

Weaving data, slicing views: a design approach to creating visual access for digital archival collections

Michele Mauri, Azzurra Pini, Daniele Ciminieri, Paolo Ciuccarelli

Politecnico di Milano, DensityDesign Research Lab

via Durando, 10, 20158 Milano – IT

{michele.mauri, daniele.ciminieri, azzurra.pini, paolo.ciuccarelli} @polimi.it

ABSTRACT

Digital archives metadata suggest a rich and complex system of relationships between the different properties of archived items, which is often not properly represented. *Lomen* is a research project aimed at exploiting the richness of digital archives, stitching up the relationships between entities and providing visual access to the system. This paper presents the design process used to create such visual access for architect Baldessari's historical archives. The research results in a digital platform that allows users to explore contents in a non-linear way, identifying patterns and fostering insight. The platform also aims at weaving together several levels of information through direct linking to archive entities such as projects, artifacts or individuals involved. Curators are also given the ability to elaborate theme-based paths, providing varied and unique entry points to the underlying data to users.

Author Keywords

Visual interfaces; information visualization; multiple views; digital archives

ACM Classification Keywords

H.5.2 [Information Interfaces and Presentation]: User Interfaces—Evaluation/methodology; Interaction styles; Input Devices and strategies; Prototyping

General Terms

Design; Experimentation.

INTRODUCTION

Providing access to digital archival collections for a general public is becoming a primary concern for curators. Cultural institutions are becoming aware that simply creating an online repository for an archive is not enough to reach a vast public, and that availability doesn't necessarily bring actual access. In fact, we are experiencing a growing need for the design of engaging interfaces that are capable of supporting the exploration of digital archives. While we

register a progressive shift from text-based interfaces to visual ones, the use of information visualization techniques is still experimental. There are several issues, including metadata inconsistencies, the representation of heterogeneous entities, the reconciliation of the practitioner and user perspective and the communication of the context to users with no specific background on the archive subject.

In the presented project, we address two main issues. The first regards the creation of a visual interface for archives that is able to represent different kinds of entities, identifying possible interactions that would allow users to visually shift between diverse types of entities. The second concerns the provision of a context by allowing an archive's curators to consolidate several layers of information in the database and build thematic paths linked to an archive's contents. While experts create their own specific paths, they perform a process of problem setting, that is, the activity of defining significant alternatives and specific means of exploring a certain situation. By way of interaction techniques, practitioners are provided with the possibility of shaping the knowledge space in a flexible, original and dynamic way [12].

Lomen is a term related to the etymology of the word 'loom', and recalls the nature of a tool and the idea of weaving and braiding information to produce new knowledge. It is a tool, as it enables people to explore information, to find new relationships and patterns created by new data structures. Also, it allows archive curators to study information from new perspectives. It is an "information loom" that allows curators to make improvements to the data and fosters the creation of new layers of information.

CONTEXT

The digitization of archival collections has created new ways to open contents to a wide public, creating new challenges and opening the field for new experimentation [14]. One of the main advantages is that it is possible to use techniques derived from data analysis and communication: between these, information visualization plays a key role in the creation of visual interfaces for archives.

One of the most coherent theories of information visualization use for digital archives was proposed by Hinton and Whitelaw [5]. In their paper, they state that these techniques can be used as a way to engage with a

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

CHI'13, September 16 - 20 2013, Trento, Italy

Copyright is held by the owner/author(s). Publication rights licensed to ACM.

ACM 978-1-4503-2061-0/13/09 \$15.00.

<http://dx.doi.org/10.1145/2499149.2499159>

wide public, providing explorative access to a collection. Presenting a case history, they describe how information visualization is being used to communicate non-scientific data, underlining how this kind of interface creates engagement with users, providing them with an overview of the archives and the ability to play with the data so they can explore it in a dynamic way. Whitelaw [16] pushes these considerations further, suggesting visual interfaces as an alternative to text-based search. In his criticism he states that by exploiting the features of digital technology, it is possible to create new kinds of interfaces, which are called *Generous Interfaces*. Generous interfaces offer users the entire contents of an archive, providing a rich overview: using a set of interactions, the user is then able to refine the contents down to a single collection item.

