the concentration, circulation, and distribution of artworks is also an aspect of art market research that has been little explored. This is especially true on an international, cross-border level, not least because, for a long time, various national art markets were not integrated in the way they are in today's globalised world.<sup>9</sup> And while this set of problems in dealing with art market mechanisms and behaviour has been widely acknowledged, a long-term, historical perspective that can describe both the processes of value formation and the dynamics of wealth remains missing.

Economists and economic historians interested in the value of artworks have focused predominantly on investigating prices and their development over time, often with the view to establishing art's profitability as an investment. For this, they have long relied on Gerald Reitlinger's work around *The Economics of Taste*, whose first volume was published in 1961.<sup>10</sup> With his impressive but selective data, which covers mainly auction transactions that span over 200 years of art market activity, Reitlinger's insights are anecdotal at best. A generation of economic historians has nevertheless built on his work due to the lack of alternative data compilations. To date, auction data has been the primary source for analysing the art market. This means that research has been limited to a fraction of the activities in which artworks change hands.

Despite the lack of large representative datasets, scholars use price indices to understand historical developments of the art market.<sup>11</sup> Studying repeat sales,

---

**8** B.R. Mandel, Art as an Investment and Conspicuous Consumption Good, in: The American Economic Review 99, 2009, pp. 1653-1663, here: p. 1653: "As a luxury good, relative art demand is an increasing function of wealth." More recently, Kim Oosterlinck has studied art as wartime investment and has proposed that certain types of art may in exceptional circumstances also be purchased for reasons to hide wealth, indicating that motives for purchasing art over time may vary. See K. Oosterlinck, Art as Wartime Investment: Conspicuous Consumption and Discretion, in: The Economic Journal 127, 2016, pp. 2665-2701.

**9** D.M. Challis, Foreign Currency Volatility and the Market for French Modern Art, Leiden 2021, pp. 6-7.

**10** G. Reitlinger, *The Economics of Taste*, Vol. I., The Rise and Fall of Picture Prices 1760-1960, London 1961; Vol. II, The Rise and Fall of Objects d'art Prices since 1950, London 1963; Vol. III, The Art Market in the 1960's, London 1970.

**11** For recent studies on prices, see e.g. A. Crotta/F. Vermeylen, Does Nudity Sell? An Econometric Analysis of the Value of Female Nudity in Modigliani Portraits, in: ACEI Working Paper Series (Association for Cultural Economics International) 2, 2020, pp. 1-32; K. Oosterlinck/A.-S. Radermecker, "The Master of..." Creating Names for Art History and the Art Market, in: Journal

for example, they track the price of one artwork at different times, locations, and with different people involved. Repeat sales indices acknowledge the fact that artworks are unique, non-interchangeable goods. The underlying assumption is that an artwork remains unchanged over time. The history of the art market has proven otherwise, however. The condition of artworks can change significantly, and even small changes can be reflected in the price, although we do not know exactly to what extent. A further limitation of such a method is that only artworks bought more than once can be considered. Expensive works that only change hands once in their lifetime, such as Rembrandt's *Nightwatch*, cannot be included due to their short provenance. This means that repeat sales indices are even less representative than other methodologies of analysing prices to understand value construction in the art market.

In contrast, the more commonly used hedonic regression can rely on more data points in any given price dataset. Here scholars have increasingly tried to identify the deciding factors in price formation. Characteristics used in hedonic regressions include object criteria (e.g. artist, medium, signature, and dimensions) and other market factors such as the seller, location, and time of the sale. Limitations of this method arise from the frequent use of dummy variables, which flatten historical complexity. Despite the sophistication of recent studies, hedonic regression also runs the risk of focusing on factors that are irrelevant or else ignoring aspects that could be important, such as the social and cultural aspects of value formation.

Social scientists have taken a different approach to the value formation of artworks. Rather than focusing solely on valuation outcomes, such as prices, these authors emphasise the importance of investigating the processes that lead to them.<sup>12</sup> With his in-depth study of galleries in the Amsterdam and New York contemporary art market in the late 20<sup>th</sup> century, the sociologist Olav Velthuis has shown, mainly through interviews, that value is a function of social context. He writes:

---

of Cultural Economics 43, 2019, pp. 57-95; Oosterlinck, Art as Wartime Investment; L. Renneboog/C. Spaenjers, Buying Beauty. On Prices and Returns in the Art Market, in: Management Science 59, 2013, pp. 36-53.

<sup>12</sup> E. Dekker, Two Approaches to Study the Value of Art and Culture, and the Emergence of a Third, in: Journal of Cultural Economics 39, 2015, p. 312.

The value of an artwork does not reside in the work itself, but is, under conditions of uncertainty, produced and constantly reproduced by artists, intermediaries, and audiences, subject to numerous conventions and codes of art worlds.<sup>13</sup>

A similar cast of actors in value construction was identified by Samuel Fraiberger et al. in their data-driven analysis of 497,796 exhibitions in 16,002 galleries, 289,777 exhibitions in 7,568 museums, and 127,208 auctions in 1,239 auction houses spanning 143 countries and 36 years (1980 to 2016).<sup>14</sup> One of the key findings of their study showed how artists who were exhibited early on in their careers in specific galleries and museums were more successful. These galleries and museums were usually located in Western art market centres and had the highest cultural and/or economic cachet. Success was measured here by higher auction prices later in life and a lower risk of dropping out of the art market altogether. The study focused on network analysis, investigating the interdependence between actors such as galleries and museums, their role in the exhibition circuit, and the effects on artists' careers and their auction prices. It demonstrated how the decision of specific galleries and museums on which artists to include in their exhibitions directly impacts their future success and the valuation of their work. These particular actors can therefore be considered gatekeepers since they influence the value of artists' work like few others do.

By not studying prices but looking at art market activity across time and space, Schich et al. have recently identified network dimensions in their important study in the field of cultural data analytics.<sup>15</sup> Using auction house data from four European markets (England, France, Belgium, and the Netherlands) between 1801 and 1820, they analysed 267,000 market transactions involving around 22,000 actors.<sup>16</sup> Despite their dataset's narrow temporal and geographical focus, they have revealed insights into social, temporal, spatial, and conceptual network dimensions. They have shown how auction activity was spread over four countries, with markets peaking at different times, measured by the number of transactions. Indeed, the European auction market cannot be fully

---

<sup>13</sup> O. Velthuis, *Talking Prices. Symbolic Meanings of Prices on the Market for Contemporary Art*, Princeton 2005, p. 160.

<sup>14</sup> S.P. Fraiberger et al., *Quantifying Reputation and Success in Art*, in: *Science* 362, 2018, pp. 825-829.

<sup>15</sup> M. Schich et al., *Network Dimensions in the Getty Provenance Index*, in: arXiv, 2017, <https://arxiv.org/abs/1706.02804v1> [physics.soc-ph], 31.10.2022.

<sup>16</sup> This data was taken from the Getty Provenance Index, the world's largest single repository of art market data. Their limited focus was due to the uneven quality of the data available, which forced them to work with a relatively complete, clean, and standardized subset.

understood by studying a single region since it is only via a comparison of four countries that it becomes evident that most protagonists stay within one region, where they stick to exclusively buying or selling. Equally, Schich et al. have described the behaviour of market participants, demonstrating the relative importance of specific dealers and the buying and selling activities of collectors. They observe: "Top sellers tend to appear on the market exactly once and top buyers tend to dominate the market for two to three years."<sup>17</sup>

Their study has also shown links between art market clusters and highlighted the existence of brokers between communities who both buy and sell. Looking at the location of sales, Schich et al.'s results have confirmed the known ranking dynamics of cultural centres in Europe; Paris is essentially synonymous with the French market (with only a few side locations, which never see auctions in consecutive years); London and Amsterdam are predominant in their countries; Belgium is a market with multicentric competition. Conceptually, their study has looked at the nationality of the artist attribution of objects recorded in auctions and related them to sales in the artist's respective countries. They found that top artist attributions function similarly to a "product category" in a supermarket. On this basis, they tried to identify sense-making time-frames of shopping cycles by looking at all transactions by a buyer in a given week, year, or in total.

