

REPORT KNAW PROJECT

Mapping Notes and Nodes in Networks

Exploring potential relationships in biographical data and cultural networks in the creative industry in Amsterdam and Rome in the Early Modern Period

Main applicant: Prof dr. Charles van den Heuvel (Huygens ING) - editor

Embedded researcher and coordinator: Leonor Álvarez Francés (Huygens ING) - text

Private partner: Lab1100

Starting date: 1st January 2014

Finishing date: 1st November 2014

Mapping Notes and Nodes in Networks

Note to the reader

This report describes the outcomes of a so-called Public-Private-Project (PPP) of the Royal Netherlands Academy of Arts and Sciences (KNAW) that makes part of a joint venture of the University of Amsterdam (UvA), Vrije University Amsterdam (VU) and the KNAW to support research in the digital humanities that focuses on collaboration between research institutions, non-profit organizations (e.g. cultural heritage institutions) and private companies in the creative industry to develop innovative digital research methods and new modes of valorisation of humanities knowledge.

The *KNAW-Public Private Project: Mapping Notes and Nodes* was a collaboration between the Huygens Institute for the History of the Netherlands (Huygens ING) and a software development company for research in the humanities Lab1100 (Pim van Bree, Geert Kessels).

Purpose of the collaboration was to develop a tool within Lab1100's platform Nodegoat to visualize the data imported from three different databases in order to establish priorities for data integration in the humanities. The embedded researcher and coordinator Leonor Álvarez Francés, together with Pim van Bree and Geert Kessels (Lab1100) experimented two days per week for a period of nine months with data import and data integration of heterogeneous datasets. This process was carried out to assess which data was relevant bringing together in the context of the history of the cultural industry. In the course of research more datasets than the originally planned were added to the project, adding more perspectives and dimensions to the analysis.

The case study for this project is the presence of networks of artists and intellectuals linking the creative industries of Rome and Amsterdam during the Early Modern period. The existence of several datasets related to this subject made for a perfect case to reflect on data integration within the humanities and the possibility this process offers for asking new questions and opening up new perspectives.

The creators and/or managers of the different datasets were part of the research committee and searched for collaboration in future projects, thus positively contributing to both the technical and content-related aspects and issues present in *Mapping Notes and Nodes*. Expert meetings were held every month counting on the fixed attendance of the main applicant Charles van den Heuvel (Huygens ING), the embedded researcher and coordinator Leonor Álvarez Francés, the software engineers of Nodegoat Pim van Bree and Geert Kessels, the creators of Ecartico Marten Jan Bok (UvA) and Harm Nijboer (UvA), one of the managers of Hadrianus, Marieke van den Doel (KNIR), the Head of Digital Data Management of the Huygens Instituut Sebastiaan Derks, and the researcher for Digital Reference Works Ronald Sluijter (Huygens ING). Finally, in the second half of the project, two Master interns joined the project and the meetings, namely Ingeborg van Vugt and Simone Wegman, both of Leiden University.

After presenting the original outline for the project, this report focuses on the methodology employed, the criteria for selecting cases and the analysis thereof, the conclusions of the research, and a reflection on possible further steps to be followed toward a better assessment of digital integration in the humanities.

The visualizations developed for the cases and the integrated data can be viewed and explored following this link: <http://mnn.nodegoat.net/viewer>. The database can also be accessed through the site: <http://nodegoat.net/>, logging in as follows:

Username: demo_mnn

Password: demo

We wish to thank the KNAW for funding and are indebted to our advisors and those scholars who greatly increased the reach and valorisation of this project by means of providing extra datasets.¹

Leonor Álvarez Francés MA
Huygens ING
Embedded researcher and coordinator (Text)

Prof. dr. Charles van den Heuvel
Huygens ING
Main applicant (Ed.)

¹ We are most thankful to the data and expertise of dr. Marten-Jan Bok and dr. Harm Nijboer (University of Amsterdam-Ecartico), dr. Marieke van den Doel and dr. Arthur Westeijn (KNIR-Hadrianus), dr. Susanna de Beer (Leiden University), Hans Brandhorst and Etienne Postumus (Mnemosyne-Iconclass), dr. Arjan de Koomen (UvA) and dr. Frits Scholten (Rijksmuseum; VU Amsterdam) for *Sculptors on the Move*, MA Sebastiaan Derks and dr. Ronald Sluijter (Huygens ING-Biographical Reference Works), dr. Alette Fleischer (Corvinus and Bentvueghels), and MA Erin Downey (Temple University Philadelphia: Patronage Guistiniani and Bentveughels).

Table of contents

1. Project description and relation to the digital humanities	5
2. Sources and approach	5
3. Method	6
a. Cases	7
b. Additional data	9
c. Technical aspects of data integration	10
i. Obstacles during data integration	10
4. Analysis	11
a. Research topics chosen for analysis	11
5. Further research	21
a. Content related additions	21
b. Technical developments	22
i. For data integration	22
ii. For data analysis	23
6. Conclusion	25

1. Project description and relation to the digital humanities

Over the past decades, numerous researchers working either within joint projects or individually have developed databases to store data that was valuable and necessary for their specific purposes. While these initiatives have brought much knowledge and experience to the humanities as a field, their potential can be augmented and reinvented by integrating them with other databases. *Mapping Notes and Nodes* aims at assessing not only how diverging databases can be brought together in order to be reused and answer new research questions, but mostly at studying in how far and under which conditions this data integration process is a fruitful enterprise.

Complete data integration consumes much work, time, and therefore resources. Our goal is to combine historical actors (nodes in networks) and contextual information (notes in networks) and map them over time and place in order to answer our main research question: what information is relevant to bring together? In order to do so, it was projected to develop a viewer that allows for visualizing the co-presence of notes and nodes of the creative industry in Amsterdam and Rome in the Early Modern period. The links binding these cities were manifold and diverse, ranging from connections via artists like Maarten van Heemskerck, engineers like Cornelis Meijer, or botanists like Hendrik de Raaf.

The availability of complementary but heterogeneous databases covering this period and these networks make it the perfect case study for this project. The fact they are complementary makes it plausible that their combination will render overlap. Their heterogeneousness allows us to experiment with the difficulties of integrating datasets that differ both in approach and structure. Indeed, since platforms in the Digital Humanities have been developed for specific datasets in the past, forthcoming integration processes will face these challenges. At the same time, these difficulties explain our ambition to develop a multi-applicable tool that serves for diverging historical enterprises and fields; the viewer above described being an instrument within the tool.

The study of the creative industries of the Dutch Republic with digital tools is the focus of scholars and research teams in many universities and organisations in the Netherlands and abroad. *Mapping Notes and Nodes* aims at contributing to this field by serving as a guideline for data integration and offering a generic tool for data analysis.

2. Sources and approach

The project started by bringing together three databases, namely The Biografisch Portaal (HI-KNAW), Ecartico (UvA) and Hadrianus (KNIR). The Biografisch Portaal contains biographies of over 80.000 people related to Dutch history. Ecartico is a database storing information about people involved in cultural industries primarily in the Dutch Republic and with a focus on Amsterdam. The project is led by dr. Harm den Boer and dr. Marten Jan Bok, both researchers at the University of Amsterdam. Finally, Hadrianus aims at gathering data on the Dutch presence in Rome and has recently been launched by the Royal Netherlands Institute in Rome. Although built up with different purposes and research agendas, these databases are complementary for the study of Dutch artists in Early Modern Rome, making them ideal object of study for *Mapping Notes and Nodes*. In the course of the project more datasets became available, contributing new dimensions to the project as a whole.

While the above mentioned databases strive for completeness, the datasets that became part of *Mapping Notes and Nodes* at a later stage were remarkably partial. This was mainly the result of pragmatic decisions such as the lack of time to prepare to curate the datasets for analysis. However, this circumstance did not make them less suitable for our goals, considering the focus of the project on the methodological implications of incremental data integration from multiple perspectives in the humanities, and not on an exhaustive study of the artistic networks linking Amsterdam and Rome. These networks are thus merely the case study serving the main purpose of achieving general conclusions for the digital humanities as a whole.

From a technical point of view, the platform Nodegoat was the main tool for our experiments. This web-based database management platform, developed by the private partner in the *Mapping Notes and Nodes* project, Lab1100, allows for incremental data integration and analysis and visualization of networks simultaneously. Lab1100 aims at constantly improving Nodegoat, and therefore seeks for close cooperation with humanities researchers that provide the necessary feedback for designing additions and correcting and refining the tool. Close collaboration between the embedded researcher and the developers within Lab1100 guaranteed avoiding technical issues to interrupt research, and fostered improvements within Nodegoat that contribute to a deeper analysis of the results of the data integration process.

