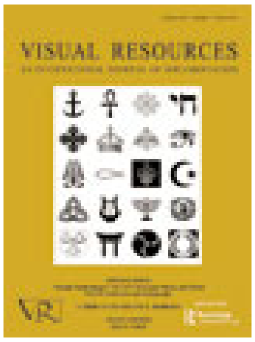

Visual Resources

an international journal on images and their uses

ISSN: 0197-3762 (Print) 1477-2809 (Online) Journal homepage: https://www.tandfonline.com/loi/gvir20

Slow Digital Art History in Action: *Project Cornelia*'s Computational Approach to Seventeenth-century Flemish Creative Communities

Koenraad Brosens, Jan Aerts, Klara Alen, Rudy Jos Beerens, Bruno Cardoso, Inez De Prekel, Anna Ivanova, Houda Lamqaddam, Geert Molenberghs, Astrid Slegten, Fred Truyen, Katlijne Van der Stighelen & Katrien Verbert

Slow Digital Art History in Action: *Project Cornelia*'s Computational Approach to Seventeenth-century Flemish Creative Communities

Koenraad Brosens, Jan Aerts, Klara Alen, Rudy Jos Beerens, Bruno Cardoso, Inez De Prekel, Anna Ivanova, Houda Lamqaddam, Geert Molenberghs, Astrid Slegten, Fred Truyen, Katlijne Van der Stighelen and Katrien Verbert

This paper presents the rationale, genesis, and applications of *Project Cornelia*, an ongoing computational art history project developed by a cross-disciplinary team at the KU Leuven (University of Leuven). It shares practical perspectives acquired while conceptualizing and unfolding the project and discusses successes as well as challenges and setbacks. In doing so, this paper is a cautionary tale for art historians entering the digital arena. However, it is also an invitation to connect to *Project Cornelia*. Art historians seeking to avoid heavy start-up costs and willing to embed their research in a larger empirical and theoretical framework can easily share their data and use *Cornelia*'s data and tools to further their and our understanding of the genesis and governance of early modern creative communities and industries.

Keywords: Digital Art History; Database Design; Data Visualization; Collaborative Research; Creative Communities and Industries; Early Modern Painting and Tapestry

It is a truism that art history has been and continues to be quite slow in exploring and benefiting from the possibilities and promises offered by digital methods and tools. There seem to be two major interwoven reasons why many art historians lag behind or simply decide to bypass new technologies altogether. Firstly, art historians tend to be soloists. Their ingrained traditionalist modus operandi prevents them from working in teams with colleagues and other experts to share raw data and develop large datasets and tools allowing for innovative analyses of art and their makers, as Jim Cuno, President and CEO of the Getty Research Institute, pointed out in 2012:

> We aren't working collaboratively and experimentally. As art historians we are still, for the most part, solo practitioners working alone in our studies and publishing in print and on line as single authors and only when the

ISSN 0197-3762
https://doi.org/10.1080/01973762.2019.1553444

> work is fully baked. We are still proprietary when it comes to our knowledge. We want sole credit for what we write.[1]

A persistent culture of academic recognition geared towards individual achievement obviously nurtures and sustains this egocentric behavior.[2] Secondly, senior art historians have not been trained to use – let alone create – digital tools properly. Even if they would really like to, they cannot develop inspiring and "dramatic proof-of-concept works" themselves.[3] Nor can they truly support graduate students interested in embracing digital methods.[4] Only "strategic funding and hiring initiatives," as Johanna Drucker stressed in this journal's 2013 special issue on digital art history, could break this Catch-22 situation.[5] Meanwhile, however, Pamela Fletcher and other advocates of computational approaches are right in claiming that "an intellectually generative digital art history" is still a work in progress,[6] and that

> there is not yet a density of practice nor a profusion of highly visible and well-received projects. Moreover, many of these pioneering endeavors are "one-off" projects – the result of years of perseverance by small teams – that would be difficult for others to replicate or build upon.[7]

This paper shows how at the University of Leuven, Belgium, a cross-disciplinary team is navigating the challenges and issues highlighted in the literature quoted above. By presenting the rationale, genesis, and applications of *Project Cornelia* and the *Cornelia* database, the aim of this paper is twofold.[8] Firstly, we want to share practical perspectives acquired while conceptualizing and developing this ongoing computational art history project. We will discuss not only successes, but also challenges and setbacks. In doing so, we hope to help art historians who are thinking of starting a computational art history project or have one in the works to bypass or mitigate a number of pitfalls and problems that they are likely to encounter as they tread the digital path. The paper's second aim is to motivate both junior and senior art historians dealing with a research agenda that is very similar to that of *Project Cornelia* to connect, thus helping them to avoid heavy start-up costs altogether.

From Criticism and Constraints to Proof of Concept, 2009–2018

In 2009, Koenraad Brosens, freshly appointed as a tenure-track assistant professor in the Art History Department of Leuven University, presented the first blueprint of what eventually would become *Project Cornelia*.[9] Inspired by the materialist perspective on art history, socio-economic literature and, in particular, Becker's *Art Worlds* (1982),[10] Brosens set out to analyze how iconographic and stylistic developments in early modern Flemish tapestries were shaped by the complex social and economic interactions between the inhabitants of the "art worlds," in particular the milieux of tapestry producers and painters, and vice-versa. In order to reconstruct "network[s] of cooperative links among participants,"[11] Brosens needed a myriad of attribution and relational archival data extracted from various archival sources, including parish records, notarial deeds, and registers of guilds and corporations. The data had to be fed into "powerful software programs that enable computations of a wide array of network measures [and that] also allow for the matrices to be converted into dynamic, interactive and readable two-D and three-D visualizations of networks."[12] For a while, Brosens dabbled with programs such as Excel, FileMaker, UCINET and NetDraw to organize data and to visualize multivariate networks of Brussels tapestry producers (i.e. networks where the nodes and edges have multidimensional attributes). These first baby steps in the digital world gave him a tantalizing glimpse of the promises hidden in a full-blown computational methodology. Yet they also revealed the manifold challenges and complexities of the approach – most of which, of course, resulted from his lack of computer skills. In short, as it was both high-risk and time-consuming, the digital agenda seemed incompatible with the tenure track requirement at the time.

Therefore, Brosens decided to spend most of his time on familiar analogue territory. A traditional tapestry scholar by day, he patiently laid the groundwork for a digital track by night. He did so by submitting applications to Belgian and European grant programs for a computational "tapestry history meets network analysis" research project. From the very start, Brosens' ambition was to develop a database and visualization tools that could have a life beyond his immediate interest in the Flemish tapestry worlds. For this reason, they would have to be designed in such a way that other researchers could also use them to process and visualize their data, which would be linked to the data already in the database. Brosens reached out to his Leuven colleagues Katlijne Van der Stighelen from the Art History Department and philosopher, statistician, technical guru and database expert Frederik Truyen from the Cultural Studies Department, who agreed to become co-advisors.