This led them to conclude that a large portion of the European auction market was "a system that functions like a super-slow supermarket on an annual grocery-cycle." Schich et al. nevertheless admit that their conclusions are made "on thin ice" since it is unclear, for example, whether the protagonists involved believed the attribution of a given artist was true or not.<sup>18</sup>

What Schich et al. have in common with the many studies based on auction data analysis is that they rely on vague catalogue descriptions. Auction cataloguing, whether analogue or digital, is a marketing tool that may or may not camouflage uncertain attributions of artworks. This means the significance of these studies for analysing art market behaviour is circumscribed, in particular, by the absence of actual artworks. How can we draw conclusions from the analysis of discrete economic decisions of buying and selling unique goods when the commodities are reduced to a set of very few characteristics? Equally unreliable is the public reporting on the results of auction sales. The common practice of auction houses and consignors, especially in times of crisis, to unofficially guarantee prices, withdraw lots, and buy back art stock, makes publicly availa-

---

<sup>17</sup> Schich et al., *Network Dimensions*, p. 6.

<sup>18</sup> *Ibid.*, p. 8.

ble auction results a poor source for studying prices or understanding buying and selling behaviour in the art market.<sup>19</sup> Furthermore, buyers, such as museums and collectors, usually enlist the help of dealers and other agents to bid at auction while they remain anonymous.<sup>20</sup>

Here is where provenance data from museums or catalogues raisonnés differs fundamentally. The information contained in provenances has been studied on a micro-level, enhanced, and vetted by art historical expertise and knowledge.<sup>21</sup> What is more, provenance data includes not only auction sales, which represent a fraction of art market activities, but also a wide variety of information on how works of art change hands, especially in non-market circumstances, such as intra-familiar ownership changes, private sales, and donations to museums. In fact, from our experience of working with provenances, we have identified at least five relatively frequent types of activity related to objects owned by an individual. The first is the sale of individual works of art at auction or to a dealer.<sup>22</sup> The second is the exchange of artworks directly with a dealer or with another collector through the intermediary help of a dealer or gallery. The third is the liquidation of an entire collection by way of auction. The fourth is the passing on of artworks through inheritance, which may or may not be followed by the liquidation of the inheritance by the receiving party for various

---

**19** See, e.g. O. Ashenfelter/K. Graddy, Art Auctions, in: R. Towse/T. Navarrete Hernández (Eds.), *Handbook of Cultural Economics*, Third edition, Cheltenham 2020, pp. 19-21; for auctions during the Great Depression, e.g. the sales of the so-called Collection of a Swiss Nobleman, Albert Figdor, or Marczell von Nemes, see L. Rother, *Kunst durch Kredit. Die Berliner Museen und ihre Erwerbungen von der Dresdner Bank 1935*, Berlin 2017, pp. 26-51.

**20** See, e.g. L. Zalewski, Creating Cultural and Commercial Value in Late Nineteenth-Century New York Art Catalogues, in: J.D. Baetens/D. Lyna (Eds.), *Art Crossing Borders. The Internationalisation of the Art Market in the Age of Nation States, 1750-1914*, Leiden 2019, p. 101.

**21** We are cognisant of the fact that provenance records also have an inherent bias. They are only produced for works considered worthwhile spending resources for research and documentation. However, as constituents such as museums, authors of catalogues raisonnés, auction houses, collectors, and artists write provenances, the quality of the data is better and more representative for the works circulating in the art market than any other data containing such interactions and transactions. For computational analysis, object and provenance data would ideally be combined with exhibition data, which has been used more frequently and is easier to structure, and has also an impact on the construction of value and the concentration, circulation, and distribution of artworks. Art criticism data has so far not been systematically recorded digitally, but would equally enhance such analysis.

**22** Many transfers recorded in provenances are complex transactions, involving not only a single sender and a single receiver of a work, but possibly two or more senders or receivers, and also agents enlisted by senders or receivers.

reasons – not least to do with tax. The fifth activity, which frequently appears in provenances, is donations to museums and institutions.

Since provenances are chains of socio-economic events – involving actors such as museums, galleries, and auction houses and describing their role in a given transaction, the location and time of the event, and, increasingly, prices – they are destined for complex network analysis. Given that their data points come from multiple potentially related events, they allow us to exploit the advantages of the repeat sales method and hedonic regression analysis. Bringing such queries together with specific object information of unique artworks (such as artist, title, subject matter, genre, style, medium, size, and date) makes aggregate provenance data an inevitable source for analysing the construction of value. This is all the more so since both the economic and social sciences have acknowledged that object *and* social factors impact the construction of value in the art market.

With the potential of analysing interdependencies of an unprecedented wide spectrum of socio-economic activities in relation to the objects, aggregate provenance data will enable us to specify the role of institutions or people and to see patterns and outliers in the concentration, circulation, and distribution of artworks. Indeed, network analysis can account for a whole host of historical factors that have an impact on the frequency and outcome of socio-economic activities such as inheritance, sales, or donations: from wars and economic crises to changes in import and export or tax regulations. The impact of gatekeepers for specific artists, subject matters, genres, styles, or media can also be analysed from social, temporal, spatial, and conceptual perspectives. To show their importance, or indeed insignificance, would not only be possible within their period of activity and geographic reach but also in comparison to their peers and across historical periods. Identifying such complex network dimensions for the understanding of art markets and collecting practices requires machine-readable data of the histories of objects, however. Since provenance data is already digitally available on a large scale thanks to the work of museums and, increasingly, catalogues raisonnés, its structuring ushers in a new era for art market studies.

### 3 From Text to Data: Structuring Provenance

The various paths of analysing aggregate provenance data that we have laid out are based on the different types of information stored in provenance texts. To address a seemingly simple inquiry, however – such as a comparative understanding of how men and women may differ in their engagement with inherited

artworks – we first need to extract the historical knowledge contained in provenance texts. In the following sections, we illustrate how deep learning is the digital method most suited to this challenge, given the syntactic and semantic peculiarities of provenance texts and the variety of information contained in them. First, we discuss which natural language processing tasks are most appropriate for extracting information from provenance texts and why. Then we describe how we implemented deep learning models to address the two most promising tasks concerning provenances: Sentence Boundary Detection and Span Categorization.

Although the study of both provenance and the art market has an established historiographical tradition, the use of computational methods for analysing large-scale phenomena in this field is still in its early and experimental stages. This is because provenances have thus far been recorded and published as texts understandable only by humans and not by machines. Many provenance repositories, like museums, adhere to guidelines on recording provenance, such as those published by the American Alliance of Museums (AAM) or guidelines loosely based on them.<sup>23</sup> According to the AAM guidelines, provenance texts are a succession of sentences, where each sentence stands for a provenance event. These provenance events are listed in chronological order, with the first event ideally identifying the creation of the object or the moment of its archaeological discovery. According to the guidelines, an event contains information about the participants involved, the method of transfer, the date, and the location. Events are divided by punctuation marks, which carry significant meanings. The guidelines determine that if two events are divided by a semicolon, there was a direct transfer between the two owners. However, if two events are separated by a period, there was no direct transfer between the two owners, and a gap in the historical reconstruction of events can be inferred. Parentheses can provide additional knowledge, such as biographical information. Further information about events, such as historical sources, can be provided in footnotes. It should be noted that, aside from this guidance on the form and structure of provenance events, the AAM guidelines leave much room for interpretation when museums come to write their provenances. This means that there may be differences between the provenance texts of different museums that conform to the AAM guidelines – an issue we will address in the next section.