Moreover, the embedded researcher was the direct contact person that could bring in the suggestions for specific improvements of the software or for additional features that were assembled from the monthly meetings and from the collaboration with the interns at the Huygens ING who were working on specific cases.

3. Method

With our main goals in mind, namely producing a joint visualization of notes and nodes and developing a generic tool for the digital humanities, the project was divided in three phases. Firstly, it was necessary to search for case studies providing the most varied range of perspectives possible. In order to make the conclusions of the present study applicable to most humanities projects, it was interesting to find data that is representative for the diverging approaches to be found in the humanities. Secondly, data integration took place within Nodegoat in what was the most technical phase of the project. Lastly, the resulting database was analysed and conclusions could be formulated.

During the first phase several case studies were discussed and selected to be part of the experiment or discarded as not adequate. The criteria for this selection were the following:

- Availability of digital (meta-)data regarding the relations relevant to the history of creative industry between Amsterdam and Rome.
- Minimizing manual work, albeit that many metadata has been added manually by the embedded researcher and interns.
- Instead of completeness within networks of actors and documents related to this history, we opted for adding more networks to explore the potential of combining multiple dimensions.
- Similarities in metadata with the large datasets of Ecartico and Hadrianus.

- Digital enrichment of existing analogue sources, such as the addresses of the Dutch artistic community in Rome, the Bentvueghels, a case study explained in the coming paragraphs.

a. Cases

The following cases were chosen to be part of the project:

- **Bentvueghels:** a society of Dutch artists who resided in Rome approximately between 1620 and 1720. This case has much to offer to *Mapping Notes and Nodes* because of its potential for overlapping with the other databases and because it offers the perspective of a group. The members of this society were a link between the Netherlands and Rome themselves, establishing and sharing artistic relations in both geographic areas and thus together holding an extensive network.

Biographical data on these artists (172 in total) had already been stored in Hadrianus. The data is based on the work of the Dutch historian Hoogewerff, who in 1938 and 1939 published several inventories and summaries about the presence of Dutch artists and scientists in Rome and indicated the locations of their residences. Several of this society's members can also be found within Ecartico and Biographical Reference Works, making it a very suitable case for testing the results of data integration of all three databases. A noteworthy difference between Ecartico and the other two main databases is its geographic component. Ecartico has stored address locations when these were available, an element missing from Hadrianus and Biographical Reference Works. During the first meeting the idea raised as to contact the KNIR and discuss the inclusion of the Bentvueghels's addresses in Rome. The researcher Hoogewerff inventoried the artists' places of residence, to be found in digitised form in the Digital Reference Works of the Huygens ING and which the KNIR agreed to store within Hadrianus and provide them with geographic coordinates.

- **Cornelis Meyer.** Meyer (1629-1701) was an engineer who had contacts with intellectuals, painters and popes both in Amsterdam and Rome. He was a member of the *Accademia Fisico-matematica* in Rome. Despite the variety of spheres present within his network that make him an interesting case, the references to his life in all three databases are scarce, making manual data introduction the only option. Accordingly, this case was assigned to an intern, Ingeborg van Vugt, who joined the project in June 1st.

- **Cosimo III de Medici:** A rich amount of correspondence from and to Cosimo III –often through his secretary Apollonio Bassetti- has arrived to our days. These letters contain abundant references to other people and objects, therefore revealing networks. In total, 937 letters were added to the project. The correspondence, introduced by Ingeborg van Vugt, was selected because of its completeness and due to the important role that those involved played within Cosimo III's network. These individuals were, apart from Cornelis Meyer, Giovachinno Guasconi (1636-1699), Pieter Blaeu (1636-1707), Daniel Nicolaas Heinsius (1620-1681), Pietro Guerrini (1651-1716), and Cassiano dal Pozzo (1588-1657).

The following cases were discarded:

- **Artus Quellinus:** this sculptor is mostly known for his work on Amsterdam's new town hall, today known as the Royal Palace. In opposition to the Bentvueghels, he offers the perspective of an individual. Because he was a sculptor, his inclusion can add plenty of new connections to the painters' networks. Besides, Quellinus had a broad international network. He can be found in Ecartico and Biographical Reference Works but not in Hadrianus, where two nephews of him are stored as Bentvueghels, which is an indication of the potential of this case for linking diverging datasets. Unfortunately, and since his records in Ecartico and Biographical Reference Works are predominantly short text documents instead of datasets, the data required manual import.

- **Maarten van Heemskerck:** this sixteenth-century painter is regarded as a key figure for Dutch-Roman contacts in the Early Modern period. Indeed, more than forty nodes of his network can be found in Hadrianus and Biographical Reference Works. An important inconvenience for incorporating this case to the project is the lack of temporal overlapping with the other cases, which cover predominantly in the seventeenth century. Besides, the data for this case would have had to be added manually.

- **Farnese family:** the Farnese were renowned patrons of the arts, therefore adding to *Mapping Notes and Nodes* the perspective of a group of Maecenas, which means networks of artists but also of members of the social elite. Moreover, the PhD candidate and co-applicant to this project Sebastiaan Derks studies the dynasty policy of the Farnese family of Margaret of Austria and could therefore offer his expertise to the project. Unfortunately, the Farneses are not registered in Ecartico or Hadrianus, and the data available through Biographical Reference Works would have to be searched for in the text and manually introduced into Nodegoat.

- **Karel van Mander:** his *Schilder-boeck* (1604), digitally available through DBNL, contains biographical information on plenty of contemporary painters from the Low Countries. This would make an excellent case to include text-mining methods in the project. However, the version digitised in DBNL is the original one, written in Old Dutch. As Piek Vossen pointed out the data curation for exploring and developing this set would cost too much time, making it inviable within this project. Although the part of the *Schilder-boeck* dealing with painters from the Low Countries has been translated into English, no digital version of this text is available anymore. Neither Hessel Miedema (the author), nor Michael Hoyle (the translator), nor Joop van Coevorden (the editor) has preserved a digital copy. For these reasons, van Mander was discarded as a case for this project.

- **Galleria Giustiniana:** an art collection (or a set of objects artistic or not) is often the subject matter of research in the humanities. Vincenzo Giustiniani's collection is therefore an interesting new perspective to add to *Mapping Notes and Nodes*. PhD candidate Erin Downey (Temple University), whose thesis focuses on the networks of the Bentvueghels, offered her database around the Giustiniana to the project. Unfortunately, and due to the fact she was working on the final version of her Doctoral thesis, we did not receive data on this case.

In the second phase, all three databases together with smaller datasets that were made available by several researchers were integrated within Nodegoat. The decision was made as to firstly import Ecartico, therefore taking its structure as the base for the integration of all other datasets. There were two reasons for this; Ecartico was not only the biggest database of all with around 20.000

entries, but was as well the best structured one. Concerning Hadrianus, Lab1100 received a dump of all data within the database, thus not only data on the Bentvueghels. The team of *Mapping Notes and Nodes* decided to import the whole of the database because this can only provide more links and does not demand more time to be spent on the import process. Given that Biographical Reference Works contains more than 80.000 entries, only a small selection was to be imported, namely around 800 entries that offered potential overlapping with the other datasets due to a similar temporal and geographic range. More specifically, we filtered by date of birth (between 1450 and 1700), by date of death (between 1500 and 1750) and selected those biographies including the word 'Rome'. This facilitated overlapping but left the link with Rome very open, for these people might have been themselves in Rome or not.

b. Additional data

During the course of the project several people wanted to participate and bring in their own collections. In that sense the project was already valorised before it finished. The consequence of these added metadata was that they sometimes went beyond the initial focus on Amsterdam-Rome relationship. We decided to add them since we are interested in partial overlapping of networks and in the question which part would be most relevant for further digitization and exploration.

The following datasets were added to the original chosen sets of the proposed research project:

- **Dutch sculptors on the move:** this database stores around 200 sculptors and their trips around Europe between the fifteenth and the seventeenth centuries. It was created by Frits Scholten (Rijksmuseum) and Arjan de Koomen (UvA) together with several students of theirs. Because the students were left to choose on a specific group of sculptors, such as those travelling to Spain, this dataset is very selective and involved more geographic areas than the Low Countries and Rome. However, as it has already been stated, completeness is not a requisite for a dataset to be part of *Mapping Notes and Nodes*.

- **Representations of Antique Rome:** Susanna de Beer (UL) has received an individual fellowship at NIAS for the project 'Mapping Humanist Visions of Rome: Sharing and Visualizing Poetic Appropriations of the Roman Physical and Literary Heritage'. This research project will combine the finishing of her VENI research with a spin-off of that same project in the field of computational humanities. At an early stage we explored with her a possible inclusion of her data in our project. She started experimenting with Nodegoat and together with Lab1100, Hans Brandhorst (Iconclass) and Etienne Posthumus (Iconclass/Arkyves) we explored the possibility of including Iconclass classifications in de Beer's storing process in Nodegoat.

