Trailblazing Metadata: a diachronic and spatial research platform for object-oriented analysis and visualisations

Pim van Bree (pim@lab1100.com) researcher and software engineer at LAB1100, Geert Kessels (geert@lab1100.com) researcher and software engineer at LAB1100.

ABSTRACT

Object-oriented approaches applied in the humanities recognise associations between human and nonhuman actors and expose relational forms of analysis. In order to make these relations meaningful, temporal and spatial attributes of the objects and of the associations between objects need to be stored and included in the analyses. To facilitate both an object-oriented form of data management as well as diachronic modes of analysis, we have developed a digital research platform. This research platform has been designed to allow scholars to determine and design relational database models. The platform can be used in multiple collaborative configurations to fuse various modes of authorship. Due to the focus on relations and associations between heterogeneous types of objects, the platform is equipped to perform analyses spanning multitudes of objects in order to gain qualitative insights.

We identify two types of data in this process: (1) apparent structured data and (2) non-apparent structured data. (1) Apparent structured data sets, such as correspondence networks or lists of published works, can be entered readily. The analysis of their metadata will reveal relational patterns, contextualised over space and time that provide forms of historical network analysis. (2) Non-apparent structured data sets require a form of pre-processing in order to manifest the inherent structures of the data set. In this process, data such as interviews, bodies of texts, pieces of music or philosophical debates are dissected in order to make the inherent structures of the data apparent. This leads to the identification of objects (a person or event named in an interview, a word or sentence structure in a text, a set of notes in a piece of music, an argument in a philosophical debate) and the relations and association between these objects (the presence of a person during an event, co-occurrence of words, repetition of sets of notes, refuting one argument by means of another). This process makes the aforementioned modes of analysis and representation available for non-apparent structured data sets.

Scholars in the humanities may utilize the platform in any aspect of the research process. As a result of the platform's relational data management capabilities it functions as an advanced card catalogue. By using the platform's analytical capacities, relational, diachronic and geospatial forms of analysis can be executed. Also, as both the raw data and the modes of analysis can be directly shared with a wide public, the platform operates as an end product of the research process, or starting point for the next.

1. INTRODUCTION

In order to facilitate an object-oriented research process for the humanities, we have developed a research platform that combines relational modes of analysis with spatial and chronological forms of contextualisation. The platform has been developed to allow scholars in the humanities to build data sets based on their own data base design. Once the data sets have been formed, multiple modes of analyses may be executed. The data as well as the modes of analysis all focus on three facets: an object-oriented relational approach, inclusion of geospatial coordinates and temporal attribution. By combining these elements within one platform, scholars will be able to process and analyse complex data sets relationally, diachronically and spatially. By embedding data in both chronological and spatial contexts, these fundamental characteristics of research in the humanities are acknowledged and included in the platform.[1]

This article will give an outline of the concept on which this platform is based. Secondly, the architecture of the platform and a number of exemplary data sets will be given. Lastly, the benefits of our approach as well as points for further development will be presented.

2. METHODOLOGY

2.1 Object-Oriented Data Design

In order to be able to execute object-oriented analyses that rely on meaningful relations and provide temporal and spatial perspectives, well curated data is a prerequisite. To obtain a suitable set of data, scholars need to decide at the very start of a research process how raw data will be structured and saved. This has a number of implications and consequences. Firstly, the designed data structure will form the basis of the complete research process. This means that during the aggregation, analyses and presentation of the data, the scholar relies on the same data structure. This continuity revises the epistemological process as the clear cut between the drawing table and the end product becomes in effect non-existent. Secondly, each object in the data set will be treated equally in order to be able to rely on a complete set of data. This means that the scholar is challenged to evaluate an object contextualised in its own network and to acknowledge objects otherwise overlooked.

For every object that will be saved, internal relations have to be identified as well as associations with other objects. Relations describe intrinsic aspects of the object (author, composer, etc) whereas associations reveal the position of the object related to other objects (member of, presence during an event, etc). To be able to temporally and spatially

[1] Although many tools for both data management, analysis and visualisation exist, none of them offer a comprehensive combination of both relational, diachronic and spatial modes of analysis. As all these facets are of vital importance for any form of research in the humanities, existing platforms rarely support the full scope of a research project.

contextualise the relations and associations, they have to be enriched with both spatial and temporal coordinates.

Following an object-oriented data design, an object is defined by its biography as initiator, while having the products of its existence – or offspring - defined as a relation or association within the produced objects. This means that when describing a type 'person', relations and associations like birth, death and domicile, or only relational definitions like spouse, are part of its biography. In other objects it is related as the author of a book or the child of an other person and associated by being employed in an institute, a member of an organisation, or the participant of an event.