To analyse provenance events on a large scale, we need to extract information from provenance texts with the help of digital methods. Automatic in-

---

<sup>23</sup> N.H. Yeide/A.L. Walsh/K. Akinsha, *The AAM Guide to Provenance Research*, Washington, DC 2001.

formation extraction is one of the purposes of Natural Language Processing (NLP). NLP is an area of research at the intersection of computer science, linguistics, and, more recently, artificial intelligence. It investigates tasks that can be performed by digital methods to automatically process and analyse natural human language on a large scale. Specifically, as in the case of provenance, the NLP problem of extracting event information from a written text is defined as event extraction, for which several NLP tasks are available – something we come to discuss.<sup>24</sup> An event can be understood as an entity represented in space and time in which the parties involved take actions that result in a change.<sup>25</sup> Where provenances are concerned, the change caused by an event affects the ownership of a work of art.

We must take four event elements into account to perform event extraction.<sup>26</sup> The first element is the *event mention*, the text portion containing a potential event, which usually forms a sentence. When extracting events from provenance, we already have the advantage that punctuation marks delimit each event according to the AAM guidelines. Therefore, each sentence is considered an event. The second element is the *event trigger*, that is, the element explicitly indicating that an event is taking place. The most common *event trigger* is a verb. This is also true for provenances since they contain verbs that indicate various methods of transfer, such as “sold”, “bequeathed”, “exchanged”, or “purchased”. We may also encounter nominal sentences using expressions such as “by descent” or “in exchange”. A peculiarity of provenance texts is that we may encounter no *event trigger* at all for a particular sentence. The text gives nothing but the object’s owner in such a case. Nevertheless, we can still consider it a provenance event since we can extract information about the owner who received the object. The third element of event extraction is the *event argument*, which is to say, the different components of an event. For example, a provenance event, in addition to the *event trigger*, may also have a date, location, and involved parties as *ar-*

---

<sup>24</sup> For event extraction from historical texts, see: A.K. Cybulska/P. Vossen, Historical Event Extraction from Text, in: Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities, Portland 2011, pp. 39-43; R. Sprugnoli/S. Tonelli, Novel Event Detection and Classification for Historical Texts, in: Computational Linguistics 45, 2019, pp. 229-265; V.D. Lai et al., Event Extraction from Historical Texts: A New Dataset for Black Rebellions, in: Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, 2021, pp. 2390-2400.

<sup>25</sup> W. Xiang/B. Wang, A Survey of Event Extraction From Text, in: IEEE Access 7, 2019, pp. 173111-173137.

<sup>26</sup> Ibid.; Linguistic Data Consortium, Ace (Automatic Content Extraction) English Annotation Guidelines for Events, Philadelphia 2005.

*guments*. It is also quite common to find other information in a provenance event referring to biographical subevents of the parties involved, such as their birth, death, or movement to another location.

The *event argument role* is the last element to consider in event extraction. Where participants are concerned, we have to determine who gives the item (sender) and who receives the item (receiver). But then again, a participant can also take the role of the intermediary (agent). It is, moreover, crucial to distinguish whether a place or date in a provenance is an *argument* referring to the main event or a biographical sub-event, such as birth.

In light of these four elements, we have divided the event extraction problem into two sub-problems. The first of these problems concerns splitting the provenance text into discrete provenance events. Such an approach allows us to preliminarily isolate the different *event mentions*. From these mentions, we can, in turn, extract the *event trigger*, *event arguments*, and *event argument roles*. To split a provenance text into discrete events, we can apply a Sentence Boundary Detection (or Disambiguation, SBD) task. This task involves recognising which characters start and end a sentence by disambiguating punctuation. We are reminded that, according to the AAM guidelines, provenance events are divided by semicolons or, if followed by an unknown event, by periods. More established digital methods to address SBD tasks include rule-based techniques, such as decision trees or regular expressions.<sup>27</sup> Although more straightforward to implement, the disadvantage of these techniques lies in their failure to disambiguate punctuation marks in some contexts. Consider, for example, the punctuation marks used for abbreviations, which are easily confused with periods and can coincide, in rare instances, with the end of sentences. More sophisticated methods, such as deep learning, allow us to address these ambiguities.

Deep learning models have recently addressed NLP tasks with impressive results.<sup>28</sup> Such digital methods, inspired by the structure of the human brain, use neural networks to extract features from digital sources, such as images or, in our case, texts. Neural networks are hierarchically organised and interconnected layers of mathematical functions, i.e. artificial neurons. From digital sources, they can identify and extract different characteristic features through a process

---

<sup>27</sup> M.D. Riley, Some Applications of Tree-based Modelling to Speech and Language, in: Proceedings of the Workshop on Speech and Natural Language, 1989, pp. 339-352; J. Aberdeen et al., MITRE. Description of the Alembic System Used for MUC-6, in: Proceedings of the 6th conference on Message Understanding, 1995, pp. 141-155.

<sup>28</sup> L. Alzubaidi et al., Review of Deep Learning. Concepts, CNN Architectures, Challenges, Applications, Future directions, in: Journal of Big Data 8, 2021, article 53.

of abstraction, from one neural network layer to the next.<sup>29</sup> This provides neural network models more flexibility than an algorithm-oriented approach, which is centred on a specific goal with certain predetermined parameters.

Extracting features means finding the correct representation of raw data.<sup>30</sup> For example, in the case of text, the representation is generated by extracting semantic and syntactic features from tokens, i.e. smaller textual units such as words. To provide a representation of a word, it is necessary to analyse the different contexts in which a word appears and infer patterns. As Firth summarises: “You shall know a word by the company it keeps.”<sup>31</sup> A neural network can infer patterns from a context through a process of multiplied, layered abstraction. However, such a network needs sufficient examples to effectively generalise a token's features across different contexts. Therefore, we need to train the model on a set of examples before we can use it on a larger scale. In addition, the examples provided for training should be annotated according to the desired output so as to calibrate the model to perform specific tasks. For instance, in the case of SBD, we need to train the model with texts in which sentence boundaries are annotated.

The effectiveness of deep learning in performing SBD has stimulated both the creation of cross-domain models (trained on large corpora of generic and consistent texts such as newspaper articles) and domain-specific models (where the text has a peculiar structure, as in the clinical and legal domains).<sup>32</sup> Performing an SBD task on provenance texts compiled according to AAM guidelines requires us to train a domain-specific model from scratch. This is because provenance texts do not always follow a logical and linguistic structure comparable to standard texts. As previously mentioned, sometimes they do not even have a verb present. Finally, according to a provenance-specific SBD model, sentences can be divided not only by periods but also by semicolons.

The second sub-problem of event extraction from provenances concerns identifying the *event trigger*, *event arguments*, and *event argument roles* for each provenance event. We can solve these three issues with three specific tasks. For

---

29 W.J. Zhang *et al.*, On Definition of Deep Learning, in: 2018 World Automation Congress (WAC), 2018, pp. 1-5.

30 Y. LeCun/Y. Bengio/G. Hinton, Deep Learning, in: Nature 521, 2015, pp. 436-444.

31 J.R. Firth, A Synopsis of Linguistic Theory, 1930-1955, in: J.R. Firth (Ed.), Studies in Linguistic Analysis, Oxford 1957, pp. 1-32.