- **Corvinus:** Arjan de Koomen brought us in touch with Alette Fleischer, who told us about Hendrik de Raaf (alias Corvinus), a Dutch apothecary who lived in Rome and had contact with artists such as the Bentvueghels. Fleischer has proof of this network and supposes artists bought their pigments from him. Although she did not have a database of Corvinus' network, the intern Simone Wegman showed great interest in this case and carried out the research within Nodegoat.

c. Technical aspects of data integration

Two specific technical deliverables were foreseen, namely:

- 1- A multi-faceted import module aimed at integrating heterogeneous data sets while preserving references to the source databases. This module allows for a semi-automated process and enables for assisted verification, thus fostering a time-saving but accurate import process in which all spatial and chronological attributes of the objects are saved.
- 2- A visualization module where nodes can be related to each other spatially and chronologically while preserving references to the source database. This viewer can be accessed here: <http://mnn.nodegoat.net/viewer>.

Both the import module and the viewer were developed during the project and are fully generic. These instruments would, according to the project description, form together a generic tool (deliverable 2). However, Lab1100's software Nodegoat showed to be an ideal platform where these modules could be built upon. The research team chose for this strategy, which allowed for further developing of an already existing functional and generic tool without having an effect on the project's goals or promises.

Within the viewer a scenario tool was created, where filters and scopes could be saved. In practice, this means the different cases we analysed can be viewed and explored online and in open access. Its public is as broad as wished thanks to the possibility to include descriptions of the visualizations on the right side. This site of the viewer is in fact a publication of our research results.

These instruments made it possible to employ the data within the project to the full of their capacity, but have broader repercussions for the digital humanities as a whole. Their flexible, generic character implies that these tools can serve for integration and visualization of heterogeneous data sets stemming from any historical discipline and for unlimited research purposes.

i. Obstacles during data integration

While the data integration process of the different databases was a success, it did not happen without obstacles. In what follows the main difficulties are presented.

- **Matching took place manually for a big part.** A crucial aspect of this research, as described above, is the overlap of data, which was mainly expected to happen through the presence of the same people in different databases. The matching and merging of the same persons from different databases met a number of obstacles. This data curation process turned out to be very time consuming.

First of all, Nodegoat finds exclusively perfect matches, that is, only suggests a match when the name is rendered in exactly the same spelling. Secondly, recognition by name in the Early Modern period is tricky business. To begin with, there was less diversity concerning names in comparison to the present times. This, together with the scarcity of biographical data in some cases, makes it difficult to conclude whether, for instance, two 'Jan Jacobsz.' should be merged. When there were indications of a match but also uncertainty, two entries were linked by the new field 'Possible identity'. Besides,

many Renaissance people signed under a Latin version of their name, such as Jan van der Straet (*Johannes Stradanus*), while some acquired another name upon arrival in another country (Caspar van Wittel was called *Gaspare Vanvitelli*). Van Wittel also had not one, but several aliases, namely *Gasparo degli Occhiali*, *Piktoors van Amersfoort* and *Toorts van Amersfoort*. In our case, which is no exception, the databases to be integrated were built following diverging criteria. While Ecartico often gives the name as recorded in official acts (baptism, marriage e.g.), Hadrianus tends to register the Italian variant. The fact Ecartico also stores alternative names only partly solved this hindrance. Thirdly, and as a consequence of the first two obstacles, a considerable part of the merging process took place manually. This means mostly an extra investment of time, but likely leads to some missed matches as well.

- **Different standards for geographic locations.** While Nodegoat takes its geographic data from the database geonames.org, most databases assign a set of coordinates to a location by means of manually pinpointing it on Google maps and copying this data into the database. Consequently, Paris in Ecartico will be located (close to but nonetheless) somewhere else than Paris in Sculptors on the Move. Within this project it was not possible to curate all these data by means of assigning a geonames code to all geographic locations. During the import of geographic locations, matches were sought automatically and merged manually, thus bringing 'Paris' under geonames coordinates and code. Because Nodegoat did not import places with less than 5000 inhabitants from geonames, small populations from the imported databases retained the coordinates they had originally and did not receive a geonames code.

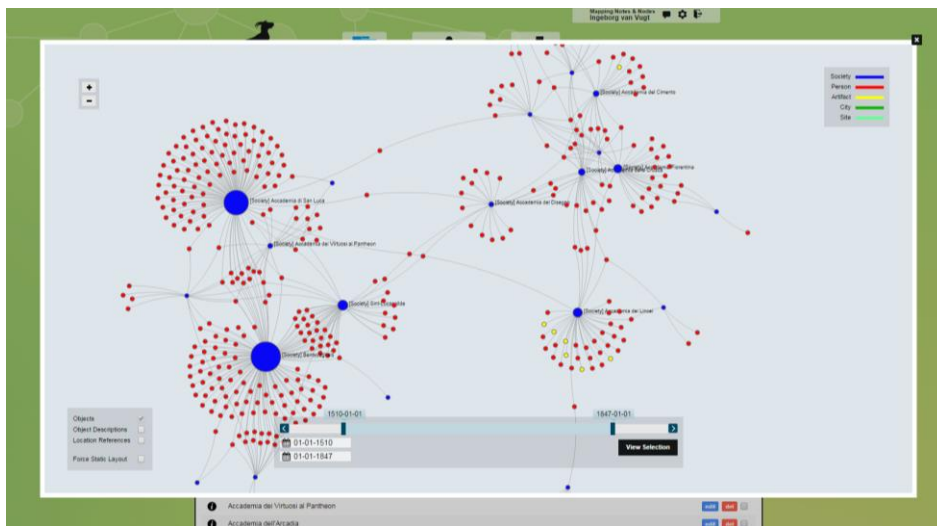
- **Historical locations.** Ecartico has harvested the addresses of plenty of artists, a very rich dataset for studying settling patterns in the creative industry through time. Some of the street names have changed and need of manual indication of coordinates until historical maps of Amsterdam and other cities containing coordinates are designed. Concerning the names that have stayed the same, although recognised automatically and thus mapped, they are not represented as precisely as desired. These addresses often contain indications as to the precise place in the street with denominations as '*Warmoesstraat in de Drie Coppen*', or even giving a house number, such as in '*Warmoesstraat nr. 178 in de Gouden Leeuw*'. Both these locations were mapped at the same spot, meaning the visualisation does not represent the data available in all its richness.

4. Analysis and technical developments

a. Research topics chosen for analysis

Once the three originally planned databases Ecartico, Hadrianus and Biografisch Portaal plus the additional data were imported into Nodegoat, it was possible to proceed to an analysis of the added value of data integration. The last three meetings were devoted to discuss what strategies were optimal for answering this question. The varied expertise present during the meetings made it possible to address our database, now containing 22.460 people, from different points of view. The following research scenarios were designed to explore the numerous possibilities our final database offers for historical enquiry. Some are derived from the previously selected cases, whilst other arose from experimenting with the data once it had been brought together.

- **Network topology: Meijer, Cosimo III and academies.** A dimension we cannot neglect is the analysis of the network as a whole, the so-called topology of the network - both because of its importance within the humanities and because our data lends itself perfectly for it. In particular we were interested in various clustering patterns of societies in hubs and in intermediary persons with few connections but linking various communities –weak ties- who were important relevant to transfer knowledge from one network to another. This aspect was studied combining data from the cases of Cornelis Meijer and Cosimo III de Medici to do research on the intellectual and artistic organisations par excellence in early modern Italy: the *Accademie*. These institutions gathered those interested in a particular field and were hubs for dissemination and promotion of culture. A visualization of networks around them shows they were rather isolated groups.