The first attempts (2010–2012) to convince funding agencies to support such a project failed miserably. Admittedly, to a certain extent the applications were marked by rather naive, "blue-sky" thinking. Interestingly, however, many reviewers did not even engage properly with the proposal. Instead, they displayed a vivid disbelief in digital approaches – while usually revealing their unfamiliarity with digital technology by grossly underestimating, for example, the level of intellectual sophistication required to conceive and implement a multidimensional data model and innovative data visualization methods.

In subsequent applications, Brosens aimed to defuse this hostile view of digital research in the field of art history. He made sure to underscore the fact that both the data and the overarching research questions driving the project were, to all intents and purposes, "traditional." In addition, he slightly downplayed the importance of digital methods and tools. With the project proposal redirected back to reviewers' methodological comfort zone, the new applications were successful. Both Leuven University and the Flemish Science Foundation-Belgium (FWO-Vlaanderen) gave initial funding for the project, then called *MapTap: Mapping the Antwerp-Brussels-Oudenaarde Complex (1600–1700) Via Network Analysis*) in the autumn of 2012.

The funding enabled Brosens to hire two PhD students for a four-year period (2012–2016). His intention to employ a computer scientist, however, was quickly curtailed. Institutional regulations forced him, as a professor belonging to the University's Humanities and Social Sciences Research Group (HSS), to enlist PhD students in the doctoral school of HSS. Master students holding a degree in Computer Science (CS) were not eligible. They could only enroll in the doctoral school of the Science, Engineering and Technology Research Group (SET). SET did not allow HSS professors to supervise PhD research in SET. As a result, Fred Truyen, head of the CS digital media lab, remained a one-man digital army, while art historian Klara Alen and historian Astrid Slegten joined the project team.

It immediately transpired, however, that the project's agenda to create a sustainable database that could host a highly varied assortment of archival data and be of interest to a wider community of scholars was challenging, to say the least. The development of the data model, ontology, and schema proved to be a long process of trial and error. When Alen, Brosens, and Slegten presented Truyen with another peculiar type of archival evidence (e.g. a probate inventory recorded in 1685 revealing that the deceased had issued a bill of exchange and a loan in 1665 and 1667 respectively), he was forced to revisit the data model. Although this time-consuming phase obviously conflicted with the "publish or perish" pressure felt by all involved, it was also a very stimulating intellectual endeavor. The attempts to process the sundry data triggered lively philosophical discussions on the nature of archival sources and archival evidence. These discussions helped the team to rethink the ways in which data can help us to model the past – or trick us in believing we can do so, as our approach highlighted a number of inconvenient truths, including the omnipresence of missing data and the resulting biases and possible misconceptions.

By the beginning of 2016, *Cornelia* was a functional Microsoft Access database. At that time, the database included records for some 4300 people from the seventeenth-century Antwerp and Brussels "tapestry worlds." In an essay published in the *Zeitschrift für Kunstgeschichte* that year, we presented *MapTap*'s research philosophy (we coined the expression "slow digital art history" to delineate our iterative methodology and inclusive approach to archival material); its core methodological framework/conceptual toolkit (formal historical network analysis); *Cornelia*'s key concepts, and, finally, a case study showing the crucial role played by women in the Flemish tapestry industry.[13] By calling ourselves "slow digital art historians," we incurred the animosity of one of the peer reviewers, who believed that we felt intellectually and morally superior to "fast" digital art historians. But we had coined the expression to stress that our approach forces us to proceed slowly. Seeing that we deal with a myriad of *sources* and include all of the *actors* who engage in one or more *events* and play one or more *roles* that can link them to one or more *actors*: *groups* (i.e. cultural, economic, political, social and/ or religious bodies); *places* (i.e. countries, town, parishes, streets, and/or houses); and/or *works of art* in *Cornelia*, populating the database is indeed a slow process. One baptism record, for example, usually gives us five *actors* producing one *event* and playing five different *roles* that link them to each other (in 10 different ways) and to a *place* (i.e. a parish in a town) (Figure 1A). We always include *all* elements and make sure to link them directly to their source. We keep digital images of all sources that are processed; in this way, data provenance, data checking, and data cleaning will never present an issue.

Our slow and inclusive approach has three more interwoven positive effects. Firstly, as we do not pretend to know who "deserves" to be included in the database and as we do not believe that we can ignore any of the roles and relationships documented in our sources, we sidestep the confirmation bias that is typical of most art-

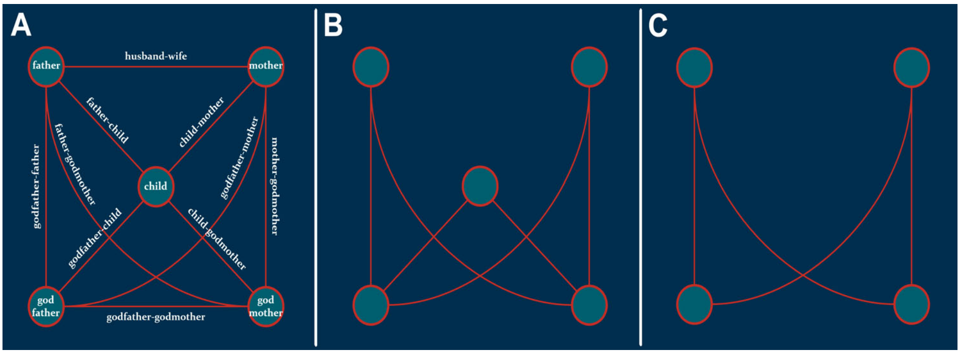

Figure 1. The transformation of a multiplex baptism network [A] into a uniplex network showing godparenthood ties [B] and a uniplex network showing godparenthood strategies [C]

historical research. In addition, by listening patiently to the archival documents, we allow for a serendipitous "record data first, ask questions later" approach. Finally, as we meticulously date, identify, and label all data, we can transform the countless multiplex and multimode networks included in *Cornelia*, which would typically be visualized as cluttered and unreadable "hairballs", into partial unimode networks (i.e. networks showing one kind of relationship) that can be analyzed properly through computation and visualization. Figure 1, for example, shows how we process crowded and surprisingly complex baptism networks [A]; and how we can transform them into simple yet truly meaningful networks showing godparenthood ties [B] or even godparenthood strategies [C] that reveal friendship and trust in and between families.[14]

An unpublished case study presented as a work-in-progress at a number of symposia in 2015 and 2016,[15] illustrates how our slow approach yields interesting research results. By aggregating networks revealing godparenthood strategies devised by Antwerp and Brussels tapestry producers throughout the seventeenth century, we get a new view on iconographic and stylistic features of Antwerp and Brussels Baroque tapestries. Figure 2 shows that while there were no godparenthood ties between the