32 D. Griffiths *et al.*, A Quantitative and Qualitative Evaluation of Sentence Boundary Detection for the Clinical Domain, in: AMIA Summits on Translational Science Proceedings 2016, pp. 88-97; G. Sanchez, Sentence Boundary Detection in Legal Text, in: N. Aletras *et al.*, Proceedings of the Natural Legal Language Processing Workshop 2019, Stroudsburg 2019, pp. 31-38.

example, we can identify the *event trigger* by performing a Part-of-Speech tagging (PoS tagging) task. This task deals with recognising each word's grammatical role in the text. For example, PoS tagging involves identifying which words are verbs, the primary *event triggers*. *Event arguments* are named entities, such as people, organisations, places, and temporal expressions. We can therefore use a Named Entity Recognition (NER) task to recognise and categorise such entities. Finally, we can perform a Relation Extraction (RE) task to identify event argument roles.<sup>33</sup> As the name suggests, relation extraction involves identifying semantic relationships within the text. In this way, we can represent the *event argument roles* in relation to the *arguments* and the *event trigger*.

PoS tagging, NER, and RE are NLP tasks that deep learning models can successfully perform. However, as with SBD, we also need to train domain-specific models for these tasks, annotating training examples for each. In addition to the SBD model, we would need to train three new models from scratch to extract provenance events. The effort involved in such an approach, however, made us consider alternative options and seek out a more efficient process that could potentially address all three tasks at the same time.

We thus experimented with identifying the *event triggers*, *event arguments*, and *event argument roles* using a Span Categorization (or Classification) task.<sup>34</sup> This task is similar to that of NER. Unlike NER, however, Span Categorization does not focus on individual tokens but spans, i.e. portions of text. Span Categorization involves recognising and classifying spans, whether they are named or unnamed entities. In addition, unlike NER, multiple spans can be recognised from the same text portion. This characteristic allows us to create a hierarchy of spans, possibly assigning more than one category to the same text portion. In this way, we can assign an *event argument role* as an additional category to a span recognised as an *event argument*.

---

33 G. Zhou et al., Exploring Various Knowledge in Relation Extraction, in: Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05), 2005, pp. 427-434.

34 M. Xu/H. Jiang/S. Watcharawittayakul, A Local Detection Approach for Named Entity Recognition and Mention Detection, in: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2017, pp. 1237-1247; M.G. Sohrab/M. Miwa, Deep Exhaustive Model for Nested Named Entity Recognition, in: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp. 2843-2849; C. Tan et al., Boundary Enhanced Neural Span Classification for Nested Named Entity Recognition, in: Proceedings of the AAAI Conference on Artificial Intelligence 34, 2020, pp. 9016-9023.

Figure 1 shows how, conceptually, Span Categorization can be applied to a provenance event.<sup>35</sup> First, we identified the *event trigger*. In this example, the *event trigger* was the term “by descent,” classified as *method* (method of transfer). Then we identified the different *event arguments*. While we tagged the span “his wife, Florene May Schoenborn (1903-1995), New York and Chicago” as *party*, “1964” indicates the *time* of the event. The advantage of Span Categorization is the ability to assign additional classifications to a party’s span. Firstly, we identified the role: the party is the one who received the object (*receiver*). In addition, we assigned the type of party, whether it be a *person* or a *group*. Finally, given our research interest in the role of women in the history of collecting, we also classified the participant as a *female party*. Through this last categorisation, we performed another NLP task by addressing the issue of gender classification.<sup>36</sup>

by descent to his wife, Florene May Schoenborn (1903–1995), New York and Chicago, from 1964;						
Method	Name	Time	Time	Location	Location	Time
		Birth	Death			
Description						
Party						
Receiver						
Person						
Female Party						

**Fig. 1:** Conceptual Example of Span Categorization Applied to a Provenance Event. Source: based on the provenance of Pablo Picasso’s *Head of Young Boy*, as published on the website of the Art Institute of Chicago (<https://www.artic.edu/artworks/158479/head-of-young-boy>, 16.07.2022); created by Fabio Mariani.

The ability to overlap multiple spans also allows us to assign categories to smaller spans within a portion of text that we have already classified. Returning to the example above, we can classify “Florene May Schoenborn” as the party’s *name*. We can also distinguish a *description* of the party, i.e. the span “his wife”.

35 The event is excerpted from the aforementioned provenance of *Pablo Picasso*, *Head of Young Boy*, as published on the website of The Art Institute of Chicago (<https://www.artic.edu/artworks/158479/head-of-young-boy>, 16.07.2022).

36 Y. Hu et al., What’s in a Name? – Gender Classification of Names with Character based Machine Learning Models, in: *Data Mining and Knowledge Discovery* 35, 2021, pp. 1537-1563. For the feasibility of the NLP task and given the inconsistent nature of the underlying historical documentation, we have opted to apply a binary gender classification. We acknowledge that gender identity is more diverse.

After extracting the data, this allows us to trace the relationship of the party to the previous owner, from whom they inherited the object. Moreover, the text provides the party's biographical dates of "1903" and "1995", which are identifiable as two spans belonging to the *time* category, the first of which can be tagged as *birth* and the second as *death*. Finally, the text portions "New York" and "Chicago" refer to the party's locations, which are both classifiable as *location*.

For all the advantages that Span Categorization offers, however, there are, of course, disadvantages. Compared to NER, Span Categorization is more computationally expensive since it does not operate on individual words but rather portions of text. When assigning categories, it, therefore, has more candidates to consider, which may or may not be overlapping. In a NER scenario, the potential candidates are equal to the number of words in the text. In a Span Categorization case, the number of candidates is

$$\frac{n*(n+1)}{2},$$

where  $n$  is the number of words in the text.<sup>37</sup> Furthermore, increasing complexity makes the task of establishing an entity's boundaries more difficult. For example, the span "Paul Rosenberg & Co." refers to an organisation and not a person. And yet this situation may prove ambiguous for a Span Categorization model, which might only recognise "Paul Rosenberg" as a person. Finally, even in the case of Span Categorization, we must still train a domain-specific deep learning model from scratch.

Despite these issues, Span Categorization, combined with SBD, nevertheless allows us to deal successfully with event extraction by having to train only two models. In the next section of this article, we illustrate how we experimented with event extraction in provenance texts by training SBD and Span Categorization models on a museum dataset.

## 4 Training SBD and Span Categorization

The Art Institute of Chicago was founded in 1879 and is a prominent museum with strong collections across multiple departments ranging from the Ancient Americas to the Arts of Africa and Asia. It is particularly known for its exceptional holdings of European Modern art, extending to Contemporary global art.

---

<sup>37</sup> Tan et al., Neural Span Classification.

It is, however, one of the few museums that have made its collection dataset available to the public, including provenance texts – be it online via the museum website or as a download.<sup>38</sup> That is why we chose the Art Institute for our experiment, hoping that the results will only inspire more museums to make their collection data available to the public. The version of the dataset we used in the experiment was downloaded on April 7<sup>th</sup>, 2022, and contained data for 122,317 objects, varying in medium, culture, and period. For the experiment, we focused on those objects with provenance texts, which counted 11,504 objects (9.4 percent of the dataset). Generally, the Art Institute's provenances follow the AAM guidelines, with the peculiarity that notes for each provenance event are given in square brackets within the text, as opposed to in the footnotes.