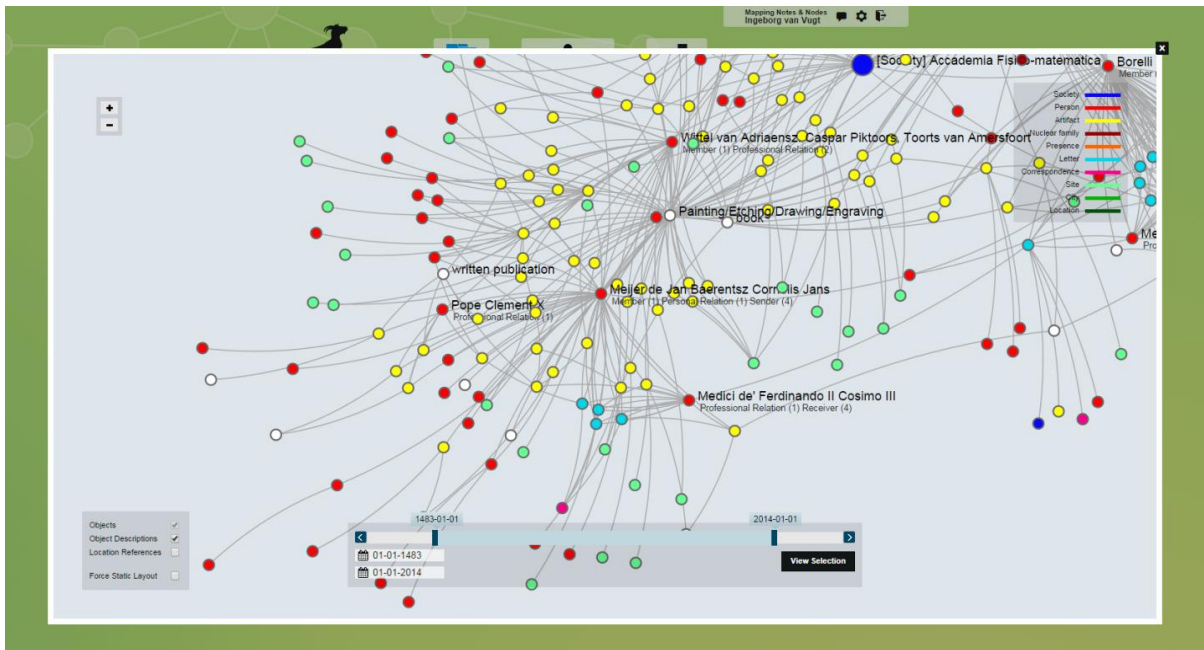


Visualization of the networks of *Accademie* in Rome and Florence showing central hubs and weak links.

The *Accademie* are the blue dots; their members are represented in red.

Despite their evident tendency to be secluded groups, these *Accademie* were connected thanks to weak links, people that functioned as bridges between them. An example of a weak link is Cornelis Meijer. Although not a weak link in the most strict sense of the term because he was acquainted with more than one member of the same society, he was positioned between popes, statesmen, intellectuals and artists. Nowadays we would not reckon his engineering work to the creative industry, but he commissioned various Dutch and Flemish artists in Rome to make drawings and engravings after his designs.

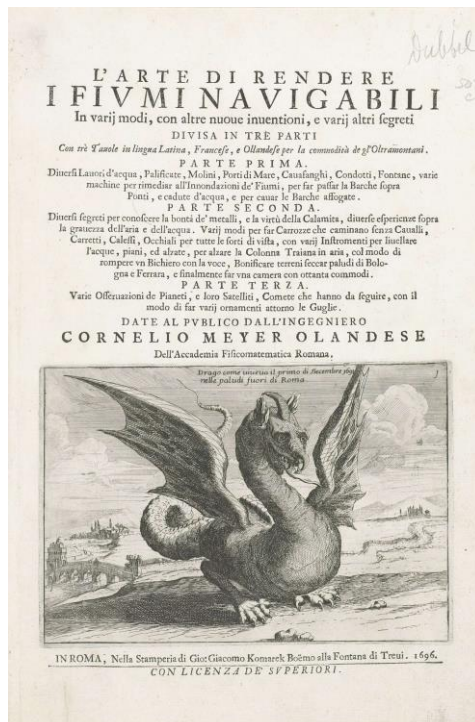
Data from the correspondence of Cosimo III included references to objects, such as books, that circulated throughout networks and reveal new nodes. When visualizing Meijer's technical publications within this correspondence including dedicatees, engravers, printers and letters mentioning Meijer's work, new nodes of his networks are shown.



Correspondence of Cosimo III showing references to Meijer's work.

Meijer is at the centre. Red dots are people; yellow dots are artefacts (e.g. publications).

New insights concerning Meijer's work were possible because Ecartico, Hadrianus and Cosimo III's correspondence were brought together. For instance, the engraving 'Dragon in the Marshes' in Meijer's '*L'arte di rendere i fiumi navigabili (...)*' (1696) was enclosed in a letter from his hand to Cosimo III in 1692.



The engraving in the book includes a text saying the dragon was found dead in 1691. The letter tells the story how the dragon was found dead in much more detail, including how the innards had to be removed due to the stink they produced. Information on the publication comes from Hadrianus, while the letter was contributed by the correspondence of Cosimo III, retrieved from the Florence State Archive. Both Ecartico and Hadrianus complete the networks around this link with prosopographical information. The publication and the letter were brought together and contextualized by means of data integration, a connection that had led to new theories. A possibility has been suggested by Ingeborg van Vugt that Meijer sent this engraving to Cosimo III due to the Grand Duke's fondness of natural oddities as a manner of searching for his patronage in the future.

- **Iconclass codes linking different datasets.** When all datasets had been combined, it became clear how the Representations of Antique Rome could play an interesting role within *Mapping Notes and Nodes*. Susanna de Beer's project focuses on images of Rome in Humanist Latin poetry where specific locations in Rome are named, especially Antique monuments. These monuments are in turn pinpointed within the Hadrianus database as represented locations in paintings. Iconclass codes include references to concrete geographic locations (such as the Forum Romanum) but also to motifs such as 'city in ruins'. This offers the possibility of exploring the representation of Antique monuments in Renaissance poetry and paintings. The language independent, numerical and letter codes of Iconclass make it possible to link descriptions of monuments to depictions of monuments without problems of semantic differences in metadata.

The Italian humanist Paolo Spinoso related in a poem the transfer of the sarcophagus of Santa Constanza from de Sant' Agnese fuori le Mura to Palazzo Venezia. During the transfer, the sarcophagus mentions the Coliseum, which brings us for example to an etching of the same monument by Hieronymus Cock (1551). Within Iconclass, this image is classified as 25I9 (Landscape with ruins) and 45L43 (devastated, ruined place or city (warfare)), as is expressed in the Latin title: *Colossaei Romani a Barbaris diruti prospectus* (View on the Roman Colosseum, destroyed by Barbarians). These codes are attached as well to Spinoso's poem, where the Colosseum is in fact described as *semirutum* (half destroyed) and the ruinous landscape of the city is also ascribed to the various barbarian invasions Rome suffered in the past. Apart from tracing cross-references throughout the database, the similarity between these two images could also be established by filtering all artifacts in the database on the Iconclass codes for Colosseum (61F(COLOSSEUM)) and 'Devastated, ruined place or city' (45L43).

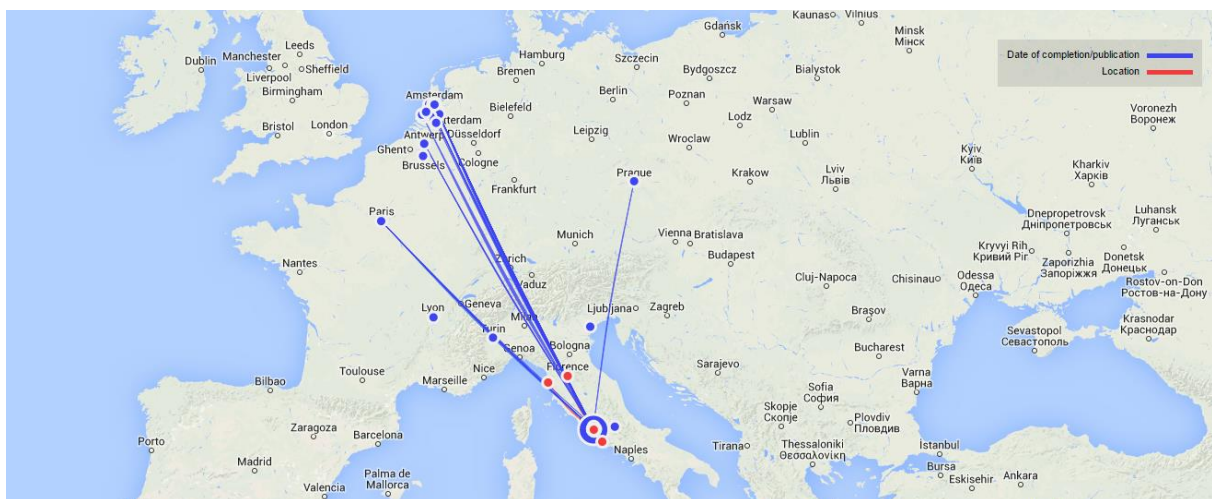
Iconclass code 25I9 (Landscape with ruins) also offers a link to the painting 'Landscape with Roman ruins' by Hermannus Postumus, that offers a kind of phantasy ruinous landscape compiled from real sites. Just as interesting is the Latin quotation *Tempus edax rerum tuque invidiosa vetustas / Omnia destruitis* (You, devouring time and jealous old age, destroy everything), which explains the causes of the Roman ruins. This topic was very common among the humanists writing about Rome and tracing this in the database via the Iconclass code 23A441 (Time causing the ruin of buildings, etc.) for example brings us to the Latin poem by Joachim Du Bellay. This poem also offers a very lively overview of the most famous ancient sculptures in the Musei Capitolini and Musei Vaticani. Finally,

the same Iconclass code also brings us to the source of the quotation on Postumus' painting: Ovid's *Metamorphoses* 15.