We define each description of an object as ‘metadata’. This means that the object ‘interview’ is described in the same manner as the object ‘event’ that may be distilled from the interview object.

2.2 Assemblage of objects

The object-oriented modes of analysis we describe are closely connected to the fundamentals of ANT.[2] We infuse this method with an ingrained focus on spatial and temporal attributes.[3] This approach identifies three levels within the process of an assemblage of objects.

The first level establishes the object by the identification of its own definitions and the definitions of its associations. These definitions may be self-evident (i.e. a name of a person or a title of a book) or relational (i.e. the child of a person or the author of a book). Within this level of the process, we emphasise the focus on identifying all objects at hand. The inclusion of apparently peripheral objects is of vital importance as the weight or importance of each object is to be determined by the relations and associations of the objects themselves. To equip the objects with attributes needed for spatial and diachronic modes of analysis, temporal coordinates and geolocations are to be stored within its associations extensively.

The second level of this process identifies the full complexity of definitions and associations directly related with the object. Within this process all available definitions need to be acknowledged in order to establish an overview of the multitude of perspectives by which the object may be defined and to be able to include conflicting or contradictory definitions. The controversies that arise from these irregularities will thoroughly expose the object.

The third and final level of this process will reveal the associations between the objects that have been identified relationally, in space and in time. These associations (i.e. the presence of a person during an event or the acquisition of a book by a library) establish

[2] Latour, Bruno. Reassembling the Social. Oxford: Oxford University Press, 2005.

[3] This approach is related to the recent development in the field of cultural topology. See further: Theory, Culture & Society, July–September 2012; 29 (4-5), Special Issue on Topologies of Culture

the contextuality by which an object is to be approached. The complex interlinkage between objects expose the relativity of the objects and allow them to define the assemblage.

2.3 Programming

To be able to store and relate objects neutrally, relational and dynamically the platform leverages a structure similar to its goal; a combination of object-oriented code and object–attribute–value–version database storage, following a reciprocity between description and definition. The platform traces search queries, object relations, or a combination of these, and recreates all the objects in the trail when requested.

3. THE PLATFORM

3.1 Data Management

The research platform functions like a normal database in the sense that it consists of sets of tables of which the columns can be defined and rows can be added. However, in order to facilitate the use of multiple chronological and spatial references per row, we have opted for an alternative terminology that defines the internal relations within the system in a logical and consistent manner. This leads to the following infrastructure.

Highest in the hierarchy are types (tables), which define and contain the data (for example: 'Letters', 'Persons', 'Interviews').

Each type contains objects (rows) of which the object descriptions can be identified (columns). These object descriptions can contain strings of data (for example: 'Name', 'Date', 'Text') or relational references to other objects (for example: 'Person', 'Location') of the same type or any other type.

Each object contains one or multiple ‘subobjects’. In the subobjects relations to other objects are defined that include both a temporal and a spatial reference. This design allows for the definition of multiple chronological and spatial instances per object. In the example of a letter this means that each object consist of two subobjects, namely the subobject 'Sender' and the subobject 'Receiver'. Both subobjects have references to the relevant persons associated with the instance of Sender or Receiver [see fig 1].

By use of this model, relations and associations between objects are always explicitly specified and always have a chronological and spatial component. In the example of a letter this means that the model reveals the nature and direction of the relation (from the subobject Sender (person A) to subobject Receiver (person B)), exposes when this relation was made, and to which locations the relation relates. All the data in the platform is therefore intrinsically relational, temporal and spatial.

3.2 Examples of data sets

The platform can host any type of data as the data structure of the research platform can be designed from the ground up and alterations can be made during the course of a project. A first prototype of the platform based on correspondence networks was developed in the scope of a project of the Study Platform on Interlocking Nationalisms at the University of Amsterdam. In this project, the connectedness of nineteenth century intellectuals, scholars and artists was analysed and visualised over time and space in order to expose the diachronic and spatial development of these networks.[4] Due to the fact that the data used in this project had an apparent structure, no pre-processing was required.

When a scholar relies on non-apparent structured data sets, the inherent but non-visible structure needs to be revealed. This process calls for a definition of objects from within the data set and the identification of relations and associations that tie these objects together. In the case of a data set consisting of interviews, this process leads to the definition of a multitude of objects cited in an interview. These objects can be diverse of nature: persons as well as events, buildings, books or other non-human actors should be identified and saved. These objects should be described exhaustively. Relations that define the intrinsic characteristics of the objects need to be defined (i.e. the architect of a building or the author of a book) as well as the associations the object has (i.e. a person attending an event or the presence of a book in a library). As result of this process, tools to perform diachronic and spatial analyses are made available to non-apparent structured data sets.