Before training the SBD and Span Categorization models, we pre-processed the data. This step is crucial in standardising texts, helping to clean up any human errors or stylistic peculiarities. Firstly, the spaces between words were standardised. We replaced inconsistent spacing caused by tabulation and removed multiple spaces, including those found before periods, semicolons, or slashes. Parentheses and quotation marks were also made uniform. Furthermore, where notes were found in curly brackets, they were replaced with square brackets. As for quotation marks, we replaced any curly quotation mark with a straight one. Different dashes used for hyphenation were also standardised as the hyphen-minus character. And finally, all HTML tags were removed. After this cleaning process, we discovered that 112 problematic provenance texts (0.97 percent of the texts) contained typos and errors, such as open/unclosed parentheses or not conforming to AAM guidelines. Since correcting such typos and errors requires human intervention, we decided to discard them altogether. The dataset we experimented on thus consisted of 11,392 objects with provenance texts, as opposed to the 11,504 we started with.

Once the provenance texts had been pre-processed, we were able to proceed with model training. For both models, we used the open-source library *Spacy*, written in Python.<sup>39</sup> *Spacy* offers a set of generic models trained to perform various NLP tasks, including those already discussed: SBD, PoS tagging, and NER.

---

<sup>38</sup> The Art Institute's Github page (<https://github.com/art-institute-of-chicago/api-data>, 08.07.2022) provides examples of the data as well as the external link to download the entire dataset. Data from a single object is also available via API (<https://www.artic.edu/open-access/public-api>, 08.07.2022), however for using the whole dataset it is recommended to download the dataset.

<sup>39</sup> *ExplosionAI*, *Spacy* 3.2.4 (<https://spacy.io/>, 18.07.2022). We conducted the experiment on a MacBook Pro (2019, 2.6 GHz 6-Core Intel Core i7), using Python 3.8.

In addition, the library offers users the opportunity to configure new neural network models and train them for custom tasks.

From the dataset, we randomly selected 6,000 objects whose provenance texts we then annotated to train a domain-specific SBD. For each text, we tagged the boundaries delimiting the provenance events.<sup>40</sup> We then divided the 6,000 annotated texts into three groups: the training set, the validation set, and the test set. We used 3,600 texts (60 percent of the corpus) in the training set. This set contained the examples with which we trained the model to predict the boundaries of a provenance event. Since training is an iterative process, finding the right number of iterations is essential, i.e. the times the model is updated. Too few iterations can cause what is known as underfitting (poor learning), while too many iterations can cause overfitting (poor generalisation ability).<sup>41</sup> For this reason, we used 1,200 texts (20 percent) for the validation set to control the model's performance during training. This so-called early stopping technique adjusts the number of iterations and stops the learning process before overfitting occurs.<sup>42</sup> Which is to say, before the model no longer improves in testing against the validation set after a predetermined number of iterations (in our case, 1,600 iterations). Our SBD training continued for 3,000 iterations before early stopping interrupted the process to prevent overfitting. Finally, we used the remaining 1,200 texts (20 percent) for the test set, which helped us to evaluate the quality of the model's predictions after training.

The model reached peak performance after 1,400 iterations, with an F1 Score of 0.99 recorded on the validation set. The F1 Score is a measure by which a model's training is evaluated. It is given by the following formula:

$$\frac{2 * Precision * Recall}{(Precision + Recall)}$$

Precision is a score that measures the quality of the model's predictions, that is, the ratio of true positives (tp) to false positives (fp), which is calculated as:

$$\frac{tp}{(tp + fp)}$$

---

**40** We annotated the texts using Doccano (<https://doccano.github.io/doccano/>, 20.09.2022). Doccano is an open source data labelling tool.

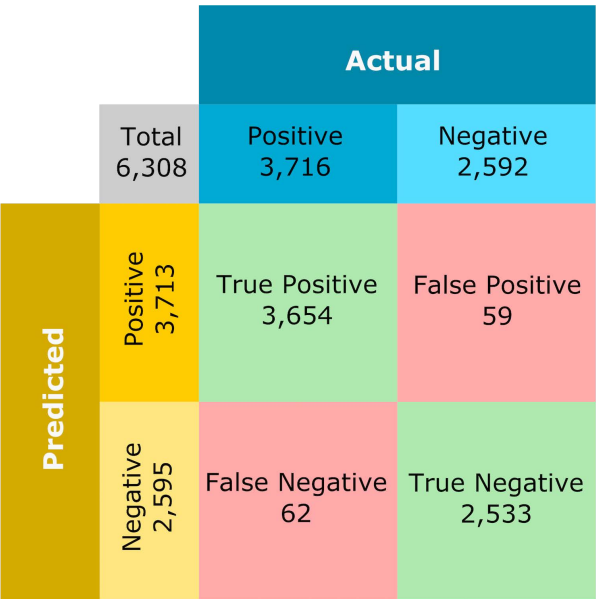
**41** X. Ying, An Overview of Overfitting and its Solutions, in: Journal of Physics: Conference Series 1168, 022022, 2019.

**42** L. Prechelt, Early Stopping – But When?, in: G.B. Orr/K.-R. Müller (Eds.), Neural Networks. Tricks of the Trade, Lecture Notes in Computer Science, Berlin 1998, pp. 55-69.

Recall rewards the number of predictions and is inversely proportional to the number of missed predictions, i.e. false negatives (fn):

$$\frac{tp}{(tp + fn)}$$

The F1 Score is the harmonic mean of these two fundamental measures. The maximum value for F1 Score, Precision, and Recall is 1.



**Fig. 2:** Confusion Matrix Showing the Model's Results of the Test Set. Source: 1,200 provenance texts from the Art Institute of Chicago, selected for testing an SBD model; created by Fabio Mariani.

The performance of the test set gave the quality assessment of the model, as summarised in figure 2. The model predicted the boundaries of 3,713 provenance events (predicted positives), 3,654 of which were correct (true positives), which means that 59 times the prediction turned out to be incorrect. We can speak of false positives in this case since the model incorrectly divided the text into provenance events. Indeed, the model failed to recognise 62 provenance events, otherwise known as false negatives. Finally, the model disambiguated 2,533 characters, which, although they take the form of a period or semicolon, do not delimit a provenance event (true negatives). Calculating Precision, Recall, and the F1 Score, the model registered 0.98 in all three measures on the test set.

Despite such an excellent result, we noticed while analysing the errors that provenance texts with notes inside square brackets (as practised by the Art Institute) created difficulties for the model. Sometimes these notes were several sentences long and created ambiguous scenarios for the SBD model. We, therefore, strengthened the model by adding an auxiliary rule which stipulated that any event boundary recognised inside brackets should not be considered. We then implemented an algorithm that corrected these predictions during the workflow of the model. Such an approach creates a hybrid model that benefits from the generalisation of deep learning and the definition of strict rules. The results of the test set for the hybrid model confirmed the efficacy of the auxiliary rule. Total predictions dropped to 3,695, compared to the previous count of 3,713. Of these predictions, those that were correct rose to 3,672, whereas they previously counted 3,654. Furthermore, false positives dropped to 23, less than half of the 59 cases mentioned above. This, in turn, led to the identification of more correct provenance events, reducing false negatives to 44. In light of these refinements, the Precision, Recall, and F1 Score values rose to 0.99. That is all to say that while the model had already achieved a high quality of efficacy through deep learning, we increased the result ever so slightly through hybridisation. This intervention, moreover, ensures that even better results are reached when applying SBD to the whole dataset.