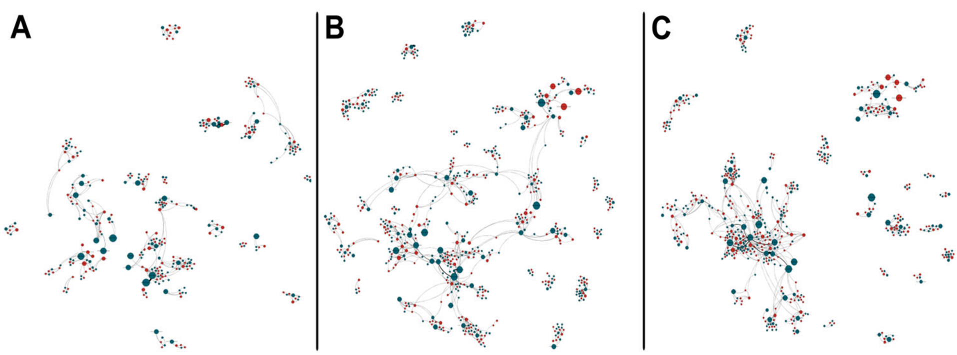

Figure 2. Godparenthood networks in the Brussels (bottom left) and Antwerp (top right) tapestry worlds between 1621 and 1640 [A], 1641 and 1660 [B] and 1661 and 1680 [C]

Antwerp and Brussels tapestry worlds in the periods 1621–1640 and 1661–1680 [A and C], this kind of intercity relationship did exist between about 1641 and 1660 [B]. Interestingly, a number of so-called Brussels tapestries woven during this period (i.e. tapestries woven after cartoons that were commissioned by Brussels tapestry producers, bearing borders typical of Brussels workshops, and sometimes bearing the Brussels city mark) are of lesser quality than usual, which is noticeable in both the materials and the weaving. While it has been argued that this was due to the inability of Brussels weavers to understand fully and operationalize the innovative designs by Rubens and his circle,[16] Figure 2 actually strongly suggests that the intercity ties of trust mitigated the rivalry between the production centers – which had been fierce throughout the first third of the seventeenth century[17] – and could very well have inspired Brussels producers to lend their cartoons (or copies of the cartoons) to Antwerp colleagues who employed less skilled weavers working with cheaper materials. Thus, Figure 2 helps us to revisit questions of attribution and dating of Antwerp and Brussels tapestries.

Figure 2[A] and Figure 2[C] also show different patterns in both Antwerp and Brussels networks of godparenthood strategies between about 1621 and 1640, on the one hand, and 1661 and 1680, on the other. The data visualizations suggest not only that in the latter period there were fewer tapestry producers, but also that they built denser networks than their predecessors. In doing so, they may have stifled competition and possibly artistic innovation. While this particular hypothesis needs testing, it is clear that what is shown in Figure 2 and similar network visualizations have great potential to contribute to the wider debate on the mechanics of collaboration, friendship, and innovation in creative communities.

By 2016 *MapTap* had proven its viability and was ready to move forward. Brosens managed to obtain additional funding from Leuven University and the Flemish Science Foundation-Belgium. *MapTap* grew into *Coral: The Interplay between Social Structure, Collaboration and Innovation in Flemish Painting and Tapestry Design (1600–1650)*. Together, these two initiatives comprise *Project Cornelia*. For this new phase, Brosens hired three PhD students: art historians Rudy Jos Beerens and Inez De Prekel, and historian Cara Pelsmaekers. Pelsmaekers was in the first class of students who graduated from Leuven University's brand new Digital Humanities program. So Truyen's digital unit doubled in size – finally. After several months, however, Pelsmaekers was headhunted by a private company.[18] Meanwhile, *Project Cornelia* had made overtures to experts in Leuven University's CS Department and Department of Electrical Engineering (ESAT). Katrien Verbert, head of the Human-Computer-Interaction research unit (HCI) in the CS Department, became interested in *Project Cornelia*'s digital ambitions, as did Jan Aerts, head of ESAT's Visual Data Analysis group. Bruno Cardoso, PhD in CS and specialized in HCI and Multimodal Systems research, and Houda Lamqaddam, master in CS, worked on a consultancy basis on the migration of the Access database to an Apache/MySQL/PHP stack (Cardoso) and on an innovative interactive data visualization tool (Lamqaddam).[19] In early 2018, both Cardoso and Lamqaddam joined *Project Cornelia* as full-time team members. In the winter of 2017–2018, Leuven University decided to lift some of the restraints on intergroup PhD research, so that Lamqaddam became the very first PhD student enrolled in the doctoral school of SET who has an advisor from both SET and HSS. Thus, *Project Cornelia* is now being developed by a cross-functional team with expertise in art-historical research, agile software development, and HCI research. As of February 2018, the *Cornelia* database included some 11,000 actors and no fewer than 320,000 time-dependent edges extracted from approximately 8500 archival entries.

Painting a Triptych

In February 2018, we launched www.projectcornelia.be. In addition to basic information on the project, blog posts, and a list of publications and talks, the website invites users to explore *Cornelia* as if it were a triptych.

The central panel is called *Actors.* In this panel, users can search for men and women recorded in one or more archival collections that can be selected from a dropdown menu. For the time being, users have access to two of *Cornelia*'s archival collections: *Parish Records (PR)* and *Guild Records (GR).* Since the Antwerp and Brussels archives hold countless volumes listing thousands of baptisms, funerals, and weddings, it is nearly impossible to process *all* parish records. Consequently, *Cornelia* is both poor and rich in parish records data, depending on the actors that users are targeting. While processing guild records, we follow a different rationale. We take *all* data included in the volumes. At this point, users have access to two volumes, both playing key roles in research on Flemish Baroque painting. One is register 818, which is preserved in the Brussels *Rijksarchief.* This volume lists hundreds of apprentices, masters, and deans of guilds recorded in the Brussels corporation of painters, gold-beaters, and stained-glass makers between 1599 and 1706.[20] The other volume is register 201, preserved in the Antwerp *Artesis Hogeschool.* This volume lists hundreds of apprentices, masters, and deans recorded in the Antwerp guild of Saint Luke between 1629 and 1660.[21] The members of this guild included a variety of "art actors," including engravers, painters, and sculptors. While entering actors into the database, we do not use any vocabularies to standardize the data, as this would force us to "erase" all ambiguity inherent in seventeenth-century archival sources. We first clean the data, i.e. we deal with inconsistencies in the spelling of names and identify unique actors, before checking them – wherever possible – to the relevant records in the Getty's *Union List of Artist Names* (ULAN) vocabulary.