Both Hinton and Whitelaw present experimental interfaces rooted in these considerations. The reported cases, however, are built around a single visualization, presenting elements with consistent metadata. It remains unclear how archival collections containing different kinds of entities could be handled [15].

CASE STUDY: BALDESSARI ARCHIVE

This project focuses on the historical archive of Luciano Baldessari, a notable Italian architect. His production is divided in three main archive groups, owned by different institutions. The main archive group is managed by Politecnico di Milano, and contains Baldessari's technical drawings and his private correspondence. The other two archives are composed of photographs and paintings. The archive is divided by architectural projects: each project is described by geographic location, the actors involved, the chronological extremes of the realization, the related bibliography, the drawings and the correspondence regarding the project. The archive has recently been fully digitized, but the metadata associated with the items lacks consistency. Projects and drawings have been digitized with the related metadata, while correspondence, which constitutes the most considerable part of the archive, doesn't have any associated metadata, and is described only by the relative project. Archive curators have highlighted the necessity of creating access for these documents because of their particular relevance for scholars. For the first prototype of the application, we processed 53 projects selected by the curators, working on a corpus of 1,568 drawings and 7,388 letters.

The aim of the project was to identify how to provide a public, web-based access to the archive capable of making the relationship between entities explicit, through the design of a visual interface that allows the exploration of the archive. The application was designed to achieve flexibility standards for managing metadata inconsistencies and future changes to the database. An additional objective of the project was to provide the curators with a back-end for managing and modifying existing records and creating new ones.

Design considerations

Baldessari's archive presents characteristics that can be extended to other archival collections, allowing general considerations on visual access for digital archives.

- The archive isn't a monolith, but is in constant evolution. Curators and scholars are always working to improve and enrich the archive.
- The archive is composed of several kinds of entities, each one characterized by specific properties.

Aiming to provide wide public access, the most relevant goal was to find a structure that users could understand, rather than translating the archive catalog. Based on our practice in developing interactive visualization applications for a general public [1], we analyzed the feasibility and benefits of a multiple-perspective interface and the importance of identifying the specific objectives from each point of view. In cooperation with the archive's curators, we identified some possible exploration paths throughout the archive in order to create a proper framing of the context. Such paths allowed us to identify the conceptual entities that a user could interact with, and the specific properties of the visual interface that are capable of representing the particular aspects of various entities.

Information Visualization for archival collections

The application of information visualization techniques to design interfaces for digital collections is a growing practice [7], but is still used mostly in an experimental way, often to create compelling interfaces rather than to simplify access to data. Far from applying an analytical approach to visualization design, the *Lomen* project aims at creating a usable and easily understandable interface for a wide public. One of the most important advantages in exploiting information visualization methods for the design of visual interfaces concerns the ability to provide a context to the archive in terms of its relative size and scale. As Hinton and Whitelaw [5] state, "Visualization enables us to literally show everything, to display large volumes of data in a way that reveals patterns and communicates context, but also provides access to the fine grain of individual elements." Information visualization is often used to create an overview of an archive that can be read on a computer screen, creating a set of interactions that allow the user to filter and drill down to the single elements of an archive. This approach is consistent with Shneiderman's 'visual information seeking mantra,' which defines the steps an interactive visualization should provide: *overview first, zoom and filter, then details-on-demand* [13]. These guidelines are widely accepted for designing interactive applications [4], and they suggest an approach that moves from a holistic vision to specific information. It also reflects an analytical use of information visualization, where data are generally consistent and share the same metadata, unlike in archival collections where it is possible to find several levels of accuracy and a general heterogeneity in

properties and data formats. The difficulty in directly applying these principles is due to the nature of the data to be processed, which requires a highly qualitative approach and brings a considerable degree of complexity.