Having completed the training on the SBD model, we turned our attention to the Span Categorization model. Again, the first step was to select and annotate data for training, validation, and testing. Since Span Categorization applies not to the entire provenance text but to each provenance event, we used the new SBD model to divide all provenance texts into events. From the 11,392 provenance texts in the Art Institute's dataset, we extracted 35,554 provenance events.

As discussed in the previous section, one of the problems of Span Categorization is the question of complexity, which is proportional to the number of words in any given text. As we have already seen with SBD training, the peculiarity of the Art Institute's provenance events, with their long notes in square brackets, creates ambiguity. For this reason, we pre-processed the texts of provenance events by extracting the notes in square brackets, transforming them into footnotes, and then replacing the missing text with their respective number in square brackets. We performed this intervention on all notes that were either longer than 20 characters or contained numbers (which usually refer to bibliographic citations).

Since Span Categorization is more complex and text-dependent than SBD, we tried to select training, test, and validation data that were as representative as possible of the entire dataset. For this reason, we clustered the provenance

events according to text similarity by calculating cosine similarity among the different texts.<sup>43</sup> Cosine similarity is a measure to calculate the similarity of two vectors and can range from 0 (totally different) to 1 (identical). In our case, we compared two provenance events as two vectors having each  $n$  dimensions, where  $n$  is the sum of the total words of the two texts. For each dimension, we assigned a value of 1 to it if the respective word was present in the text represented by the vector or 0 if not. Once both provenance texts were represented as vectors, we calculated their cosine similarity. This calculation allowed us to collect event texts with a cosine similarity of 0.5 or greater into 6,531 clusters. For each cluster, we randomly selected one event, creating an annotation corpus of 6,531 items.

We tagged 17 span categories in the provenance events of the annotation corpus. First, we annotated the *method*, *time*, *location*, and *party*. As discussed in the previous section, in Span Categorization, it is possible to have multiple categories for a span or its sub-portions. For example, each span annotated as *party* was also categorised by type (*group* or *person*), role (*sender*, *receiver*, or *agent*), and gender (*female party*). Within each party, we also tagged the span representing biographical information, such as *name* (even more than one, e.g. maiden name), *birth* and *death*, with the associated *time* and *location* span. In addition, we annotated any life *location* within the *party* span. Finally, through the *description* tag, we annotated portions of text providing additional information about the *party*, such as their profession or relationship with other parties. Where relationships were concerned, it was also possible to annotate any parties mentioned within the description. We maintained a similar approach for groups of people acting together (e.g. couples), tagging them as a group-type party containing multiple person-type parties. For some events, especially auctions, where the inventory or lot number is often given, we annotated with the *inventory* tag. Given the possibility of overlapping annotations, we used a *vagueness* tag to categorise approximate *time* or *location* spans (e.g. “circa 1945” or “near Paris”) and an *incompleteness* tag for *party* spans whose information was incomplete (e.g. “unknown collector”). During the annotation stage, 80 provenance events were discarded as SBD errors (1.2 percent of the annotation corpus).

Training of the Span Categorization model was carried out along similar lines to that of the SBD model, dividing the annotation corpus into three datasets: the training set (3,871 texts making up 60 percent), the validation set, and the test set (each containing 1,290 texts amounting to 20 percent). Training

---

<sup>43</sup> W.H. Gomaa/A.A. Fahmy, A Survey of Text Similarity Approaches, in: International Journal of Computer Applications 68, 2013, pp. 13-18.

continued for 5,800 iterations, reaching peak performance on the validation set after the 4,200th iteration, with an F1 Score of 0.95. By evaluating the model on the test set, we obtained Precision and Recall results of 0.95 and 0.93, respectively, and an F1 Score of 0.94. While these results amount to overall good performance, they are lower than the SBD model's, which performed a more straightforward task. For instance, we can see that the Span Categorization model has a lower Recall result than it does Precision, which means that it fails to identify categories for some spans. Still, when it does, it is fairly accurate. This shows that the model did not encounter any difficulties when handling entity boundaries, however, which we previously identified as one of the drawbacks in dealing with the Span Categorization task. If it had encountered such difficulties, we would have found ourselves with a model with a Precision result lower than its Recall result, which would have meant that it was more adept at identifying spans but more prone to errors.

Looking back at the excellent results of the experiment with the Art Institute's data, we are confident that the method can be applied on a larger scale, training models using provenance texts from several institutions. By using deep learning to address the two NLP tasks of SBD and Span Categorization, we are able to identify discrete provenance events and the various elements presented in them with reliable accuracy. Moreover, Span Categorization offers the possibility to go beyond event extraction, as it allows us to extract layered and complex information on individual provenance elements, especially those involving parties. With the help of Span Categorization, researchers analysing data can query more complex phenomena, which depend on individuals occupying multiple roles at once. For example, they will be able to identify a person of a certain age, gender, and relation to another person or group, not to mention acting in a certain place, time, and role.

## 5 Preliminary Analysis

On the basis of the publicly available collection data from the Art Institute of Chicago that we structured through our experiment, we were able to undertake preliminary data analysis. This analysis allowed us to identify not only certain dimensions of the historic acquisition patterns of the museum but also gender differences regarding the question of how individuals that inherit artworks engage with them subsequently.

The collections of museums are not only their most important source for attracting and educating visitors, but they are also a signifier of their wealth,

which is ultimately the cumulative result of a string of economic decisions regarding collecting strategies that can be traced from their founding years to the present. Their standing and recognition rely on the capacity to acquire works of art, either by spending money (or offering other works) or by attracting donors. Across aggregate provenance data, it is possible to discern whether or not the decision of a museum to acquire a certain artist – through purchase, gift, donation, or some other method of transfer – has had a measurable effect on their performance with collectors and/or other museums.<sup>44</sup>

Any acquisition occurs in a market setting where museums, dealers, and collectors compete for artworks. When a museum acquires an artwork, this can have multiple effects. We already know from Fraiberger et al. that the inclusion of an artist in an exhibition by a (specific) museum positively affects their reputation and success. When museums acquire specific works, they exert even more influence on the recognition of objects and, ultimately, how the canon is formed. When an object enters a museum collection permanently, the object is automatically elevated to the level of so-called museum quality, which at once marks it as cultural heritage – worth collecting, storing, and preserving.

When a museum department buys an object, there is always a decision to purchase one object over other objects. This is because museum departments are usually restricted by financial constraints or infrastructural factors, such as limited storage. While a lack of storage also concerns donations and bequests, their acceptance does not affect museums' generally scarce acquisition budgets. Moreover, with their expert judgements, we can assume that curators are more invested in and have a bigger influence on the selection process of what museums buy and when than they do on donations and bequests. It is, therefore, logical to assume that of all the acquisition methods, purchases by museums have the highest symbolic meaning. This can, in turn, affect the growing reputation and value of an artist, style, or a related group of objects when observed on a micro-level by market participants. Although we would like to acknowledge the many nuances of acquisition processes, such as the influence of curators on future donors, for our preliminary analysis on a macro-level, we have focussed on the simplified categorisations of *active* and *passive* purchases.<sup>45</sup>

---

<sup>44</sup> An example of an analysis of the impact of museums on the art market can be found in W.P. Pommerehne/L.P. Feld, *The Impact of Museum Purchase on the Auction Prices for Paintings*, in: *Journal of Cultural Economics* 21, 1997, pp. 249-271.