This approach can also be employed in less obvious applications. A philosophical debate spanning multiple centuries and a vast region can be dissected logically in order to create an independent instance per argument. This allows the scholar to perform a relational analysis on the coherence (or incoherence) of the argumentation structure. Due to the inherent local and temporal coordinates of the instances, the development of the debate can be mapped over space and time.

3.3 Analyses

The object-oriented approach that lies at the basis of the platform facilitates numerous applications. As relations between objects are constantly saved, described and given chronological and geographical coordinates, well connected objects surface based on the value of their connectedness. This process allows the objects to self-define their role in the data set (peripheral, central, hub). Also, as the data set grows, geographical and chronological patterns will reveal themselves.

The relations and associations connecting the objects allow for a complex mapping of the interlinkage between objects. These relations and associations can span over one or

[4] J.T. Leerssen, P. van Bree, G. Kessels, M. Witberg. *Spin Time: Dynamically Visualizing How Diffusion Patterns Evolve over Space and Time.* Working Papers European Studies. Amsterdam. 2012.

multiple types. In the example of a letter, objects in the type of 'person' are connected to each other via the type of 'letter' [see fig 2]. From the perspective of the object 'person A', this means that an association is established with object 'person B' via the object 'letter X'. From the perspective of the object 'letter X', this means that two relations are made: one to the sender of the letter (object 'person A') and one to the receiver of the letter (object 'person B'). This signifies that a single link can be both a relation and an association, depending on the chosen perspective.

When employing a multitude of the object letter consisting of the same sender and receiver, spatial and diachronic developments will expose themselves. Transformations in the localities disclose movements of the objects of 'person A' and/or 'person B'. Transformations in the intensity of the associations reveal the developments in the connectedness between the objects 'person A' and 'person B'.

By employing a multitude of the object letter and a multitude of the object person, but limiting the selection to one sender and a multitude of receivers, a different range of spatial and diachronic developments are exposed. In this scenario, the network of one person is mapped over time and space. Such a mode of analysis will reveal transformations in the locality of the sender as well as transformations in the regional scope of the receivers. Regions of influence and and of interest will be exposed in this manner. The diachronic perspective will be able to show the growth and reduction of the network of the sender.

If a selection is made of a multitude of the object letter as well as a multitude of the object person, the network will be exposed in its totality. In this scenario, the complete network can be analysed. Paths defined by the relations and associations of the objects within the network can be identified. These paths may connect one object with another, as the object of a person is linked to the object of a letter. A path may connect one object to another object via a third object, as the object of a person is connected to another object of a person via the object of a letter. Or a path may connect one object to another object via a multitude of other objects, as the object of a person is connected to the object of another person via a set of objects of letters and persons.[5]

In another example, relations are mapped over a multitude of types. When analysing the distribution of paintings over sets of museums, the relation between the object of 'Painter' and the object of 'Curator' can be defined via the objects of 'Painting', 'Museum', 'Collection', and 'Traveling Exhibition' [see fig 3]. This analysis can be executed by means of the aforementioned modes and complexities.

3.4 Visualisations

[5] A full elaboration of this application can be found in: J.T. Leerssen, P. van Bree, G. Kessels, M. Witberg. *Spin Time: Dynamically Visualizing How Diffusion Patterns Evolve over Space and Time.* Working Papers European Studies. Amsterdam. 2012.

Every relation that is entered into the platform can be visualised applying a set of different models. Diachronic network analyses convey overviews of the development of the interconnectedness between objects [see fig 4]. Diachronic spatial visualisations make the expansion or decline of influence insightful [see fig 5]. As the complexity of data sets grow, new modes of presentation emerge. Relations spanning multiple types will reveal highly differentiated qualitative patterns. Depending on the scale and completeness of the data set, these patterns will reveal themselves, rather than merely confirming the hypotheses of the scholar.

3.5 Collaboration

The object-oriented approach propagates a research process based on co-creation and collaboration. As all the data is saved in a structured fashion, multiple scholars of different disciplines can work on the same data set.

This process can take shape in the form of a nonhierarchical and a hierarchical structure. In the former situation scholars work together on equal footing, in the latter scholars are able to form groups of users that can enter data into the platform that is only used in the platform after is has been reviewed.