In the *Lomen* project, the application is built on an architectural archive composed of projects defined by geographic location, the year of realization, the people involved and the technical drawings made by architects. So much information is available that identifying the best way to represent it is a complex issue: geography implies maps, dates require a visual representation of time, representing relationships among people requires networks and graphs. It is not possible to find a unique visual solution that clearly shows all the information due to two main issues. First, the units involved are very different from each other: while representing a project's sequence over time, it becomes impossible to represent the actors; while representing Baldessari's social network, it is not possible to show information about technical drawings. Second, the visualization of all the properties of even a single kind of unit is problematic. Again, while representing the temporal sequence of projects, it is difficult to also show their geographical distribution or the people involved.

Multiple views

As previously shown, even if it were possible to encode any kind of information in a single visualization, it would lead to an over-complicated interface, inhibiting the potential of the visual representation.

To cope with this, it is necessary to create different perspectives on the database, giving users the ability to shift from one to the other in a coherent way. For this reason, we will call each visualization *view*: a perspective from a vantage point [4] on the information contained in the database. Views are linked together through interaction patterns, allowing the user to seamlessly browse the archive's information, collecting new insight from movement through it.

Specific literature [2] generally associates multiple views with multiple coordinated visualizations. This approach applies the *navigational slaving* interaction technique [2], which implies that the actions and the movements performed in one view are automatically propagated to other views.

The *Lomen* project, in contrast, employs a system of alternative views, where the user can see the archive through one view at time. Having multiple simple views linked with clear rules, the user is able to quickly understand the function of each visualization and create his or her own exploration strategies. Also, as each visualization is distinct from the others, it becomes extremely simple for developers to modify a single visualization or to add a new one. In each view, elements are represented individually, so the user is always provided a direct mapping between data objects and visual elements.

Application development

The identification of the most appropriate views was the focal point of the project. Rather than developing ad-hoc visualizations mixing relevant variables of multiple entities, we chose to develop a modular design process that could be expanded over time. Three main issues influenced the choice of a modular design. First, at the time of the project launch, the archive database was still under construction. This means it was possible that new entities and new metadata would be added to it, making the prototype rapidly obsolete. The second issue is that views are in fact representations of a specific task identified by archive curators: through the manipulation of the prototype, new ideas can arise, requiring additional views. Finally, we want to identify a design process that can be extended to other cases, for example, other historical archives.

The design process can be summarized in three main phases:

- Identifying entities
- Designing views
- Identifying links

Each step could be expanded and enriched during the prototype development, independently from the others. *Entities* refers to units that share the same metadata and are intended to be quickly understandable by users. *Views* are built upon a specific kind of entity, and focus on a specific set of the entity's properties. *Links* are relationships between different entities (e.g., person *x* involved in project *y*). As visualizations are built on a specific entity, links between entities will be mostly used as hyperlinks to move from one visualization to another.

Entities

As previously stated, the application was designed starting from ideal exploration paths identified in consultation with archive curators. The selected paths are: Baldessari's social network, the architect's stylistic evolution, the geographical distribution and the visual comparison of drawings. Starting from these, we found three main entities along which the corpus is structured: projects, documents and actors.

Projects can be thought of as the pivotal elements of the archive, as every document and person in the collection refers to one of them. Every project is characterized by its chronological and spatial coordinates, a brief textual description, the actors involved (e.g., collaborators, correspondents, clients, etc.), the linked documents and a bibliography.

Documents include every work made by or related to Luciano Baldessari: sketches, drawings, letters, photographs, models and so on. Every document is characterized by a signature, a date, a digital representation (in JPEG format) and a set of physical attributes and descriptions (support, instruments used, scale, preservation status, etc.).

Finally, actors are defined as every person who was in contact with Luciano Baldessari for a particular project. They are characterized by their name and their role in the project.