Nodegoat also offers the possibility to visualize the route that the sarcophagus follows in Spinoso's poem, in which Santa Constanza laments the ruinous state of the city she has not seen since she was buried in the 4th century AD. From there she is transferred to the Palazzo Venezia, commissioned by Pope Paul II and at the moment of writing still under construction. By this same means also much longer 'literary walks' through the city can be visualized. This can, for example, be helpful in establishing the relationship between 'real' Rome and 'literary' Rome in passages that were modeled after the walk over the future site of Rome by Aeneas in *Aeneid* 8.

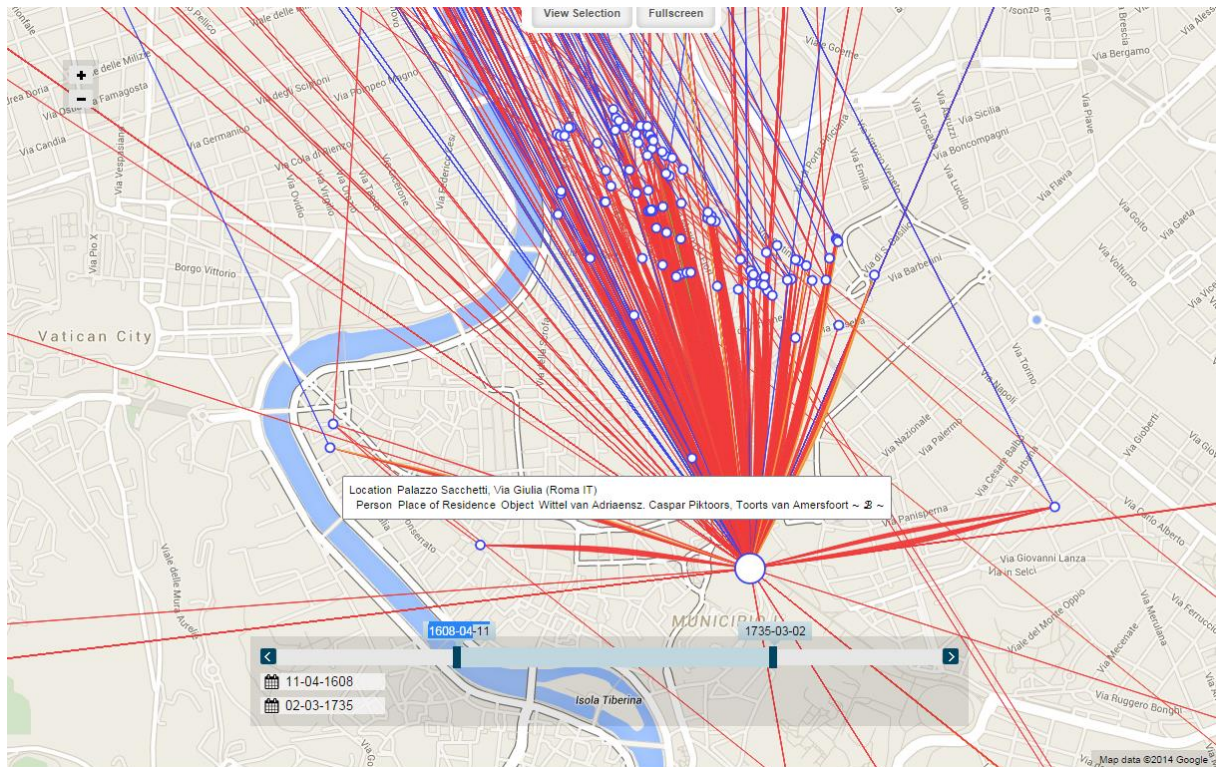
- **Connections between represented objects and the artist's location: Casper van Wittel.** Thanks to data on migratory movements and artefacts that we gathered from different databases, it is possible to establish whether an artist was in fact in Rome when he painted the Coliseum. Indeed, while *Sculptors on the Move* and *Ecartico* store rich data on residential locations, *Hadrianus* is missing this type of information, but contains a collection of paintings where places (for a great majority in Rome) are depicted. This makes it possible to assess whether works of art were composed *ad vivo* or at the studio, a question especially interesting when dealing with landscapes, cityscapes and monuments.

It is important to take into account that artists very rarely painted outside because paint in tubes did not exist yet, and that many artists travelled to make sketches of objects that would serve as models at the studio.



Many Italian locations (red dots) were depicted somewhere else (blue dots)

These circumstances explain why the visualisation shows that most representations were painted in a different geographic location and make exceptions (when red and blue dots are placed together) utterly interesting. An example is Casper van Wittel, a painter from Amersfoort who lived in several addresses in Rome in the seventeenth and eighteenth centuries. Van Wittel was member of the Bentvueghels, but his settlement in Rome greatly differs from that of the other Bentvueghels.



Overview of all places of residence of the Bentvueghels available in the Mapping Nodes and Notes dataset, showing the long century in which the Bentvueghel artists were active in Rome (1620-1720). The blue dots south-west from the rest pertained to Casper van Wittel.

Thanks to data integration we can identify Van Wittel as an exception in his time. He was an artist painting monuments as vivo and displays very different residing patterns than his fellow Bentvueghels. Indeed, he stands out because he was successful in working for the Roman art market and finally settled in Rome for the rest of his career. This gave him the chance to portray the Eternal City while living in it and to reside in different parts of the city where he might be more integrated with the local population.

- Residential patterns in Rome in relation to origin. This is a case inspired by the data gathered by the Hadrianus team about the addresses of the Bentvueghels in Rome.

The Bentvueghels settled within a specific area in the city, just south of the Piazza del Popolo, near the northern city gates. This pattern triggered the question: why did they move into the same space in Rome? Did they know each other before arrival? While these data are based on the partial research of Hoogewerff and thus cannot be used for general statements on Netherlandish artists in Rome, the phenomenon of members of the Bentvueghels living in the same streets and even sharing houses is an interesting fact for further research.

In order to answer these questions, the addresses in Rome were plotted along with addresses of the Bentvueghels in Amsterdam (Ecartico), migratory flows (Sculptors on the Move) and data from the additional datasets provided by Ingeborg van Vugt, Susanna de Beer and Simone Wegman. Undoubtedly, clustering might have played a determinant role in these patterns. This neighbourhood

was possibly known as the artists' area, or even the Netherlandish artists' area. Other possible explanations are that this was a cheap neighbourhood, or a neighbourhood recipient of many immigrants.