An important advantage of this procedure is the ability to reuse sets of data. Firstly, ‘raw’ data sets such as ‘authors’, ‘historical persons’ or ‘locations’ can be shared among scholars. Secondly, once a research process is finalised and has been formed into multiple modes of analysis, other scholars have the ability to continue to build on the outcomes. This workflow leads to a viable process of co-creation in which multiple forms of authorship are acknowledged.

4. CONCLUSIONS

In this article we have infused the debate concerning digital humanities with the appliance of a diachronically and geospatially elaborated form of ANT. The object-oriented nature of this approach may be of high value for material as well as non-material forms of research in the humanities. Moreover, by including both temporal and geospatial coordinates, diachronic developments, essential to research in the humanities, can be extensively mapped and analysed. This approach also affects the research process as forms of collaborations between scholars can be easily established.

In this process, scholars act like trailblazers within bodies of associative data in which the new paths they create function as the starting point of new research. This approach can be connected to the primordial concept of trailblazing, as envisioned by Vannagar Bush in 1945. A scholar trailblazes by creating trails in a maze of intrinsically associative objects. The trails allow one to interact, annotate and collaborate on any type of data, making it a

practice of active transclusion; a continuous distribution of authority between the author and the maze.[6]

The object-oriented approach poses itself in contrast to a number of current trends in digital humanities in which scholars passively follow algorithms designed by programmers to, for example, analyse bodies of texts. On the platform we propose, the scholar has complete control the design of the data structures, and thus, is in full control of the research process.

Eventually, as the complexity of the modes of analysis and forms of presentation expands, textual representations may become superfluous.[7] Instead, by trailblazing through sets of associative and relational data, evanescent paths are generated that stimulate forms of reuse and collaboration.

5. ACKNOWLEDGEMENTS

The authors thank Joep Leerssen for his support.

6. REFERENCES

A. Rigney. When the monograph is no longer the medium. History and Theory 49:100-117, 2010

J.T. Leerssen, P. van Bree, G. Kessels, M. Witberg. Spin Time: Dynamically Visualizing How Diffusion Patterns Evolve over Space and Time. Working Papers European Studies. Amsterdam. 2012.

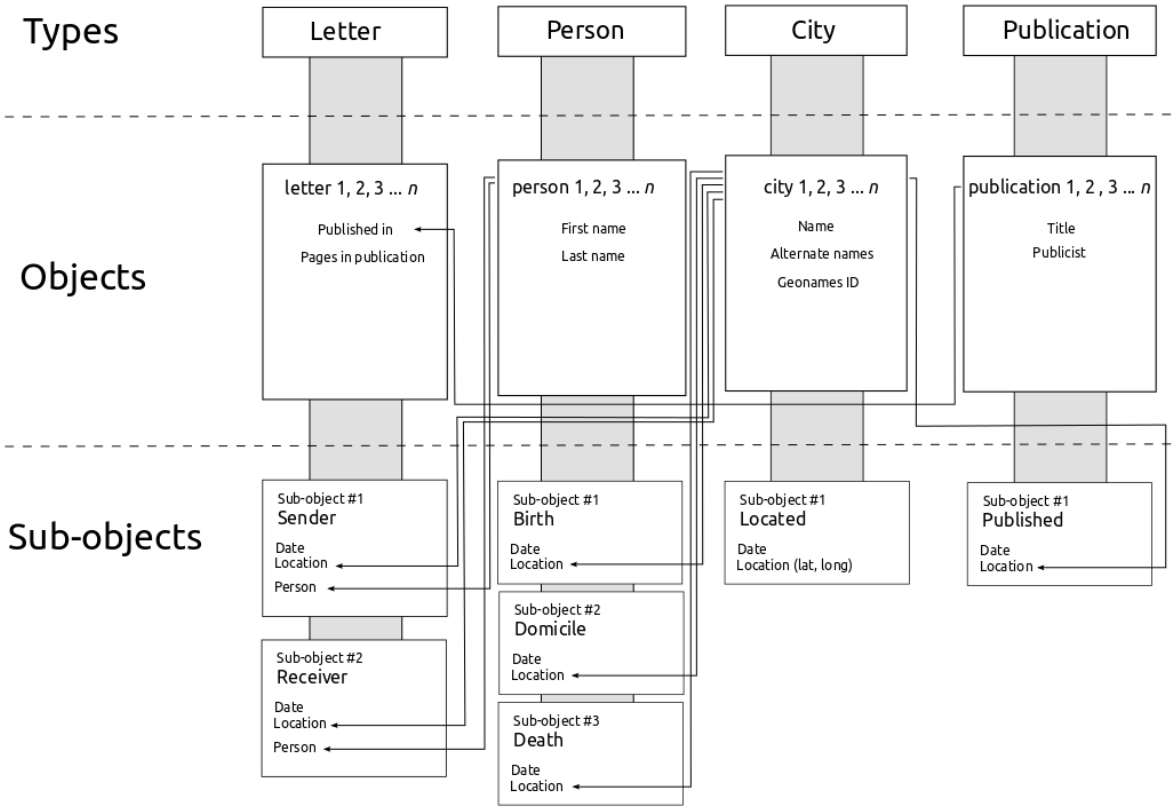
Latour, Bruno. *Reassembling the Social.* Oxford: Oxford University Press, 2005: 1-156.