Database building

Once the application structure was identified, the database creation followed. The overall *Lomen* logic is based on entities, each one defined by a set of properties. While all the properties are set for each entity, none of them is compulsory; in this way, the database can handle different levels of accuracy and possible inconsistencies. These inconsistencies provide useful information for scholars, and must be kept.

In the original database, only two kinds of entities were described: projects and documents.

Involved people were recorded in the project's metadata. In order to obtain a viable structure for visualizing the architect's network of contacts, though, they were treated as separate entities associated with a specific project based on their role. This choice was also made considering other possible implementations of the *Lomen* framework in which external actors could play a more important role.

The document entity is an abstract class with two sub-entities related to projects: drawings/photographs and letters. Drawings and photographs refer to paintings, sketches, models and prints related to the architectural works of Baldessari. Letters are used to catalog notes, correspondence and reports related to a specific project and are not necessarily written by the architect. The metadata associated with documents have been modeled in order to apply them to a vast range of artifacts while still taking into account the archival notation.

Besides the data obtained from the archive, we defined another entity that can be used by registered users in order to describe thematic paths (as later discussed). Each path is composed of a set of theme details, which may be seen as the steps of a general essay. Each one of these steps is composed of a title, a textual description and an image with a related caption, and has a chronological frame so it can be arranged in a timeline-based visualization. This entity is also used in the Baldessari implementation to compose the architect's biography.

INTERFACE DESCRIPTION

The *Lomen* framework is structured around entities extracted from data and identified as the main elements of the corpus. For every entity, we defined a set of suitable views, each one focusing on one or more characterizing attributes. Moreover, for every entity, we defined a group of custom filters that allow the user to refine the corpus and narrow the set of entities being shown in the current view.

The main advantage of this approach is that it enables efficient and flexible navigation throughout the corpus, allowing the user to see the same entity from different

points of view. This way, data can be analyzed in several contexts, and patterns concerning specific characteristics of an entity can be easily discovered. This is especially useful for expert users who may experience a new methodology for researching a set of archives, giving them a powerful tool for analyzing the distribution of information [9]. For a common or occasional user, *Lomen* represents an innovative and simple approach to exploring the corpus and looking for a specific datum.

Views

As stated above, we defined a set of suitable views for each entity in order to expand the exploration modalities. Each view accesses one kind of entity at a time (with the exception of the actors network), providing visual features and interactions specifically designed for them. Each layout is designed to be robust in case some elements lack one or more properties.

Views for projects

Concerning the projects, we developed four different visualization models in order to cover the complexity of the entity:

- *List*: This model offers a simple, general and efficient way of exploring the projects in the archive, providing a small thumbnail of a related document and a brief textual description.



Figure 1. The list view

- *Gallery*: The gallery view follows a visually driven approach to project exploration, made of a grid of stacked items where every cell is occupied by a thumbnail of a project and its title. Arranged in chronological order, the grid view allows the user to see at a glance the different styles used by the architect in his projects, find similarities and contrasts or look for a work characterized by a particular visual style. Such a view is based on the collection montage discussed in [7], and was introduced to the project as a tool for visualizing the portion of the collection selected by the users in an intuitive and comprehensive way, enabling direct comparison among images and avoiding over-complicated visualizations. By focusing on the chronological dimension, we offer a clear overview of the dataset

and the possibility of directly comparing a large quantity of images, exploiting the space-efficient placement of the items.

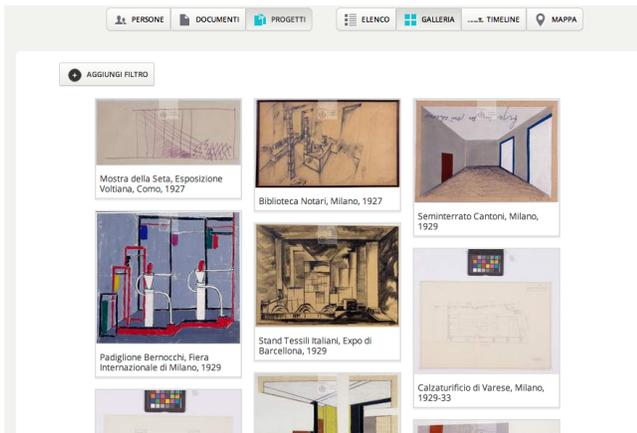


Figure 2. The gallery view

- **Timeline:** The timeline view focuses on the chronological ordering of the projects: the story of Baldessari's documented works is represented as a horizontal scrollable panel in which every item is denoted by a marker and its reference dates. This visualization is crucial to analyzing the development of the architect's career over time and finding periods characterized by similar activities or project types.