As of this writing, we are processing additional volumes and plan to open up more data and collections in the near future. However, we continue to move slowly, and in fact our pace has decreased slightly. This is because, as more actors find their way to *Cornelia,* the inconsistencies in the spelling of names and the fact that many people had the same or very similar names, makes the phase of data cleaning increasingly time-consuming (Figure 3). The alternative, however, would be to present dirty data, which, of course, would confuse rather than illuminate users and justify the reservations of the naysayers.

Users are invited to enter a text string (i.e. a first or last name) in a search box. While typing the name, a dropdown list suggests possible matches. If they wish, users can ignore these suggestions and opt for a free search. *Cornelia* then produces a list of names that are identical or similar to the search term. The piece of code generating this list takes both sound and spelling into consideration. For example, the

Figure 3. Slow digital art history at work: data cleaning

search term *wouter* triggers a list that includes not only actors whose first name is *wouter*, but also actors whose family name is *wouters*, *wauters* and *wautiers. Cornelia* further suggests names such as *coster*, *winter* and *zoute*, as parts of these text strings are identical to the search term. We intentionally put the threshold for matches quite low, as this enables users to catch typographical errors and to use or discover name variants (*hubert*, for example, will also return *huybrecht* and even *robert*). This, of course, also inspires serendipitous searches.

For each search request, *Cornelia* returns an identity card. As there are now two collections that are accessible, there are two types of cards. *PR* cards include baptism, wedding, and funeral dates for an actor, together with the locations of the events, the names of the men and women who were witnesses at the baptism (i.e. the parents and godparents) and the wedding(s), and a reference to the archival source. *PR* cards further list children and godchildren and the years in which they were baptized. If *Cornelia* has more than one parish record event for an actor, the card shows a biographical timeline. Seeing that *Cornelia* does not include *all* parish records, *PR* cards typically have one or more empty fields.

GR cards show the dates or years actors were enlisted as apprentice, master, and/or dean. The cards further list the name(s) of the actor's teacher(s) and student(s) and his or her occupation. *GR* cards can also include pieces of biographical data (i.e. father, children and birthplace), as guild records sometimes mention these elements. Finally, the cards show a timeline recording guild events. Since not all actors had "full" careers, and because not all entries in a volume are equally detailed (because different officials recorded entries over the years), *GR* cards, like *PR* cards, usually have one or more empty fields.

By clicking on the names of children, godchildren, teachers, and students, or by entering a new name in the search box, users can produce new cards. These are shown above older ones, enabling users to scroll up and down through their search history. Thus, *Cornelia*'s central panel not only allows users to retrieve essential biographical and professional data on actors, but also to catch a first glimpse of their family and professional networks.

The left wing of the triptych, *Networks*, will handle these constantly changing structures in detail. While we used and still use Gephi, a powerful open graph viz. platform,[22] to reconstruct networks (as seen in Figure 2), we often find that Gephi, just like other existing data visualization tools, frequently falls short of our expectations as we try to make our multivariate networks more accessible and understandable. Therefore, we are developing a new interactive tool that will allow users to explore the dynamics within the family, social, and economic networks of their choosing. Almost by definition, these networks are very complex, for actors continuously enter and leave the system while different types of relationships emerge, continue, fade away, or vanish over time. In order to facilitate the readability of the networks and to maximize the analytical potential of *Cornelia*'s relational data, the new tool will allow users to access three complementary views. Thus, *Cornelia*'s left wing will become a triptych within a triptych.

The wings of the triptych will be based on models that can readily be found in data visualization libraries. One wing will show a traditional genealogical tree. This familiar view of a family network has the advantage of clearness and definiteness. This clarity, however, comes with sacrifices and potential disadvantages. To ensure readability, for example, the number of both actors and relationships needs to be limited. In addition, the clear-cut stratification of time, depicted as a sequence of generations, is a somewhat misleading attempt to straitjacket what in essence was a lottery of life spans. The other wing will present a graph view not unlike those presented by other computational art history projects that deal with networks.[23] This view allows users to discover both clusters and people bridging different networks, thus revealing or at least suggesting "importance." However, graph views tend not only to suspend or collapse time, but also to conflate different types of ties. In addition, "importance" has many meanings, depending on the type(s) of relationships that produce the view.

The central panel, which as of this writing only exists as a demo with mock data but is being transformed into a full-fledged tool, presents an innovative view that aims to tackle the most important caveats and pitfalls of the tree and graph views.[24] The tool allows users to select an actor of interest. It then visualizes the actor surrounded by parents, siblings and children, together with partner(s) and the latter's parents (Figure 4). All actors have bars that depict their life span. Icons show the gender and occupation of the actors. When the user hovers over the names basic biographical information is provided. Basically, the initial view presents a slightly enriched yet essentially familiar genealogical tree rotated 90 degrees counterclockwise.

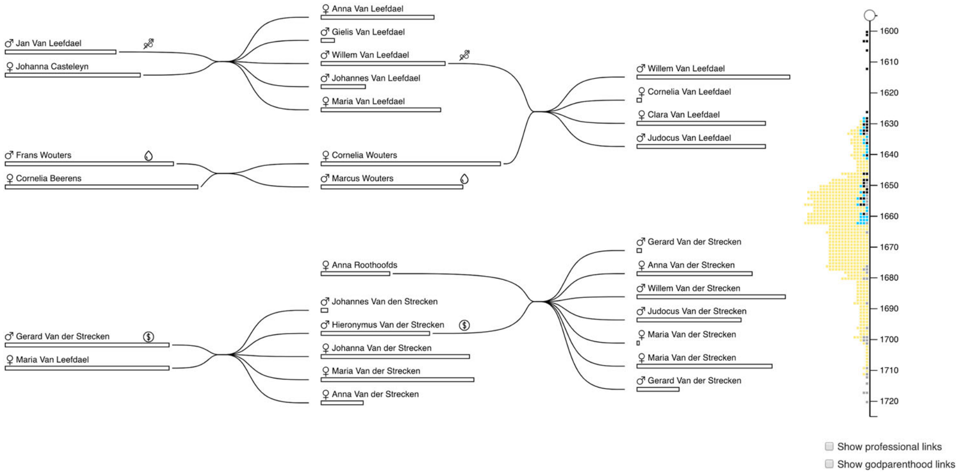

Figure 4. *Project Cornelia*'s network tool: default view

This view, however, is just the starting point. Users can explore the *dynamics* of the network – the intermingled effects of time and economic and social behavior – in different ways. They can, for example, select and depict professional relationships and/or godparenthood relationships linking the actors. Other relationships, such as neighborhood ties and ties revealing conflicts between actors, will be added in the future. Depending on the number of selected relationships included in *Cornelia*, however, the visualization can quickly become difficult if not impossible to read, not in the least because time is conflated.