<sup>45</sup> For a discussion of acquisition categorisation, see I. MacDonald, *Counting When, Who and How. Visualizing the British Museum's History of Acquisition Through Collection Data, 1753-2019*, in: *Journal of the History of Collections*, 2022, fhac034, p. 7.

So how has the Art Institute of Chicago collected objects since its founding? From their collection of circa 300,000 objects, 122,317 object records are publicly available through their data dump. We could draw from 10,776 of these objects and their accompanying provenance texts for our preliminary analysis.<sup>46</sup> Our main objective was to differentiate between objects that were purchased actively by the Art Institute (as indicated by the terms *sold* and *purchased*) or acquired passively through donation (indicated by the terms *given*, *gifted* or *gift*) or bequest (indicated by *bequeathed*).<sup>47</sup>

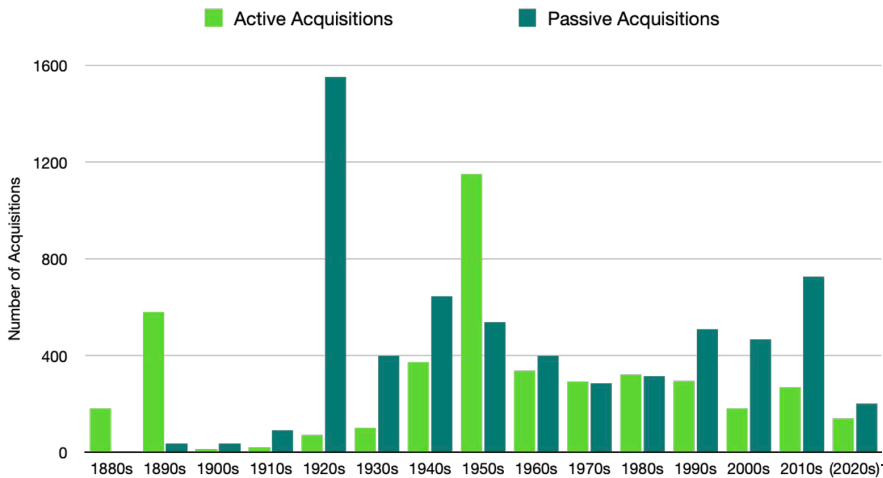
Such a birdseye view of a collection's history of acquisition patterns provides insights into the institution's history (fig. 3). Where the Art Institute is concerned, the institution began with an active purchasing period, which was followed by an increasing reliance on donations in the first decades of the 20<sup>th</sup> century. From the 1930s to the 1950s, the institution then increased its share of purchases compared to gifts, followed by a period reaching into the 1980s of relative stability in the ratio of purchased to donated objects. Since the 1980s, the institution's activities have been marked by an increase in passive acquisitions rather than purchases.

What emerges from the data beyond this broad view are, of course, outlier events that punctuate the institution's history of acquisitions. Of particular interest are spikes in purchasing that are easily identifiable. The first peak of purchases occurred in the 1890s. Then, after steady growth, a second peak occurred in the 1950s. This latter peak also marks the decade in which the highest number of objects were directly purchased by the museum. A closer look at the data reveals that these two peaks can be attributed to specific departments, namely, the Arts of Africa in the 1890s and the Arts of the Americas in the 1950s. Both purchasing peaks were related to singular and exceptional circumstances in which collections of objects became available on the market or were offered, and the museum decided to act. It should be noted that the numbers here reflect the number of objects rather than their prices. A graph representing the monetary value of purchases and donations may look much different due to single objects, such as modern paintings, being worth multiple times that of objects from other collecting areas.

---

<sup>46</sup> A total of 616 provenance texts could not be included in this analysis as they did not contain an explicit record of the Art Institute as a receiver of the object. The analysis therefore considered 10,776 provenance texts from the total dataset of 122,317 object records.

<sup>47</sup> Of the 10,776 acquisitions made by the Art Institute of Chicago, 4,329 are a purchase, 5,686 relate to a donation, and 523 are a bequest. The remaining 238 acquisitions refer to a variety of less frequent methods of transfer, excluded from this analysis.



**Fig. 3:** Comparison of Active and Passive Acquisitions of the Art Institute of Chicago by Decades from the 1880s to the 2020s. Source: 10,538 acquisition events extracted from provenance texts from the Art Institute of Chicago dataset downloaded on April 7th, 2022; created by Fabio Mariani.

From the perspective of researchers, such preliminary insights into acquisition patterns already offer a range of questions to be investigated. For example, researchers could explore the wider significance of the 1955 purchase by the Art of the Americas department of 571 objects from an illustrious collection of Pre-Columbian objects. Did the event change the museum's collecting strategy? Was it an event unique to the Art Institute, or did similar events with similar provenance details occur at other museums? Did the purchase impact the wider collecting of Pre-Columbian objects in the US? Did the purchase (or the potential simultaneity of similar purchases) have lasting effects on the valuation of Pre-Columbian artefacts?

Our second preliminary analysis relates to the market behaviour of specific people. From experience, we know that provenances usually record the locations of owners rather than the actual locations of their objects. Aggregate provenance data thus provides a general indication of the capacity for conspicuous consumption at a particular location. Conspicuous consumption in the form of collecting can also function as an indicator of elite behaviour and can support a social and economic analysis of the differentiated developments of the elite and, indeed, their wealth.

On an individual level, provenances contain varying degrees of information about the fates of objects at the hands of their owners or at the hands of those who receive works of art due to death or divorce, for example. With aggregate

provenance data, we can also contextualise the sale of an artwork and better understand whether this activity is part of a collector's larger, long-term collecting strategy or not. The ability to observe such behaviours allows the art historian to describe, for example, the changing shape of any given collection, which, when analysed together with the collecting activity of other collectors, can amount to a history of taste. For the economic and social historian, such a sequence can indicate the consolidation of wealth or the liquidation of one type of asset for another, i.e. art for money. From aggregate provenance data, we can thus gain insights into not only the relationship between collecting and larger macroeconomic trends but also the historically shifting role of art as an asset class within those trends.

Where aggregate provenance data records interdependencies of socioeconomic events, an individual can appear not only in one activity (e.g. as a sender or receiver) but also in multiple activities across provenances, where one activity (e.g. inheritance) relates to another (e.g. selling or donating). Inheritances are especially promising for the analysis of provenances, as they are not usually recorded in art market data. And yet inheritances remain a key social phenomenon, where gender and wealth, in particular, intersect in meaningful and quantifiable ways. To study gender is to study mechanisms of exclusion within the art market.

A closer look at the behaviour of women who inherit objects can provide knowledge of how women have historically adapted to their positions as inheritors. With aggregate provenance data, it will be possible to study their attitudes toward the wealth passed on to them, for example, and compare their attitudes with those of men. Are they more or less likely to sell their inherited artworks on the market, pass them on to family members, or gift them to institutions? Regarding the latter, it would also be possible to measure gender disparities in philanthropic engagement and how they have developed over time and across geographies.