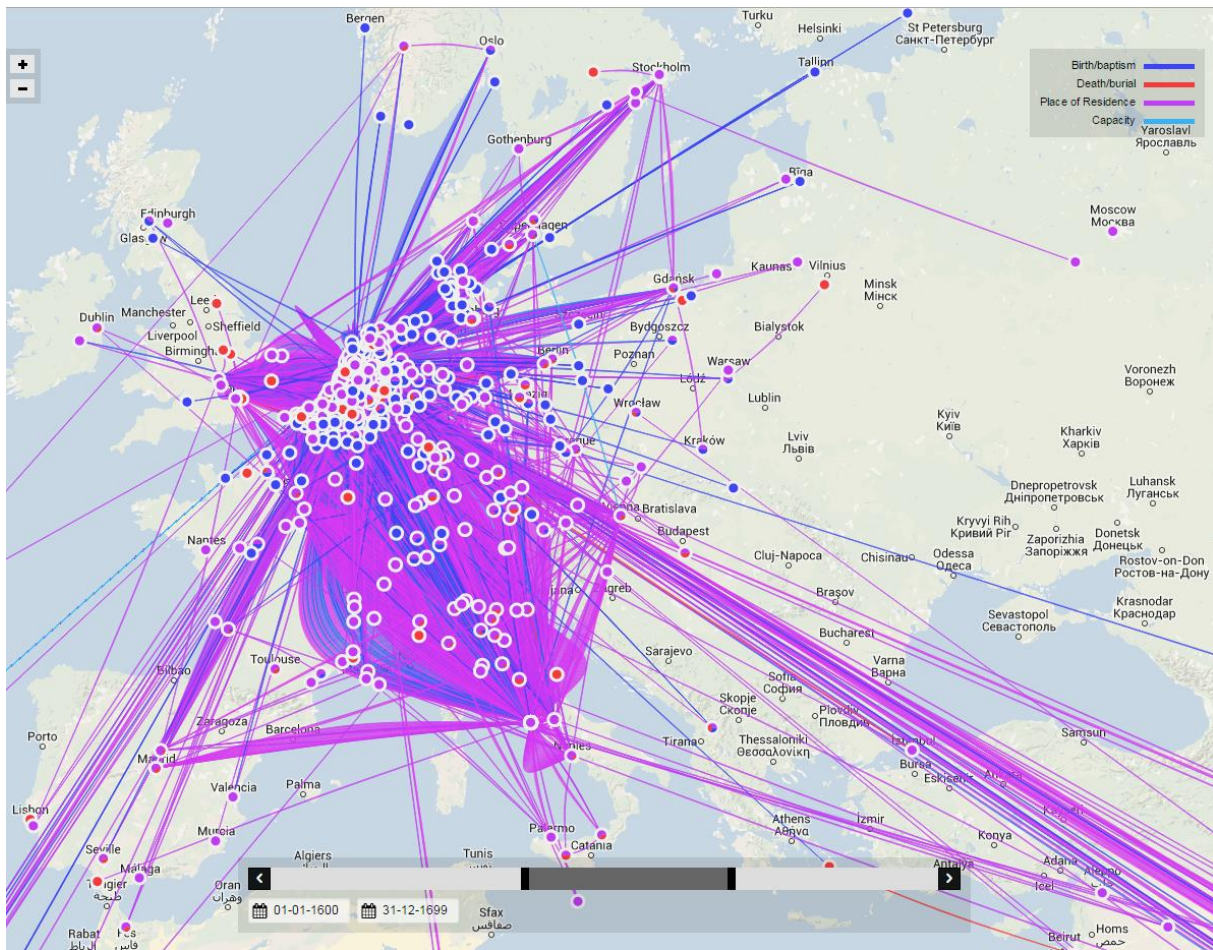
However, evidence shows that networks played a role in these residing patterns as well. An interesting case is that of the three founders of the Bentvueghels, Wouter Crabeth, Wybrand de Geest and Leonard Bramer. All three of them were in Aix-en-Provence in 1616, where we know for a fact that de Geest and Bramer met. Crabeth and Bramer lived in the same house in Rome in Via della Croce (1618) and later on in Via del Pontifici (1620-1621). This points out that not only residential patterns in Rome can be related to contacts before arriving in Rome, but that the creation of the Bentvueghels itself might be founded on networks existing before in other geographic areas.

While comparison of addresses in Rome and Amsterdam did not render significant results, there are examples linking geographic origin to closeness in Rome. Bramer lived in the same house in the Via Bocca de Leone with another Bentvueghel-to-be, namely Pieter Anthonisz. Van Groenwegen already in 1615-1616, thus many years before the brotherhood came into being (around 1620). Both were born in Delft in around the same period (Bramer in 1596, Groenwegen in 1599). Philips le Petit, Bartholomeus Appelman and Willem Doudijns, all from The Hague, lived together in the Strada Laurina, 8, between 1656 and 1658.

This case combined information from Ecartico (origins, travels) with residential and network information from Hadrianus.

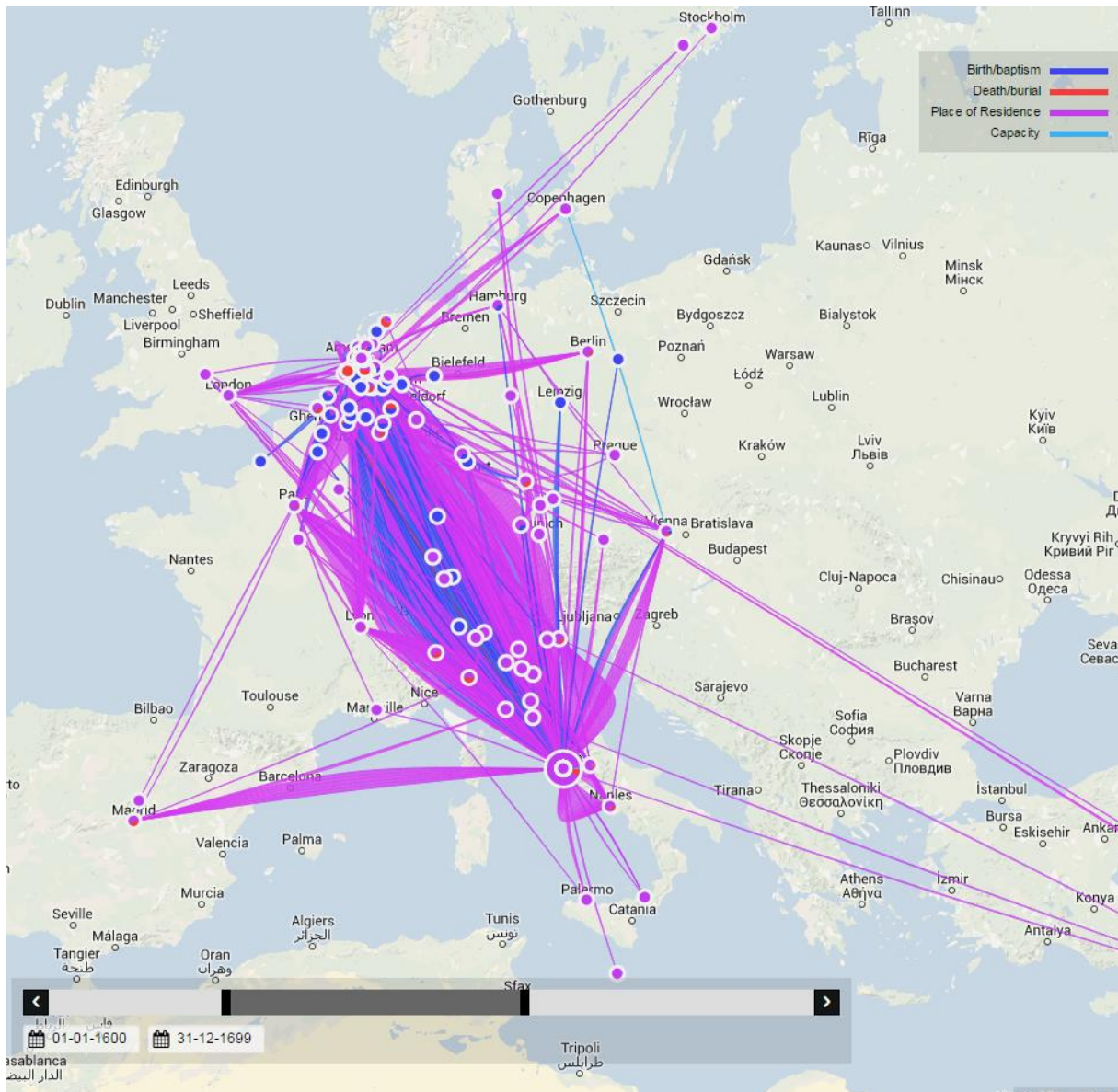
- **Migratory patterns.** The great amount of information on migration that Sculptors on the move, Hadrianus and Ecartico contributed stimulated the idea of exploring the migratory patterns shown by artists in the seventeenth century, the period they all contain data about. While the Mapping Notes and Nodes database is partial, it is interesting to explore how employing different datasets leads to diverging answers to the same question: where did artists travel the most in the seventeenth century?

This first visualization employed data from Ecartico and shows much traffic toward North Germany, followed by the Northern areas of Italy and Central Europe. France, Scandinavia and England were also popular destinations but to a lesser extent.