A time slider on the left solves this problem, as it allows users to walk slowly through time. By moving forward in time, the bars indicating the actor's lives are filled gradually – that is, the actors are born and age. After their death, they are not removed from the network, but are depicted in gray, for their memory and/or influence can last. By walking through the seventeenth century, users not only discover the network configuration of a specific year, but also witness how this configuration changes from year to year. Users can also travel back in time, take screenshots, and use them as working documents.

We designed the time slider so that it helps users to identify interesting years and periods. Right next to the slider, we show small color-coded cubes that represent the number of different links that are included in *Cornelia* per year (Figure 5). Put differently, by simply looking at the time slider and the color-coded cubes, users can immediately see *when* particular types of relationships and networks emerged, blossomed, and faded away. Even without manipulating the slider and reconstructing networks, users can easily discover the sequence of different types of relationships and the pace or rhythm with which they appear. Figure 5, for example, shows that in this case professional relationships (blue) preceded godparenthood relationships (yellow). It further reveals that for about a decade, collaboration went hand in hand with the development of godparenthood relationships, while after

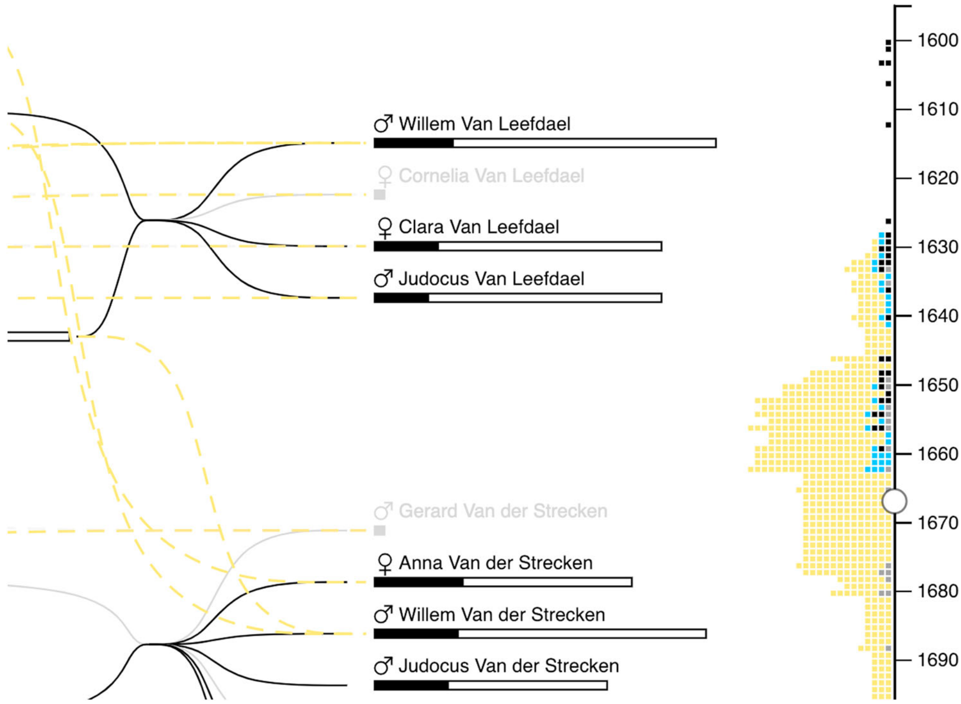

Figure 5. *Project Cornelia*'s network tool: color coded cubes representing different ties included in the database

1663 the families were still linked by godparenthood relationships, but no longer collaborated.

Finally, *Cornelia*'s right wing, called *Frameworks*, provides a platform for longitudinal and quantitative analyses of creative communities and industries as a whole. Again, our slow approach and *Cornelia*'s data model can transform complex sources, such as the Brussels register 818 listing hundreds of apprentices, masters, and deans recorded in the Brussels corporation of painters, goldbeaters, and stained-glass makers between 1599 and 1706, into an accessible and powerful data set. By aggregating data on the careers and profiles of all actors in register 818, for example, we get a better understanding of the professional mobility and the career choices and strategies of artists and craftsmen in Brussels.[25] The analysis shows, among other things, that what we might call a "complete" career – one in which a single actor took on each of the available roles (apprentice, master, teacher, dean) at least once – was far from the norm. Only one in four apprentice painters (at least, among those who were not masters' sons) became masters, and they took an average of 10 years to make this transition. Only 39% of all masters took on one or more apprentices. A relatively small proportion of all masters, just 10%, trained over half of all apprentices. Only one in five masters, most of them masters' sons, served on the board for one or more terms. Again, data visualizations, in this case made in Tableau, help both us and other users to understand and engage with the data (Figure 6).

Figure 6. Enlistment of apprentices in the register of the Brussels guild of painters, goldbeaters and stained-glass makers, by occupational group and by year, 1599–1706. The number of apprentice painters clearly decreased throughout the century as opposed to the other occupations. The fall in the number of apprentice painters does not necessarily mean that workshops shrank, or that their output declined. Former apprentices, of whom there was an abundant supply given the large numbers who had enrolled earlier in the century, could always be hired as assistants.

Obviously, as we continue to enter more data in *Cornelia*, the power of this and similar analyses will grow stronger. By processing sources similar to the Brussels register, such as that of the Antwerp guild of St Luke, for example, we are adding comparative data that will help us to arrive at an even better understanding of the development of artistic careers and the population of artists in different cities.

In sum, the *Project Cornelia* triptych aims to operationalize and optimize the analytic potential of *Cornelia*'s data by giving users the flexibility needed to address their various research questions. In addition, it accommodates users trying to make serendipitous discoveries and to find unexpected questions. However, as most art historians are not trained to operate this kind of data-driven toolbox – and might even get lost in the triptych – we will study their behavior while exploring and using all the panels. This human–computer interaction research will help us to refine the tools and to develop an online environment that will allow users to reconstruct and explore data and networks economically, efficiently, and intuitively.

However, while we keep on populating the database and constructing the triptych, there is one thorny issue that grows along with Cornelia: missing data. If we arranged our data set in a matrix (resembling a rectangular Excel sheet), where the 11,000 rows correspond to *Cornelia*'s actors and the hundreds of columns correspond to as many events (or variables) (such as "baptized in," "registration as an apprentice painter," "registration as a master painter," "recorded as dean of the guild of painters," and "buried in"), there would be thousands of elements (or cells) showing values. These could reveal, for example, that actor number 157 was recorded as an apprentice painter in 1623; that he became a master painter in 1649; and that he became dean of the guild of painters in 1667. However, other cells, such as "baptized in" and "buried in," could be empty, together with thousands of other cells pertaining to other actors. The pattern of empty cells would be arbitrary. This is by no means exceptional. In fact, "missing data is one of the most important statistical and design problems in research,"[26] especially in research fueled by longitudinal data of networks and behavior.[27] Though not all empty cells are literally missing data (for example, not all 11,000 actors embarked on a career as painter, so we cannot expect values for all "registration as an apprentice painter" or "registration as a master painter" variables), it is easy to understand that missing data poses a major threat to our ambition to study the dynamics of multivariate networks. If too many variables have no values, we cannot pinpoint when nodes appear and disappear, nor when relationships were established, changed, and terminated. How, then, can we address this issue?