Bush, Vannevar. 'As We May Think'. *Atlantic Monthly*, 176(1), 1945: 101-108. <http://www.theatlantic.com/doc/194507/bush>

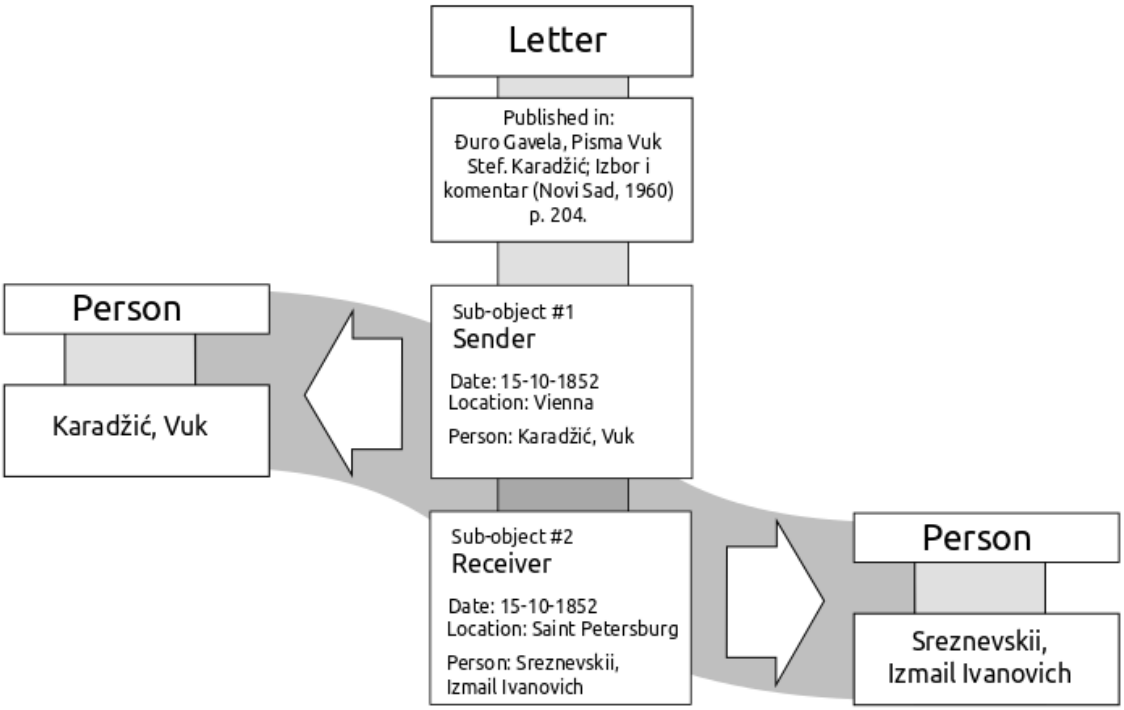
[6] Bush, Vannevar. 'As We May Think'. Atlantic Monthly, 176(1), 1945: 101-108. <http://www.theatlantic.com/doc/194507/bush>

[7] A. Rigney. When the monograph is no longer the medium. History and Theory 49:100-117, 2010