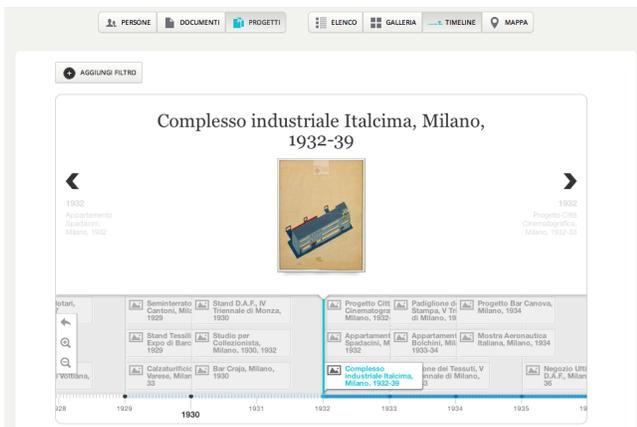


Figure 3. The timeline view

- **Map:** This view offers a geographic representation of project distribution. The user may interact with the map by zooming in and out and moving the region of interest. Depending on the zoom level, projects located in close areas are represented as clusters, which can be exploded and viewed as single items. This visual tool is useful for finding places and countries where Baldessari most frequently worked, or for discovering projects in uncommon localities with respect to the architect's life.

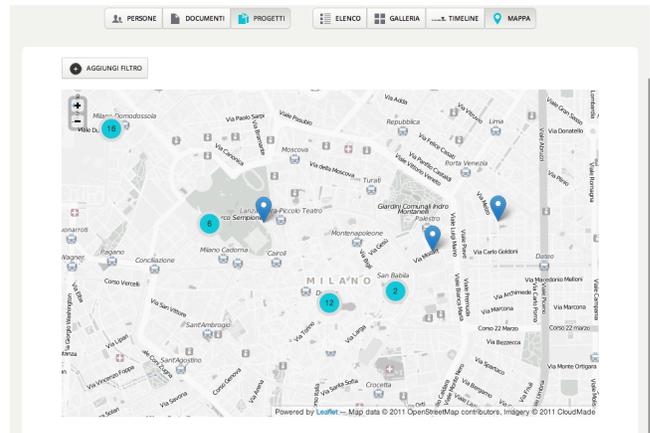


Figure 4. The map view

In any of the above-mentioned views, the user can select a single project and see its detail page, which describes its metadata in depth and offers a gallery of all the related documents.

Views for the documents

The documents entity is characterized more by its digital visual representation than its metadata. The main interest of the user is typically to access the image file of the document; hence, the view should follow a thumbnail-based layout. Starting from this assumption, we decided to use the list and gallery views defined for the project entity, mainly for their suitability when applied to this case. In particular, we redefined them as follows:

- **List:** Every item of the list is composed of a thumbnail of the document, followed by the related metadata: related project, signature, technique, description, scale, instrument, support and notes. Some of the metadata strings act as *quick filters*; e.g., clicking on the "pencil" technique, the view is refreshed showing only the document realized with a pencil.