Filling in the Blanks

Art history traditionally relies on two intuitive methods to deal with the issue of missing data. These methods are also used in other social sciences and humanities research.[28] Firstly, a great deal of art-historical scholarship focuses on a limited number of artists and famous if not iconic works of art. In doing so, the field has the tendency to see data sparsity and missing data as a nuisance rather than a threat, and to delete – albeit tacitly and sometimes unwittingly – actors and/or sets of variables that have a significant amount of missing data. Seeing that it is our ambition to reconstruct and study multivariate networks that include all inhabitants of the seventeenth-century Antwerp and Brussels art worlds as they appear in the sources, including actors whose voice sounds weaker (such as women), we avoid this procedure and potential confirmation bias.

Secondly, art historians frequently apply best-guess imputations, i.e. the use of available data to formulate an educated guess for a missing value. Best guesses do not sacrifice data points and usually seem fairly straightforward and accurate, although art historians tend to produce best guesses inside a personal "black box" containing insights gained from literature, expertise, gut feeling, and "common sense." As a result, best guesses for art-historical data points are often imprecise and divergent, producing "noise" in a data set. The authoritative database of the *Rijksbureau voor Kunsthistorische Documentatie* (RKD) in The Hague, for example, states that the Brussels painter and artistic powerhouse Antoon Sallaert was born "ca. 1580–1585", whereas the equally respected *Union List of Artist Names* compiled by the Getty Research Institute (GRI) and Grove Art Online (GAO) (via Oxford Art Online) give "ca. 1590" and "c. 1580–90" respectively. The RKD, the GRI, and GAO probably used the year in which Sallaert became a master painter (1613) and/or the year in which he died (1650) to guesstimate his birth year. Thus, these tools highlight our problematic understanding of seemingly straightforward yet essentially complex issues such as career trajectories and life expectancy in early modern times.[29] It is possible to mitigate these problems by producing a best guess for Sallaert's date of birth that is based on the best guesses found in RKD, GRI and GAO – which could be “01/01/1585,” for example. However, it is clear that handpicking best guesses is a cumbersome process that will still leave hundreds of empty data points unaddressed.

Thus, while art history’s traditional ad hoc strategies and methods to tackle missing data can certainly be used to a certain extent to address qualitative and micro-historical questions, they fall short of meeting our data-driven research agenda. Consequently, we turned to the field of statistical computing.

First, we identified smaller, homogenous subsets of data – we reduced the number of rows and columns of the matrix – as this enabled us to test out different data imputation methods. We decided to focus on a subset of seventeenth-century Brussels painters with five variables (“year of birth,” “registration as an apprentice painter,” “registration as a master painter,” “registration of first apprentice painter,” and “year of death”), as most painters have missing values for one or more of these variables. Also, we dealt with the issue of missing values by deleting, so to speak, a number of data points. This allowed us to assess the quality of the computationally derived outcomes.

We then applied single imputation in R.[30] This method substitutes a missing value with a suitable replacement value. There are different ways to produce replacement values, including mean imputation, regression imputation, and stochastic regression imputation. Though statistical literature warns against using mean substitution, especially when dealing with multivariate data sets,[31] we decided to explore this method, as it echoes art history’s familiar and much-used best-guess imputation. This method replaces missing values for “year of birth,” for example, with the average of the observed values for that variable. To calculate the average, we can use the time lag between “year of birth” and one of the four other variables, such as “registration as an apprentice painter.”

By using the latter variable, mean substitution predicted, for example, 1590 as the year of birth of Antoon Sallaert, which happens to coincide with the GRI’s approximate date of birth for this artist. However, when we used the three other variables (i.e. “registration as a master painter;” “registration of first apprentice painter;” “year of death”), the model predicted 1586, 1579, and 1580 respectively. While these results may tie in with the best guesses discussed above, they also underline the fact that mean imputation presents serious problems. If Sallaert had entered the network in 1579, he must be regarded as a contemporary of the omnipresent star player of Flemish Baroque painting, Peter Paul Rubens (1577–1640). But if he entered in the network in 1590, he must be seen as belonging to the first generation of painters following in the wake of Rubens. These, of course, are two very different settings guiding our understanding of Sallaert’s position and ties within the network of Baroque art.

Interestingly, *both* settings and the resulting networks, and especially the first one, would be misleading. In the case of Sallaert, we faked missing values. His baptism record is included in *Cornelia*. It shows that the artist was born in 1594. This means that, if we had allowed Sallaert to enter the network in 1579, relationships that could have been expected would have been absent (as in actual fact Sallaert entered the network 15 years later). This could easily have led to interesting yet also inaccurate claims and hypotheses, such as: "Sallaert had no ties with Rubens or his Antwerp contemporaries, which suggests that both artists developed the new Baroque style independently in Antwerp and Brussels."

As we saw that single imputation increased rather than reduced the level of uncertainty in the data set, in 2017 we decided to reach out to yet another team of domain experts, i.e. the Leuven Biostatistics and Statistical Bioinformatics Center (L-BioStat) directed by Geert Molenberghs. Together with Ana Ivanova, one of Molenberghs' PhD students, we first tried to understand the nature and distribution of "missingness" in the database in order to develop the best possible imputation method. It transpired that *Cornelia* had all three types of missing data identified by missing data experts: data that is *Missing Completely At Random* (MCAR; the propensity for a data point to be missing is independent of any values in the data set, observed or missing), *Missing At Random* (MAR; the propensity for a data point to be missing is not related to the missing data, but it is related to some of the observed data); and, finally, *Missing Not At Random* (MNAR; the propensity for a data point to be missing is related to the reason why it is missing).[32] A page that is missing from an otherwise complete register of baptisms, for example, creates missing baptism data that is MCAR. A somewhat sloppy official who failed to detail all information in the register of the Brussels guild of painters while he was in charge of keeping the register produced data that is MAR. We can identify the official as a variable that can be used to predict missing data: as long as we observe this official at work (through his handwriting), we can predict that the data will be incomplete. Finally, a fraudulent painter working in Antwerp who refused to comply with the rules imposed by the guild of Saint Luke and never enrolled as a master produced missing data that is MNAR.