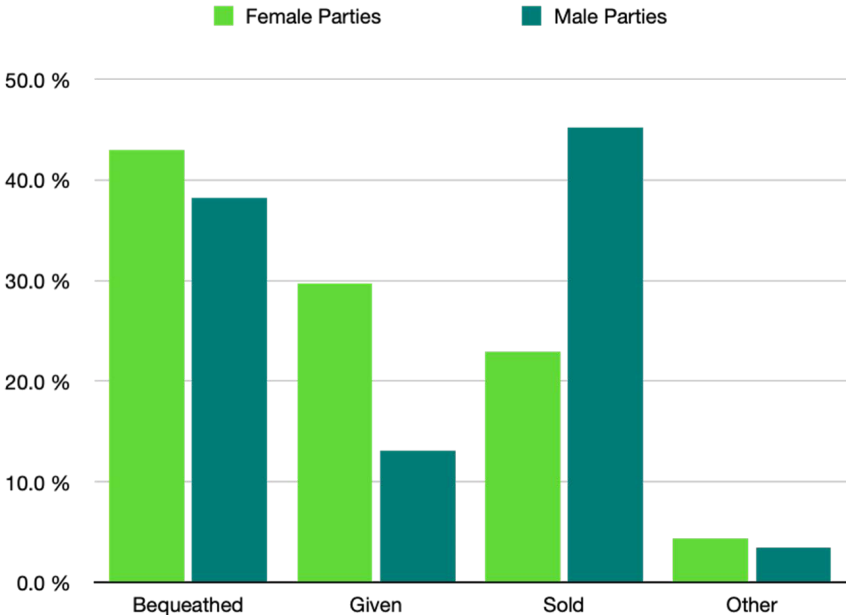
For our preliminary analysis, we identified a total of 3,151 inheritance events in which an owner passed an object to another owner by one of the several types of activity that fall under the umbrella of inheritance.<sup>48</sup> We then analysed the category of the event following the inheritance event. We found that in 41.3 percent of the cases, those parties who inherited a work of art subsequently sold it. In 11.5 percent of the cases, they made a bequest; in 13.5 percent of the cases, they made a donation (96 percent of which went to the Art Institute); and

---

<sup>48</sup> The terms we gathered under the category of "inheritance" are: bequeathed, descended, bequested, inherited, by descent, by bequest, by descendant, by inheritance, bequest, and variations thereof with typographic errors.

in 18.4 percent of the cases, it was the museum itself that inherited the artwork, thereby bringing an end to the object’s provenance.<sup>49</sup>

Based on this information, we were able to analyse the data further. Of the 3,151 inheritance events, only 1,056 occurrences involved individuals, and for 831 of those, we can identify the following event.<sup>50</sup> These 831 events could then be broken down further by gender (fig. 4). If we consider the actions of men (403 events), visualised in figure 4, we see that they are most likely to sell an object they inherited (45.2 percent of the time), followed by making a bequest (38.2 percent) and, lastly, making a donation (13.1 percent).<sup>51</sup>



**Fig. 4:** Comparison of Different Methods of Transfer Following an Inheritance and Their Relative Importance According to Gender. Source: 831 events following an inheritance involving an individual, extracted from provenance texts from the Art Institute of Chicago dataset downloaded on April 7th, 2022; created by Fabio Mariani.

<sup>49</sup> The remaining 15.2 percent were divided among other methods (3.7 percent), the complete absence of information on the next event in the original provenance text (6.4 percent), and the absence of information on the activity of transfer to the immediately following owner (5.1 percent).

<sup>50</sup> We excluded groups of people because they did not represent individual decisions.

<sup>51</sup> Both groups engage in other types of activities as well, but we have grouped them under *other*, given their statistical insignificance.

For women (428 events), the ranking looks different. They are most likely to bequeath an inherited object (43 percent), followed by donating it (29.7 percent). Selling an inherited object is ranked third in their list of activities (22.9 percent), thus highlighting significant behavioural differences between men and women.

While the underlying data for such statistics was taken from one specific collection and can therefore not be treated as representative in the way that a much larger, cross-institutional and geographically diverse set of aggregate provenance data could be, the results nevertheless point out the way for specific avenues of further research. Indeed, by comparing men and women alone and their actions when passing on an inherited object, we can observe stark gender differences. This may, in turn, indicate a broader difference in gendered attitudes towards direct engagement in the market. To what extent such a difference is a product of external factors, such as barriers when accessing the market, or other cultural and societal factors, is a question that would have to be addressed with further research. Suffice it to say, with the help of aggregate provenance data as presented in this paper, we can not only explore large-scale gender dynamics of art circulation but also bring them into relief.

## 6 Conclusion

Transforming provenances into sources for computational analysis by historians is not only a promising endeavour for interdisciplinary research, spanning art, social, and economic history but also technologically feasible. As we have demonstrated, making provenances available as a source for quantitative analysis depends on structuring the information contained within them. Only when we have structured provenances into machine-readable data can we analyse aggregate provenance data. By describing how we can do just that with the help of digital methods, our paper hopes to bridge the gap between computer science, the humanities, and the social sciences.

Our paper has spotlighted several related issues for social and economic historians to consider, which can be addressed with the help of aggregate provenance data: from the socio-economic construction of value to questions of wealth, and the role of gender contained therein. At the same time, we have shown that, based on aggregate provenance data, any element in provenances can be analysed, be it parties, locations, time periods, or methods of transfer. With this possibility to pursue narrow, specialised inquiries into the fates of particular artists, collectors, or institutions, aggregate provenance data also

brings the potential for complex comparative analyses on much larger scales. Moreover, the discreteness of the information found in provenance descriptions allows us to map a whole host of networks across time and geography.

Lastly, we acknowledge that the act of making the information contained within provenances searchable and analysable through digital methods is but a preliminary step in the development of provenance data. Ultimately, efforts to digitise provenances and structure provenance data will deliver a different kind of infrastructure altogether: provenance linked open data. In its connectivity across institutions – including museums, libraries, archives, and other repositories – provenance linked open data promises to deliver unprecedented insights into the circulation of artworks.<sup>52</sup> It is towards this vision of provenance that our paper means to contribute.

**Acknowledgement:** The authors would like to thank the Art Institute of Chicago for making publicly available and downloadable detailed object data on its collection, including provenance information. We are grateful, in particular, to Amanda Block and Jennifer Cohen. We extend our gratitude to the anonymous peer reviewer and to Liza Weber for her rigorous and incisive editing of multiple versions of this article.

## Bionotes

### Lynn Rother

is the Lichtenberg-Professor for Provenance Studies and the Director of the Provenance Lab at Leuphana University Lüneburg. Prior to this appointment, she held research positions at The Museum of Modern Art in New York and the Berlin State Museums working on 20<sup>th</sup>-century provenance and digital initiatives. Her research agenda on provenance brings together art history, the social sciences and digital humanities. Her most recent publication, together with Fabio Mariani and Max Koss, is “Taking Care of History: Toward a Politics of Provenance Linked Open Data in Museums” (Art Institute of Chicago, 2022).

### Fabio Mariani

is Digital Humanities Research Associate at Leuphana University Lüneburg, where he is also PhD candidate. He holds degrees in History and Digital Humanities from the University of Bologna. His current research focuses on semantic web technologies applied to complex historical information. His most recent publication, together with Lynn Rother and Max Koss, is

---

<sup>52</sup> L. Rother/M. Koss/F. Mariani, *Taking Care of History. Toward Provenance Linked Open Data in Museums*, in: E. Canning/E. Fry (Eds.), *Perspectives on Data*, Chicago 2022.

“Taking Care of History: Toward a Politics of Provenance Linked Open Data in Museums” (Art Institute of Chicago, 2022).

**Max Koss**

is a Research Associate at Leuphana University Lüneburg. Max holds an MA and PhD in Art History from the University of Chicago and degrees from the Courtauld Institute of Art (MA) and the London School of Economics and Political Science (BSc). Their research interests are periodical studies, material histories of modern art, and the social and economic history of art. Max’s most recent publication, together with Lynn Rother and Fabio Mariani, is “Taking Care of History: Toward a Politics of Provenance Linked Open Data in Museums” (Art Institute of Chicago, 2022).