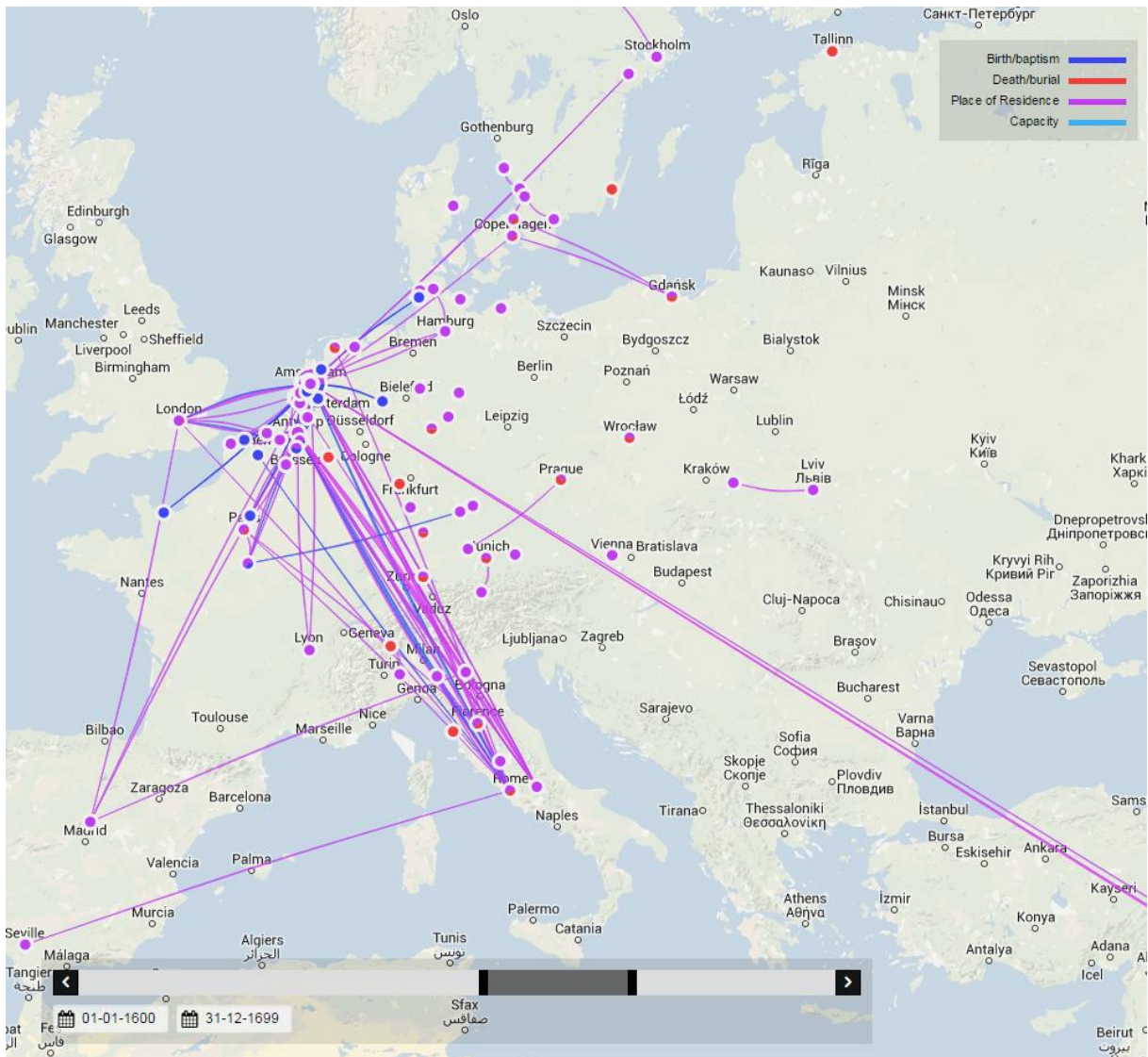
Migratory patterns of artists present in the Ecartico database (1600-1699)

For the second exploration only the Bentvueghels were mapped. They showed an intense affluence toward Northern areas of Italy as well, but the North of Germany is for them a minor destination. Central Europe, France, Scandinavia and England enjoy the same popularity for the Bentvueghels as for the artists stored in Ecartico.



Migratory patterns of the Bentvueghels (1600-1699)

Lastly, the Sculptors on the move were tested for their migratory patterns and the result is that they coincide with the Bentvueghels; thus showing a preference for northern Italy, and very little interest in the North of Germany, which was the top destination for Ecartico artists.



Migratory patterns of the Sculptors on the Move (1600-1699)

Data integration therefore allows for contemplating multiple perspectives of one same matter instead of ambitioning or pretending completeness and thus lay closer to methods of digital hermeneutics.

- **Reversed classification.** Nodegoat offers the possibility to, instead of identify classifications and assign these to objects in a dataset (like 'sculptor' or 'German', define a multi-faceted filter spanning multiple datasets in which they define any number of parameters that are associated with a classification. This reverses the classifying process as the definition of the classification is identified by the exchange between parameters of the classification and attributes of the object. For the classification of 'Artist' this allows for an inclusion of persons who yield positive on the filter ('pupil of person who produced paintings' AND 'has had paintings in an exhibition'). No external labelling or self-labelling of the classification of 'Artist' is needed.

LAB1100 developed a scenario that profits from this implement by dynamically selecting all persons who are associated with the Bentvueghels. Persons are labelled Bentvueghel when they match a

reversed classification with a configuration that checks whether a person is connected to the society 'Bentveughel' (via the sub-object 'Membership' of this object) and checks whether a note describing this person contains the strings 'lid' (member in Dutch) and 'Bentvueghel'. As a result, notes (texts indicating someone's affiliation to the Bentvueghels) and nodes (links) are brought together to filter information in the database regardless of their characteristics. In other words, people related to the society according to notes but lacking a relational link to the society are filtered together with members who do have a direct link to the society. This allows for the different datasets to complete each other; Ecartico provided links to the Bentvueghels through text, while Hadrianus gave links to the society. The result is a filtered dataset that was not available before data integration and did not require the extra work of linking members from Ecartico to the society 'Bentvueghels'.

Besides, a reversed classification can be configured to be shown at all times in the database, thus indicating a characteristic of the object that the researcher finds significant while exploring the database. For instance, because the sign ~ **B** ~ was placed next to those related to the Bentvueghels as indicated above, membership to the society is visible while carrying out research about migration patterns of artists, thus allowing for comparing migration patterns of Bentvueghels to those of artists who were not related to this society.

5. Further research

While *Mapping Notes and Nodes* has more than satisfied our ambitions, many experiences and aspects led to the formulation of wishes for the future along the way. Because of the success achieved in this project and the new questions it triggered, efforts are being made as to continue the lines of inquiry that lay at its core. Charles van den Heuvel has handed in a proposal for lecturing at the Digital Humanities Conference that will be held in Sydney, Australia in the summer of 2015. Moreover, he applied for funding for a Pre-PhD project from CREATE (UvA), a platform created for research on Amsterdam's cultural industries throughout the centuries employing digital tools and methods. Leonor Álvarez Francés is writing an article together with Charles van den Heuvel, Pim van Bree and Geert Kessels, and will seek for holding presentations on the project as means for assuring dissemination of the insights it brought along. Susanna de Beer, who at the moment works at the NIAS as a fellow researcher, has organised a workshop on March 3rd and 4th in order to bring researchers together who focus on Rome and employ digital tools. Finally, Lab1100 has expressed its desire to continue working on spin-offs of *Mapping Notes and Nodes*.

This section reflects on additional data that would lead to new knowledge on the study of Roman and/or Netherlandish cultural industries as well as on future developments that would foster research on digital humanities at a technical level.

a. Content-related additions

- **Systematic data on residence in cities.** If the data on addresses in Rome we have from Hoogewerff would be completed with addresses in Rome of other social groups, such as for instance botanists, patrons or Italian painters, new questions could we asked in regard to networks and patterns in this Italian city. Moreover, when complemented with similar data, addresses of people involved in

Amsterdam's cultural industry could be related to residence patterns in Rome. It would then be possible to compare patterns in different cities and find new links and networks binding them.

- **Structured data about arts of work** and analysis on the role of networks for professional success. Our database is short on networks of patrons that would make it possible to study the mechanics of assignments. For example, data on the Giustiniani as patrons as researched by Erin Downey would add a new dimension to the database, allowing for analysis on the secrets behind professional success and failure in the Early Modern artistic world.
- **Links through Iconclass codes.** Iconclass, as a systematized but natural language independent classification system offers numerous possibilities for linking objects of diverging characteristics by using a code of numbers and letters. It is flexible because it is always open to additions. Within *Mapping Notes and Nodes* we experimented with poetry and paintings as sources, but other possibilities include sculpture, narrative sources, music or architecture. Iconclass codes can in the future be used to identify similar representations of, for example, persons, national tales and mythology, narrative motifs, personifications and symbols.
- **Corrections on Hoogewerff.** Scholars specialised in the presence of people from the Low Countries in Rome in the Early Modern period agree that a considerable amount of data collected and published by an authority on the artistic relations between Italy and the Low Countries, Hoogewerff, is incorrect. Introducing corrections could lead to the discovery of new links and networks.

b. Technical developments

i. For data integration

- **Tool for identifying people** despite spelling variations and employing biographical data. Algorithms capable of finding a name despite spelling variations are already common in the digital humanities. However, when working with big amounts of data being integrated from different sources, recognition of similar biographical data would make the work much more efficient and avoid manual import to be the main method for finding matches.

Indeed, a known difficulty is the fact many people wore the same name back in the Early Modern period. Another obstacle for finding a match only based on spelling similarities is the existence of translated names and aliases. A tool suggesting matches between persons from different databases that would not only base itself on the name but also on similar biographical data would undoubtedly quicken and improve the efficiency of data integration.