However, since most of *Cornelia*'s missing data is MAR, we can address the issue by applying multiple imputation (MI).[33] MI is regarded as one of most effective methods for missing data handling and analysis in many fields. Moreover, the technique is still being refined.[34] However, it is anything but widely used in the humanities, not least because of the high level of mathematical thinking required to understand fully and apply MI. Yet, on a more conceptual level, the MI framework is quite straightforward. First, missing values are imputed m times from a distribution of similar records. Usually a small number of imputations (5 to 10) is sufficient, unless the amount of missing values is exceptionally high or the data set does not include a large amount of information to model the probabilities of "missingness." The first step results in m completed data sets that all have plausible values for missing data points. By computing the mean over the m completed data sets, its variance, and its p value, MI produces a final and statistically valid result.

However, the idiosyncratic nature of the *Cornelia* data and the high levels of "missingness" make MI challenging at times. Our first attempts generated predictions that were both slightly off the mark and highly accurate. For example, the model was spot on in claiming that Antoon Sallaert was born in 1594. As the quality of the predictions will get better as the data set grows, we intend to refine and develop our use of MI in the near future.

Conclusion

We hope that we have shown that computational art history, and in particular our kind of "slow" digital art history, which aims to operationalize large amounts of mostly unused archival data in order to understand the socioeconomic reality that shapes artistic developments and processes (and vice versa), can be rewarding on many levels. We also hope that, by underscoring that *Project Cornelia* was, and to a certain extent still is, a process of trial and error, we can help aspiring digital art historians to avoid making the same mistakes.

While one of the goals of this essay is to advocate digital art history, we realize that this paper can also be read as a "Don't Try This at Home!" warning. This is because we do not want to minimize the many interwoven challenges that early and mid-career art historians in academia will have to deal with if they embark on a truly digital adventure. They will need time to overcome numerous problems and setbacks that are typical of any innovative research that clashes with the "publish – as a single author, and in print – or perish" dynamics. They will have to reach out to colleagues and develop cross-disciplinary teams, while the culture of academic recognition and even institutional restraints frequently undermine if not outright block incentives to collaborate. They will have to develop interactional expertise – that is, they will have to master the language of a specialist domain (such as CS, HCI, data visualization or statistics) – in order to create a fruitful group dynamic leading to hybrid research questions and research outcomes. They will have to exercise both methodological diplomacy and stubborn persistence, as they will have to convince evaluating and funding committees that are likely to be populated with outspoken non-believers who feel that a great deal of digital art history is gimmicky rather than fundamental. In short, they will experience the reality that the digital arena is expensive in both time and resources.[35]

This is one of the reasons why we invite both junior and senior researchers to connect to *Project Cornelia* and avoid heavy start-up costs. An easy-to-use back end is accessible online. This allows users to process archival data themselves and to use *Cornelia*'s data and tools. As we understand that art historians tend to feel the need to be the "owners" of their data – at least for a couple of months or years – and/or to be acknowledged as the discoverers or re-discoverers of data, "ownership" and participation can be claimed and signaled in different ways.[36] Perhaps we were naive back in 2009. Perhaps we are naive now. But we hope that sometime in the future, tools like *Cornelia* will play a key role in a truly international collaborative effort to reconstruct and analyze networks and frameworks fueling creative communities and industries in early modern Europe.

Acknowledgements

The authors wish to thank Koenraad Matthys (University of Leuven, Centre for Sociological Research) for his help and support throughout the years.

Disclosure Statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the University of Leuven [OT/12/022 and C14/16/016] and the FWO-Vlaanderen [G044513N and G057117N].

KOENRAAD BROSENS is a Professor of Art History and Chair of the Art History Department, University of Leuven.

JAN AERTS is a bio-engineer and Professor in the Visual Data Analysis Group of the ESAT/Stadius Center for Dynamical Systems, Signal Processing and Data Analytics, University of Leuven.

KLARA ALEN is an art historian in the Art History Department, University of Leuven.

RUDY JOS BEERENS is a PhD student in the Art History Department, University of Leuven.

BRUNO CARDOSO is a postdoctoral researcher in Computer Sciences (Human-Computer Interaction) affiliated with the Art History Department, University of Leuven.

INEZ DE PREKEL is a PhD student in the Art History Department, University of Leuven.

ANNA IVANOVA is a postdoctoral researcher in Biostatistics at UHasselt (University of Hasselt) and the University of Leuven.

HOUDA LAMQADDAM is a PhD student in the Computer Sciences and Art History Departments, University of Leuven.

GEERT MOLENBERGHS is a Professor of Biostatistics at University of Hasselt and the University of Leuven.

ASTRID SLEGTEN is an historian and globetrotter.

FRED TRUYEN is a Professor of Cultural Studies, University of Leuven.

KATLIJNE VAN DER STIGHELEN is a Professor of Art History, University of Leuven.

KATRIEN VERBERT is a Professor of Computer Sciences, University of Leuven.

Notes

1 James Cuno, "How Art History is Failing at the Internet," *The Daily Dot* (19 November 2012) <https://www.dailydot.com/via/art-history-failing-internet/>.

2 Georg Schelbert, "Art History in the World of Digital Humanities: Aspects of a Difficult Relationship," *kunsttexte.de – E-Journal für Kunst- und Bildgeschichte / Renaissance* 4 (2017): 7 <http://dx.doi.org/10.18452/18694>.

3 Johanna Drucker, "Is There a 'Digital' Art History?," *Visual Resources* 29, no. 1-2 (2013): 12.

4 Johanna Drucker, Anne Helmreich, Matthew Lincoln, and Francesca Rose, "Digital Art History: The American Scene," *Perspective* (2015): 12.