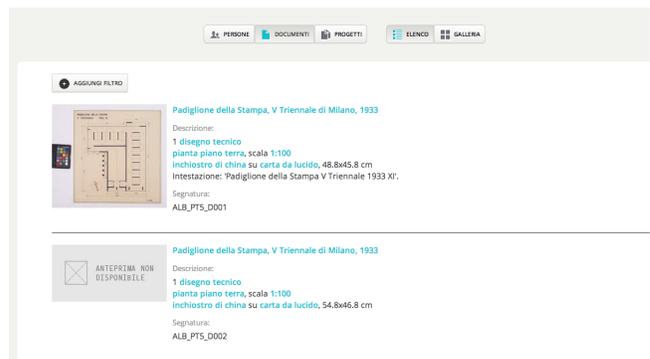


Figure 5. The list view applied to documents

- **Gallery:** A grid of stacked documents, each one showing a thumbnail, the signature of the item and its related project.

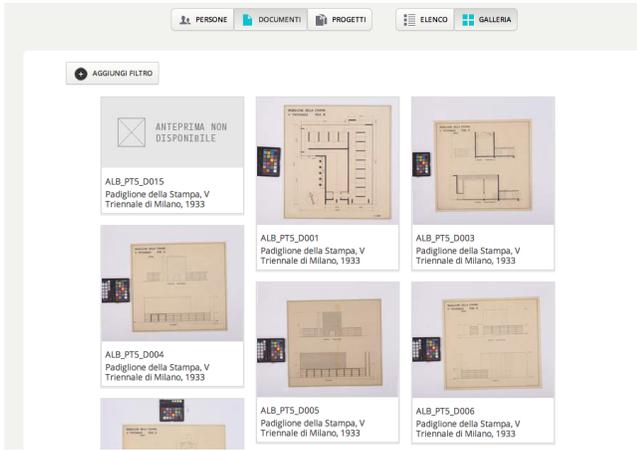


Figure 6. The gallery view applied to documents

Views for the actors

In the case of the actor entity, the user is mainly interested in the relationships between Baldessari and other people and organizations related to him and his works. Based on these assumptions, we implemented an interactive network graph where the nodes are divided into two categories: actors and projects. An actor node can only be connected to a project node, and vice versa. Each node size depends on the number of edges connected to it. The resulting visualization, developed with a force-directed drawing algorithm [3], allows the user to explore the collaborations of the architect, highlighting the most important relations and projects of his career.

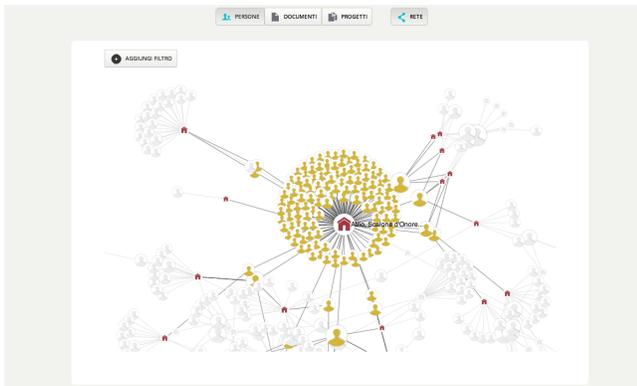


Figure 7. The actors network graph

The user can interact with the graph by zooming in and out, moving the center of the view and selecting a particular node. In the last case, only a subset of the relations is highlighted, starting from the clicked element and going through the connections up to the third level; i.e., if the user clicks on an actor, the highlighted items will be the starting actor a , the set $B = \{b_1, b_2, \dots, b_n\}$ composed of every project in which a is involved, and finally the set $C = \{c_1, c_2, \dots, c_n\}$ composed of every actor participating in a project in set B .

This visual tool enables a deep exploration of the relations: recognizing the most important projects and actors is immediate, and so is the discovery of small, isolated ones.

Filters

For every entity defined above, we built a set of related filters to narrow the search and dynamically create a subset of data. These filters can be stacked one above the other, resulting in a chain where every filter is applied consecutively performing a logical “AND” operation (e.g., a project view filtered by name and date will return entries containing the search string in the title and starting from the specified starting date). Moreover, filters can be dynamically added or removed without the need to reload the page. We defined three global filter typologies, each one suitable for a specific kind of data:

- *Search box*: This type of filter is intended for free text metadata such as titles, descriptions, etc. The filtering is implemented in an instant search fashion and is aided by autocomplete choices, which help the user find the exact term he or she is looking for.