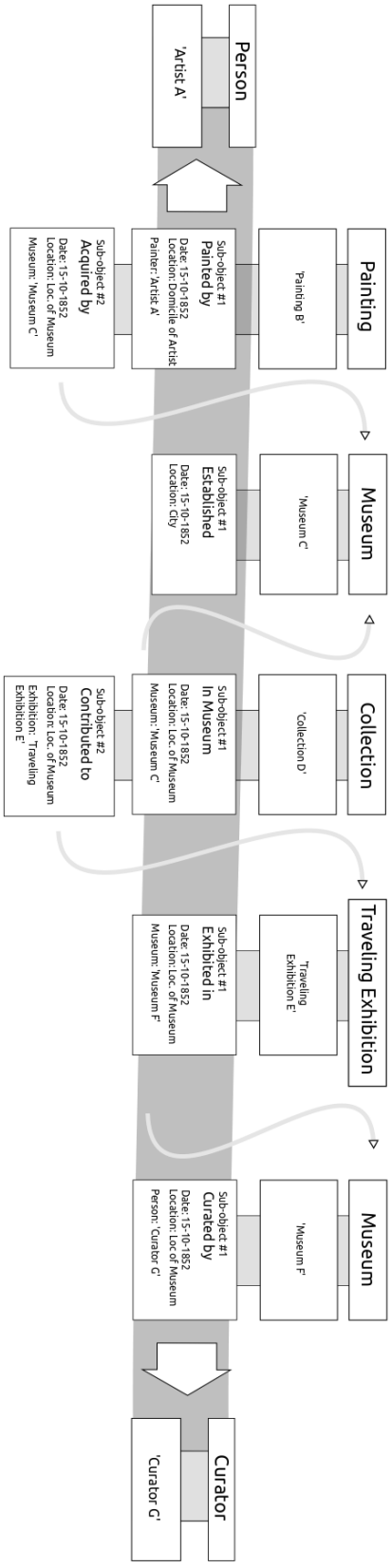
[Fig. 1]


Types
Letter
Person
City
Publication
Objects
letter 1, 2, 3 ... n
Published in
Pages in publication
person 1, 2, 3 ... n
First name
Last name
city 1, 2, 3 ... n
Name
Alternate names
Geonames ID
publication 1, 2 , 3 ... n
Title
Publicist
Sub-objects
Sub-object #1
Sender
Date
Location
Person
Sub-object #2
Receiver
Date
Location
Person
Sub-object #1
Birth
Date
Location
Sub-object #2
Domicile
Date
Location
Sub-object #3
Death
Date
Location
Sub-object #1
Located
Date
Location (lat, long)
Sub-object #1
Published
Date
Location

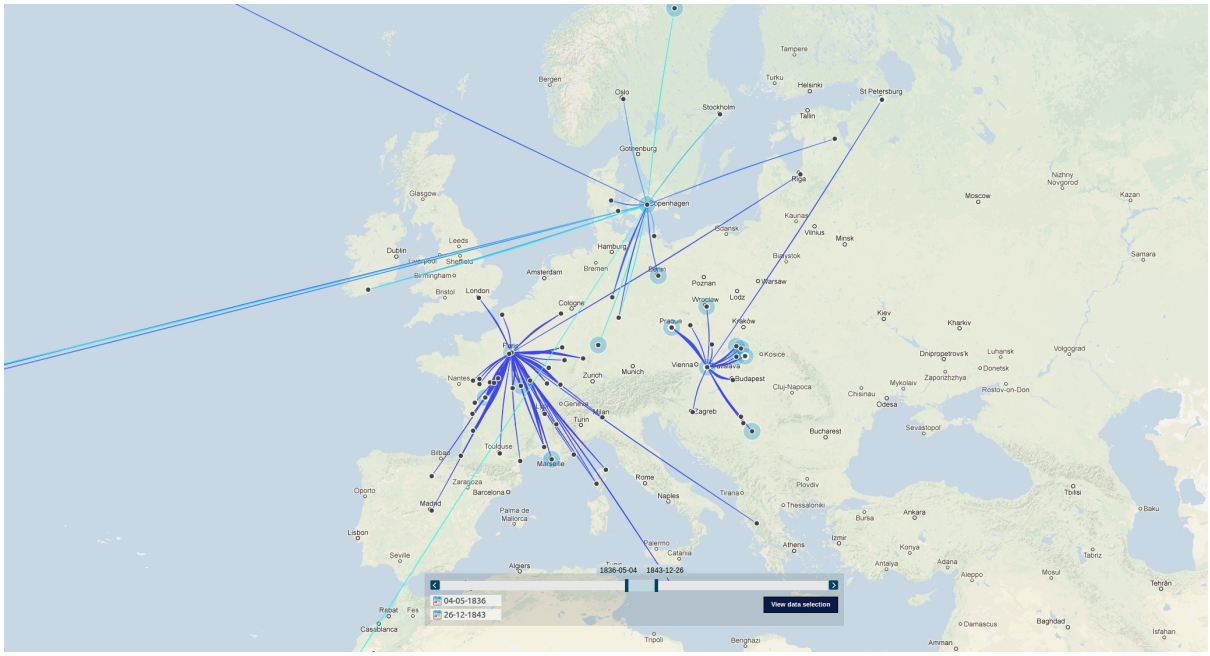
[Fig. 2]