5 Drucker, “Is There A ‘Digital’ Art History,” 12.
6 Pamela Fletcher, “Reflections on Digital Art History,” *caa.reviews* (18 June 2015) caareviews.org/reviews/2726.
7 Drucker et al., “Digital Art History,” 11.
8 www.projectcornelia.be.
9 His paper was presented at the *Family Ties* symposium (University of Leuven, 10–11 December 2009). It was published in 2012 as Koenraad Brosens, “Can Tapestry Research Benefit from Economic Sociology and Social Network Analysis?,” in Koenraad Brosens, Leen Kelchtermans, and Katlijne Van der Stighelen, eds., *Family Ties: Art Production and Kinship Patterns in the Early Modern Low Countries* (Turnhout: Brepols), 43–52.
10 Howard S. Becker, *Art Worlds* (Berkeley: University of California Press, 1982).
11 Becker, *Art Worlds*, 35.
12 Brosens, “Can Tapestry Research,” 47.
13 Koenraad Brosens, Klara Alen, Astrid Slegten, and Fred Truyen, “MapTap and Cornelia: Slow Digital Art History and Formal Art Historical Social Network Research,” *Zeitschrift für Kunstgeschichte* 79 (2016): 315–330.
14 Guido Alfani, Philippe Castagnetti, and Vincent Gourdon, eds., *Baptiser: Pratique sacramentelle, pratique sociale (XVIe–XXe siècles)* (Saint-Étienne: Université de Saint-Étienne, 2009).
15 Koenraad Brosens, Klara Alen, Astrid Slegten, and Fred Truyen, “Cornelia Loves Data: A Networked Approach to Early Modern Flemish Tapestry and Creative Communities” (paper presented at Methodik zwischen Theorie und Praxis, Universität Bonn, October 2–4, 2015); Koenraad Brosens, “Slow Digital Art History” (paper presented at Digital Art History, Universität Regensburg, April 7–9, 2016); Koenraad Brosens, Klara Alen, Astrid Slegten, and Fred Truyen, “Removing the Gender Bias. The Cornelia Database, Formal Art Historical Network Research and the Agency of Women in the Antwerp and Brussels Tapestry Complex 1650–1700” (paper presented at Intersectionality in Digital Humanities, University of Leuven, September, 15–17 2016).
16 Thomas P. Campbell, “New Centers of Production and the Recovery of the Netherlandish Tapestry Industry, 1600–1620,” in Thomas P. Campbell, ed., *Threads of Splendor: Tapestry in the Baroque* (New York: The Metropolitan Museum of Art, 2007), 74.
17 Koenraad Brosens, Klara Alen, and Astrid Slegten, “Claiming Commerce, Quality and Credit: *Raisons d’être* of the Antwerp and Brussels *tapissierspanden* (Sixteenth–Eighteenth Centuries),” *Textile History* 49 (2018): 5–21.
18 All her classmates quickly landed excellent and often high-paying jobs, showing the labor market’s need for professionals with hybrid digital humanities profiles.
19 The Apache HTTP Server is an open-source cross-platform web server. MySQL is a relational database management system. PHP is a general-purpose scripting language.
20 This register was published by Alexandre Pinchart, “La corporation des peintres de Bruxelles,” *Messager des sciences historiques, ou Archives des arts et de la bibliographie de Belgique* (1877): 289–331; (1878): 315–32, 475–82. For a recent analysis of this source, see Koenraad Brosens, Rudy Jos Beerens, Bruno Cardoso and Fred Truyen, “The Brussels Guild of Painters, Goldbeaters and Stained-glass Makers, 1599–1706: A Prismatic Analysis,” *Zeitschrift für Kunstgeschichte* (forthcoming).
21 This register was published by Philippe Rombouts and Théodore Van Lerius, *De Liggeren en andere historische archieven der Antwerpsche Sint Lucasgilde* (Antwerp: Martinus Nijhoff, 1872), vol. 2, 1–310.

22 https://gephi.org/.
23 E.g. http://www.sixdegreesoffrancisbacon.com/ and http://www.vondel.humanities.uva.nl/ecartico/.
24 Houda Lamqaddam, Katrien Verbert, Koenraad Brosens, and Jan Aerts, "NAHR: A Visual Representation of Social Networks as Support for Art History Research," *Proceedings of the 10th Nordic Conference on Human-Computer Interaction - NordiCHI* 18 (2018): 634–41.
25 Brosens et al., "The Brussels Guild."
26 William Shadish quoted in Beth Azar, "Finding a Solution for Missing Data: New Ways to Analyze Research to Account for Missing Data Could Significantly Change Research Results in the Next 10 Years," *Monitor on Psychology* 33 (2002): 70.
27 E.g. Gary King, James Honaker, Anne Joseph, and Kenneth Scheve, "Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation," *American Political Science Review* 95 (2001): 49–69; John W. Graham, "Missing Data Analysis: Making It Work in the Real World," *Annual Review of Psychology* 60 (2009): 549–76; Patrick E. Mcknight and Katherine M. Mcknight, *Missing Data in Psychological Science* (Oxford: Oxford University Press, 2013).
28 E.g. Joseph L. Schafer and John W. Graham, "Missing Data: Our View of the State of the Art," *Psychological Methods* 7 (2002): 147–77; James Honaker and Gary King, "What to Do about Missing Values in Time-Series Cross-Section Data," *American Journal of Political Science* 54 (2010): 561–81.
29 For the methodological complexities, see M.A. Jonker and A.W. van der Vaart, "Correcting Missing-Data Bias in Historical Demography," *Population Studies* 6 (2007): 99–113; Frans van Poppel, Dirk J. van de Kaa, and Govert E. Bijwaard, "Life Expectancy of Artists in the Low Countries from the Fifteenth to the Twentieth Century," *Population Studies* 67 (2013): 275–92; Fereshta Mirzada, Anouk S. Schimberg, Frouke M. Engelaer, Govert E. Bijwaard, David van Bodegom, Rudi G.J. Westendorp, and Frans W.A. van Poppel, "Arts and Ageing; Life Expectancy of Historical Artists in the Low Countries," *PLoS ONE* 9 (2014): e82721, https://doi.org/10.1371/journal.pone.0082721.
30 R is a free software environment for statistical computing and graphics: https://www.r-project.org/
31 Trivellore E. Raghunathan, "What Do We Do with Missing Data? Some Options for Analysis of Incomplete Data," *Annual Review of Public Health* 25 (2004): 99–117; Amanda N. Baraldi and Craig K. Enders, "An Introduction to Modern Missing Data Analyses," *Journal of School Psychology* 48 (2010): 5–37.
32 For a concise introduction to MCAR, MAR, and MNAR, see Garrett M. Fitzmaurice, Michael G. Kenward, Geert Molenberghs, Geert Verbeke, and Anastasios A. Tsiatis, eds., *Handbook of Missing Data Methodology* (London: Routledge, 2015), 3–22.
33 On MI, see Donald B. Rubin, *Multiple Imputation for Nonresponse in Surveys* (New York: Wiley, 1987); Donald B. Rubin, "Multiple Imputation After 18+ Years," *Journal of the American Statistical Association* 91 (1996): 473–89; Joe Schafer, "Multiple Imputation: A Primer," *Statistical Methods in Medical Research* 8 (1999): 3–15; Stef van Buuren, *Flexible Imputation of Missing Data* (Boca Raton: Chapman & Hall, 2012).
34 E.g. Kate Tilling, Elizabeth J. Williamson, Michael Spratt, Jonathan A.C. Sterne, and James R. Carpenter, "Appropriate Inclusion of Interactions Was Needed To Avoid Bias in Multiple Imputation," *Journal of Clinical Epidemiology* 80 (2016): 107–15.

35 As is rightfully stressed by Stephanie Porras, "Keeping Our Eyes Open: Visualizing Networks and Art History," *Artl@s Bulletin* 3 (2017): 48.

36 As all newly added data is linked to the researcher entering the data, we can easily refrain from showing it online. In addition, blog posts on the website can help to claim and secure "authorship" of data and ongoing research fueled by the data. Finally, as shown by this essay, we feel that people who are willing to share their data, energy and/or expertise have every right to be credited as co-authors of publications.