- **Standardization** of geographic locations and people. Finding common codes for identifying the same elements in the digital humanities worldwide is not a new need, but it is a still unsatisfied one. This demand is notably urgent when dealing with locations and people. Standardization with for instance Geonames and VIAF would save time, work, and therefore huge amounts of resources.

- **Automatic updates** from the final database into individual sources. While data integration can lead to plenty of new discoveries in the humanities, as *Mapping Notes and Nodes* has shown, it is likely every research group providing data wishes to keep its own dataset active and running independently. This does not mean they are not interested in update taking place in other databases. For this reason it is interesting to consider automatic updates happening in the source data when it takes place in the common database.

ii. For data analysis

- **Visualize uncertainty.** Two databases integrated within this project, namely Ecartico and Hadrianus, contain degrees of certainty for specific elements. Besides, during the data import process it became necessary to include an extra uncertain element, the likelihood of a person coming for database 2 to be the same as a person in database 1. Accordingly, a new field was added, 'Possible identity'. It is possible to visualize this connection on network visualizations. Last but not least, uncertainty concerning dates is as well common in the humanities. Nodegoat did not offer a satisfactory solution to the problem of how to provide an *Accademia* in Rome with dates when these are unknown. In most digital tools, and this is the same for Nodegoat, lack of information leads to invisibility. An *Accademia* without dates will not appear in any visualization. Giving a fixed date when there is uncertainty concerning the data is obviously not an ideal solution. For this reason Lab1100 has worked on a conditional formatting instrument that makes it possible to visualize fuzziness. Birth and death dates that contain notes are represented in a different fashion. The content of the notes can be visualized by placing the cursor on it. This way the researcher does not miss data that, although fuzzy, can be useful; nor does he get deceived by data that seems definitive when in fact there are doubts about its accuracy.

- **Automatized merging of unique data** imported from different sources. Historic research is characterized by the potential existence of two (or more) answers to one question. Even when concerning data that does have a unique possible answer, which is the case for the exact dates someone was born and died. While Nodegoat allows for accommodation several birthdays for one person (a persistent phenomenon in *Mapping Notes and Nodes*), this leads to equivocal results and analysis. This is due to the fact both birthdays add up to the importance of a geographic location on a map in which we seek to answer the question: where were the most people in my database born? Ideally, Nodegoat would make one period out of the scope covered by both birthdays, thus avoiding misleading visualisations while preserving the fuzziness of the data at hand. Otherwise, the choice can be made as to which database is considered as the most trustworthy and import unique data from that one.

- **Source of data visible** at all times. Most researchers employ several sources for a project, but do not equally rely on all of them. Therefore, it is important to be able to identify the origin of the data being used at all times. Whether this is done by means of different colours, letters or symbols, this information can best be shown in all spaces of the interface.

- **Identify different relations within networks.** It is not unusual to add codes to network visualizations in order to distinguish diverging relations such as parental or professional connections. Another distinction relevant for the study of cultural industries but also, for instance, war negotiations, is the quality of this relation. Making a difference between a friendship, a relation

of competition or enmity brings network visualizations to life. By reflecting the characteristics of a relation, analysis of this data can go further, revealing unexpected (weak) links or patterns.

- **Representing historical locations.** When seeking patterns in cities or outside them, it is necessary to count on maps that identify and locate historical indications as those named in pages 10 and 11. For representing addresses that still exist nowadays and include house numbers, a tool placing them at different distances would be an addition, but would ideally rely on a historical map as well.

6. Conclusion

Mapping Notes and Nodes aimed at creating generic import and visualization modules and at assessing which data were worth bringing together within the Digital Humanities. The instruments created are integrated within Lab1100's platform Nodegoat, and have not only served the project's specific aims but are designed to import and visualize any heterogeneous group of data sets. By selecting cases that offered as many perspectives as was possible within the scope of this project, we reached conclusions relevant for many scenarios of data integration. This allows us to offer new insights on efficient data integration in the humanities.

(Un)useful data: what to bring together

Most data imported within Nodegoat were useful for analysis of historical phenomena in one or more of the case studies. Biographic data gained in networks and residential places from Ecartico, *Sculptors on the Move* and Cosimo III's correspondence, and was complemented with works of art and concrete addresses in Rome thanks to Hadrianus. Integration of this sort of data led to new insights concerning cultural industries as shown by the cases on the Bentvueghels, Cornelis Meijer and Casper van Wittel. The Iconclass case showed that data integration of two datasets with similar objects but missing common classifications can be enriched by importing a set of codes that links them. However, some data were not employed despite being susceptible to produce insightful results. As we shall see, this is due to the presence or absence of these data in the different datasets and whether they share the same structure or not.

Data present in different databases but slightly different can be worth bringing together. This is the case for geographic locations. Paris might have received slightly different coordinates within Ecartico than from the makers of *Sculptors on the Move*, but representation in a big scale map will not show this difference and interpretation can take this aspect into account. While radically different classifications make filtering impossible, slight difference such as coordinates on a big scale map do not necessarily interfere with proper, scientific analysis.

Data present in different databases but with conflicting structures do not offer the possibility of combined analysis. Ecartico stored sorts of 'painters' that differ from those to be found in *Sculptors on the Move*. For instance, 'amateur painter' or 'miniature painter' can only be found in Ecartico, and 'painter of altarpieces' is a classification in *Sculptors on the Move* alone. In case researchers wish to pursue an inquiry on the genres painters in these databases belong to, data curation unifying terms needs to precede data integration. The possibility also exists as to import these data without performing curation tasks. This is a more realistic solution given curation will not always be a possibility due to the lack of time and/or resources. Besides, and more interestingly, this approach lies closer to digital hermeneutics. At the end of the day, all databases in the humanities are a selection of sorts, even when carefully planned to cover the whole spectrum of a topic.

Contrarily, data present in only one database will not lead to this conflict; they offer more information about common elements across databases. The uncertainty degrees of Ecartico and Hadrianus are a good example of this. Ecartico stored data on how certain researchers are on a connection between two people, while Hadrianus reflected on the reliability of the Bentvueghels' addresses in Rome.

Data unique in nature but fuzzy and thus diverging in different datasets, such as birthdates, contributes to analysis but presents risks. A tool being able to process and unify these data into one unique attribute is desirable. This would avoid representing several birthdates for one person and therefore misinterpretation.

Datasets apparently greatly diverging in essence, like poetry and paintings, can render new discoveries when data curation provides a common code linking them. Such was the case for the scenario employing Iconclass codes.

Notes and Nodes: how to bring data together

In general nodes are data that lend themselves better for visualizations than notes. Nodes are links that a relational database stores as such, and both network and geographic visualisations are prepared to render connections in a clear manner. Notes can be stored in two very different manners: as objects related to two nodes (such a book lent by A to B), or as mere text. The first ones are easily displayed for the reason described above. The case of the correspondence network showing Meijer's publications as nodes is a very good example of this. Notes stored as text are in general more reluctant to be displayed in a uniform manner that the eye can interpret. Fuzziness in dates ('between May and October 1574') can be displayed without interrupting global data visualization by showing dates containing a note in a different fashion. However, this is a representation of a characteristic of the note (its fuzziness) and not of the note itself.

A further consideration is the fact nodes are usually displayed in a standard manner, namely as a link between A and B. However, the characteristics of this link can be of outmost interest when researching cultural industries. When visualizing art networks it is very much relevant to know that the connection between two people is collaboration on a project, or precisely the opposite: competition. Indeed, Ecartico stores whether a relation is 'professional' or 'personal', which is a manner of creating a node complemented with a note; a half-way between a note and a node. Relations such as patron-artist or artist-pupil are often stored as links as well, but the characteristics of such relation like 'friendly relation' or 'A is critical about B's book' are not common. It seems unnatural to create a scale reflecting the positive or negative characteristics of a relation, but avoiding it makes us miss information in the visualization.

As a matter of fact and especially within the humanities, nodes are simplified notes. The node 'Govert Flinck was a pupil of Rembrandt' is a simplified version of the real connection between these people, missing temporal and emotional characteristics. Temporal characteristics are easily added to a link in a database by means of a date, but emotional classifications, although not less relevant, remain elusive.

When bringing databases together this aspect becomes the more evident. While Ecartico stored a relation of a painter portraying somebody as a node, Hadrianus did so by means of a note. The tendency is then to reduce the notes to nodes because of the easiness nodes pose for visualization. However, this step should not be taken lightly: information will be missed by doing so.

In conclusion, data curation can better be invested on achieving deeper research insights than on more comfortable data integration. If the goal of bringing diverging datasets together is reaching results otherwise elusive, why choosing for missing information on the way? Data curation in the

humanities needs to respond to the complexity of interests rising from the field. Therefore, functionalities blurring the differences between notes and nodes need to be developed.