Letter
Published in:
Đuro Gavela, Pisma Vuk Stef. Karadžić; Izbor i komentar (Novi Sad, 1960) p. 204.
Person
Karadžić, Vuk
Sub-object #1
Sender
Date: 15-10-1852
Location: Vienna
Person: Karadžić, Vuk
Sub-object #2
Receiver
Date: 15-10-1852
Location: Saint Petersburg
Person: Sreznevskii, Izmail Ivanovich
Person
Sreznevskii, Izmail Ivanovich

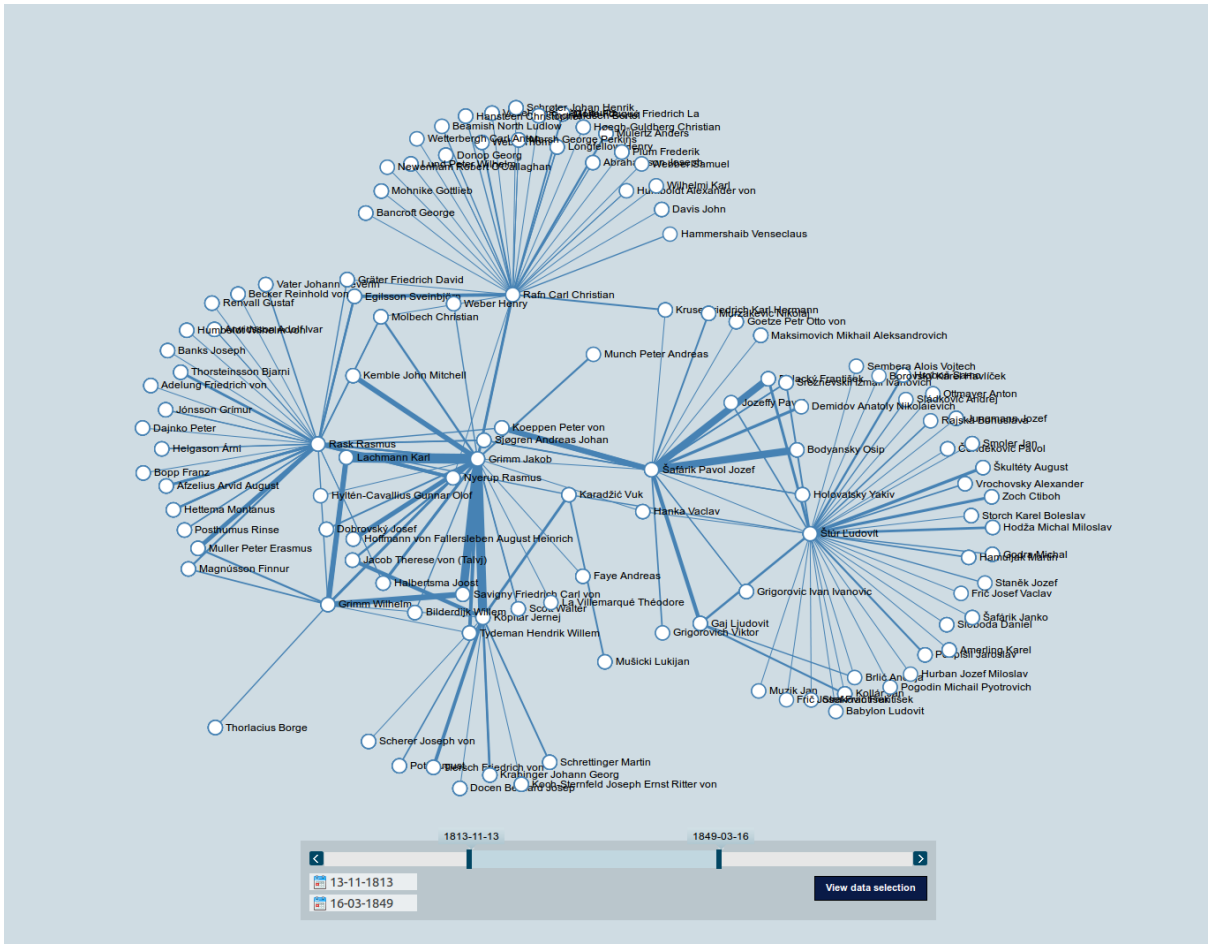
[Fig. 3]