Figure 8. The search box filter

- *Multiple-choice checkboxes*: This filter is designed for data belonging to a closed, narrow set. It is used, for example, to filter documents by their related activity (architecture, painting, sculpture). Every checkbox is added consecutively to the search performing a logical “OR” operation (e.g., if we check *architecture* and *painting*, the view will return all the documents related to architecture and/or painting).

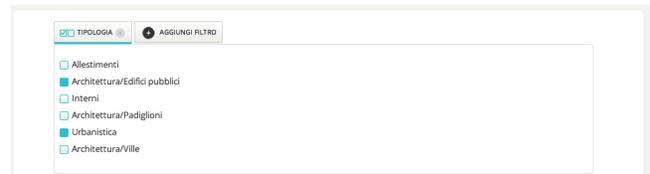


Figure 9. The multiple-choice filter

- *Sliders*: This filter type is used to select a range between two extremes, which may be dates, geographical coordinates, etc.



Figure 10. The sliders filter

The platform can be extended to use new kinds of filters, depending on the needs of the corpus to which it is applied.

As we defined different views for each entity, we thought of the filters as consistent elements throughout the

visualizations. In this way, the user can create a subset of the corpus related to a specific type of data and analyze it in different contexts. For instance, if we set a *sliders* filter on the projects map, we can see the same results in the timeline just by switching the view.

Thematic paths

As stated above, one of the main goals of the *Lomen* platform is to allow the user to extract new information starting from the initial data, to find interesting patterns and clusters of homogeneous entities and to create new valuable content.

In order to facilitate this process, the platform offers archive curators a tool for creating thematic paths based on the data present in the archives and shown in the views. Each thematic path is composed of an arbitrary number of steps, and each step is composed of a title, an image and a descriptive text. The steps are shown in chronological order on a timeline-based visualization similar to the related projects view, and the user may browse them freely by scrolling the horizontal panel.

Referencing entities

The entities previously defined can be referenced in the steps of the thematic paths by using a language similar to the wiki markup language used by Wikipedia. The format to be used follows the form `[[name to be shown | entity type: entity ID]]`. For example, a reference to the actor Marcello Nizzoli may be specified in the following way: `[[Marcello Nizzoli | person: 43]]`. This syntax is simple enough to be understood by users with no technical background, and may be familiar to students, researchers and expert users in general due to its similarity to Wikipedia's syntax. Moreover, since the links are qualified and refer to specific entities, it is possible to return this information visually by adding a small, text-sized icon just after the link.

Visual recaps

An additional advantage of using qualified links is the possibility of showing a visual summary of the current step in the thematic path. At the side of the descriptive text, a set of boxes shows the user all the entities cited on the page: projects, actors, documents and other themes. In the case of documents, instead of a textual reference, a thumbnail for each item is shown. This synthesis may be used to provide a context for the arguments discussed or to quickly find some interesting referenced items at a glance.

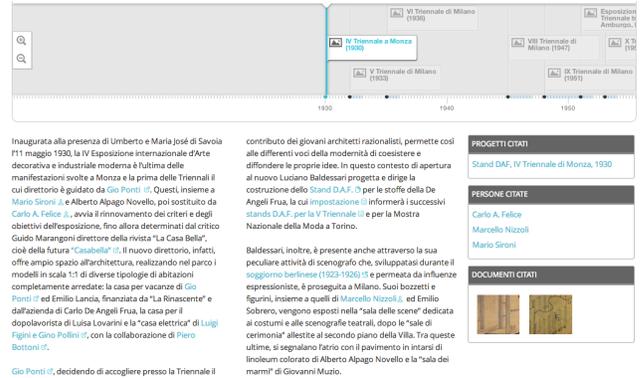


Figure 11. An example of a thematic path. On the right, it is possible to see the visual recap of lined entities

CONCLUSIONS

In this paper we presented a visual, exploratory interface for browsing archives with heterogeneous entities. Through the development of the *Lomen* framework, we have demonstrated the possibility of adopting a modular design process by creating dedicated views for each kind of entity.