Person
'Artist A'
Painting
'Painting B'
Sub-object #1
Painted by
Date: 15-10-1852
Location: Domicile of Artist
Painter: 'Artist A'
Sub-object #2
Acquired by
Date: 15-10-1852
Location: Loc. of Museum
Museum: 'Museum C'
Museum
'Museum C'
Sub-object #1
Established
Date: 15-10-1852
Location: City
Collection
'Collection D'
Sub-object #1
In Museum
Date: 15-10-1852
Location: Loc. of Museum
Museum: 'Museum C'
Sub-object #2
Contributed to
Date: 15-10-1852
Location: Loc. of Museum
Exhibition: 'Traveling Exhibition E'
Traveling Exhibition
'Traveling Exhibition E'
Sub-object #1
Exhibited in
Date: 15-10-1852
Location: Loc. of Museum
Museum: 'Museum F'
Museum
'Museum F'
Sub-object #1
Curated by
Date: 15-10-1852
Location: Loc. of Museum
Person: 'Curator G'
Curator
'Curator G'

[Fig. 4]


Tampere
Bergen
Oslo
Turku
Helsinki
St.Petersburg
Stockholm
Tallin
Gothenburg
Riga
Glasgow
Copenhagen
Moscow
Nizhniy Novgorod
Kazan
Kaunas
Vilnius
Minsk
Gdansk
Dublin
Leeds
Liverpool
Sheffield
Hamburg
Amsterdam
Bremen
Berlin
Birmingham
Bialystok
Samara
Poznan
Warsaw
Bristol
London
Wroclaw
Lodz
Cologne
Kiev
Kharkiv
Prague
Krakow
Paris
Kosice
Dnipropetrovsk
Luhansk
Volgograd
Vienna
Bratislava
Budapest
Zurich
Munich
Donetsk
Nantes
Cluj-Napoca
Zaporizhzhya
Rostov-on-Don
Chisinau
Mykolaiv
Odessa
Geneva
Milan
Zagreb
Turin
Bucharest
Sevastopol
Bilbao
Toulouse
Marseille
Rome
Naples
Tirana
Plovdiv
Tbilisi
Oporto
Zaragoza
Barcelona
Madrid
Palma de Mallorca
Thessaloniki
Baku
Bursa
Ankara
Lisbon
Izmir
Konya
Athens
Antalya
Adana
Tabriz
Seville
Palermo
Catania
Aleppo
Mosul
Tehran
Algiers
1836-05-04
1843-12-26
04-05-1836
26-12-1843
View data selection
Rabat
Fes
Casablanca
Damascus
Baghdad
Tripoli
Benghazi
Amman
Isfahan

[Fig. 5]


Schroter Johan Henrik
Hanaelen Christian Friedrich La
Besmish North Ludlow
Heegh Guldberg Christian
Wetterbergh Carl Anton
Mullertz Anders
Donop Georg
Plum Frederik
Abrah
Mohnike Gottlieb
Wilhelm Karl
Bancroft George
Davis John
Hammershaib Venseclaus
Grater Friedrich David
Vater Johann
Becker Reinhold von
Egilsson Sveinbjo
Renvall Gustaf
Weber Henry
Rafn Carl Christian
Molbech Christian
Goetze Petr Otto von
Maksimovich Mikhail Aleksandrovich
Banks Joseph
Munch Peter Andreas
Thorsteinsson Bjarni
Kemble John Mitchell
Adelung Friedrich von
Jonsson Grimur
Dajnko Peter
Koeppen Peter von
Rask Rasmus
Sjogren Andreas Johan
Helgason Arni
Lachmann Karl
Grimm Jakob
Bodyansky Osip
Bopp Franz
Nyerup Rasmus
Safarik Pavol Jozef
Afzelius Arvid August
Hytten-Cavallius Gunnar Olof
Karadzic Vuk
Holovatsky Yakiv
Hetterna Monanus
Hanka Vaclav
Posthumus Rinse
Dobrovsky Josef
Sturr Ludovit
Muller Peter Erasmus
Hoffmann von Fallersleben August Heinrich
Magnusson Finnur
Jacob Therese von (Talvj)
Faye Andreas
Halbertsma Jacob
Savigny Friedrich Carl von
Grigorovic Ivan Ivanovic
Grimm Wilhelm
Villemarqué Théodore
Gaj Ljudovit
Grigorovich Viktor
Tydeman Hendrik Willem
Mušicki Lukijan
Thorlacius Borge
Scherer Joseph von
Schrettinger Martin
Kraitinger Johann Georg
Stemfeld Joseph Ernst Ritter von
1813-11-13
1849-03-16
13-11-1813
16-03-1849
View data selection