The project served an experimental purpose, addressing both a general and expert public, where the latter is actively involved in the continuous design of the archive itself. The design process takes the characteristics of archival collections into account and characterizes them, adapting to different levels of accuracy in the metadata.

The application offers multiple perspectives of the archive corpus, each highlighting specific features of the represented items. Information visualization techniques have been exploited to create a non-intrusive visual interface, avoiding overly complex solutions and devoting interactions to expressing relationships among entities.

Finally, the application supplies a tool for archive curators to consolidate new levels of information. By creating thematic paths directly linked to the exploration interface, we provide a context for the archive data to users. Adopting an authorial approach to thematic path creation, the application becomes a tool for supporting new studies and their divulgation.

REFERENCES

1. Azzi, M., Caviglia, G., Ricci, D., Ciuccarelli, P., Bontempi, L. & Bonetti, E. Dust: A Visualization Tool Supporting Parents' School Choice Evaluation Process in *Parsons Journal for Information Mapping*, Vol. 3 n° 4. 2010.
2. Baldonado, M.Q.W., Woodruff, A. and Kuchinsky, A. Guidelines for using multiple views in information visualization. In *Proceedings of the working conference on Advanced visual interfaces*. ACM, 2000.
3. Bostock, M., Ogievetsky, V., & Heer, J. D³ Data-Driven Documents. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12), (2011) 2301-2309.
4. Craft, B., & Cairns, P. Beyond guidelines: what can we learn from the visual information seeking mantra?. In *Proceedings. Ninth International Conference on Information Visualisation* (2005) (pp. 110-118). IEEE.
5. Hinton, S., & Whitelaw, M. Exploring the digital commons: an approach to the visualisation of large heritage datasets. *Electronic visualisation and the arts (EVA 2010)*. 2010.
6. Kramer-Smyth, J., Nishigaki, M., and Anglade, T. ArchivesZ: Visualizing Archival Collections. 2007.
7. Manovich, L. Museum Without Walls, Art History Without Names: Visualization Methods for Humanities and Media Studies. *Oxford Handbook of Sound and Image in Digital Media*, ed. Carol Vernallis (Oxford University Press, in press). 2012.
8. North, C, & Shneiderman B. Snap-together visualization: a user interface for coordinating visualizations via relational schemata. *Proceedings of the working conference on Advanced visual interfaces*. ACM, 2000.
9. Pike, W. A., Stasko, J., Chang, R., & O'Connell, T. A. "The science of interaction". *Information Visualization*, 8(4), (2009) 263-274
10. Ricci, D. Seeing what they are saying: Diagrams for socio-technical controversies. *Proceedings of DRS 2010*.
11. Roberts, J. C. State of the art: Coordinated & multiple views in exploratory visualization. *Coordinated and Multiple Views in Exploratory Visualization, 2007. CMV'07. Fifth International Conference on*. IEEE, 2007.
12. Schön, D. A. *The Reflective Practitioner: How Professionals Think In Action*. Aldershot: Avebury, 1983
13. Shneiderman, B. The eyes have it: A task by data type taxonomy for information visualizations. In *Visual Languages, 1996. Proceedings., IEEE Symposium on* (1996) (pp. 336-343). IEEE.
14. Smith, A. Why digitize?. *Microform & imaging review* 28.4 (1999): 110-119.
15. Urban, R. J., Adamczyk P. & Twidale M. B. Cultural Heritage Information Dashboards. (2010).
16. Whitelaw, M. Towards Generous Interfaces for Archival Collections *International Council on Archives Congress*, 20-24 August 2012, Brisbane, Australia (August 2012)