



Engineering Annotations to Support Analytical Provenance in Visual Exploration Processes

Maroua Tikat, Aline Menin, Michel Buffa, Marco Winckler

► To cite this version:

Maroua Tikat, Aline Menin, Michel Buffa, Marco Winckler. Engineering Annotations to Support Analytical Provenance in Visual Exploration Processes. ICWE 2022 - 22nd International Conference of Web Engineering, Jul 2022, Bari, Italy. pp.1-16, 10.1007/978-3-031-09917-5_14 . hal-03779349

HAL Id: hal-03779349

<https://hal.science/hal-03779349>

Submitted on 16 Sep 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Engineering Annotations to Support Analytical Provenance in Visual Exploration Processes

Maroua Tikat^[0000–0002–7989–5518], Aline Menin^[0000–0002–9345–3994], Michel Buffa^[0000–0003–1900–0871], and Marco Winckler^[0000–0002–0756–6934]

University Côte d’Azur, CNRS, Inria, I3S (UMR 7271) France
[maroua.tikat, aline.menin, michel.buffa, marco.winckler]@inria.fr

Abstract. This paper focuses on the fundamental role played by annotations to support provenance analysis in visual exploration processes of large datasets. Particularly, we investigate the use of annotations during the visual exploration of semantic datasets assisted by chained visualization techniques. In this paper, we identify three potential uses of annotations: (i) documenting findings (including errors in the dataset), (ii) supporting collaborative reasoning among teammates, and (iii) analysing provenance during the exploratory process. To demonstrate the feasibility of our approach, we implemented it as a tool support, while illustrating its usage and effectiveness through a series of use case scenarios. We identify the attributes and meta-data that describe the dependencies between annotations and visual representations, and we illustrate these dependencies through a domain-specific model.

Keywords: annotations · provenance analysis · visual exploration · information visualization · data quality.

1 Introduction

The amount and complexity of digital data has increased exponentially during the last couple of decades. These data contain valuable information to support decision-making processes in several application domains. Nevertheless, the value of these data depends on the ability of decision makers to find relevant information that describes the phenomenon embedded in data. In this context, visual analytics is a suitable and widely adopted solution as it supports human reasoning through interactive tools that embed visual representations to highlight and reveal any relationships within data. Some examples of widely adopted tools for data analytic processes are kibana [9] and Tableau [20]. Even with the help of visual analytics tools, the exploration and analysis of large data sets is not a straightforward process. Although, visual analytics helps to show the tendencies and patterns within the data, data analysts must be able to interpret the findings in order to produce decisions. Moreover, it is not unusual that, during the exploration process, a data analyst is confronted with many sorts of errors (e.g., missing, duplicated, inconsistent data) [11], which should be fixed in order to

complete the analysis. In this context, it becomes part of the data analyst’s duties the task of checking the integrity and validity of data. For this reason, some authors [4] suggests that “the expertise to analyse and make informed decisions about these information-rich datasets is often best accomplished by a team”. In fact, most of real-world analytical environments require the participation of multiple analysts of different backgrounds and that play different roles during the analytical process [13].

Annotations are a suitable solution to record and process design decisions made by teams as they can compliment existing information with a rationale. Previous works [1, 12, 17] have shown that annotations are useful as a means to disclosure hidden relationships between data, to record the results of discussion and decisions made by team members, gather internal and external feedback on artefacts produced, to connect pieces of information such as results of usability evaluations and the design artefacts, to document and to justify design choices by describing them retrospectively.

Nonetheless, the ISO standard 9241-210 is very silent about how to record and process design decisions. To the best of our knowledge, there is no study investigating the use of annotations to record the visual exploration process of data. Inspired by these previous works we investigate the use of annotations to support decision-making processes through the visual exploration. We show that annotations can support analytical provenance studies by allowing data analysts to reason with various evidences collected during the exploration process and also trace back the results to the source of findings [15, 21]. Despite the concept of annotation being simple to understand, implementing has many implications on the design and development of tools. In this paper, we focus on annotations that can be embedded with provenance data, thus, supporting the tracking of use context of annotations, while supporting collaborative tasks around annotations. Our approach is implemented as a plug-in embedded into a visualization tool called MGExplorer [15]. The usage of our tools are illustrated through a set of use case scenarios describing the exploration of the Wasabi dataset [2], a large dataset describing music data.

The remainder of this document is organized as follows. The section 2 presents the concepts of annotation and provenance analysis. The 3 provides an overview of similar works highlighting the lack of studies investigating the use of annotations in visual exploratory processes. The section 4 introduces the rationale of our approach and presents a proposal for extending the W3C annotation model to support the idiosyncrasies of exploration process using visualization tools. The section 5 describes a set of relevant use case scenarios (including annotation of multiple types of views and objects, and collaborative use of annotations) that demonstrate the utility of our approach. Finally, section 6 presents our conclusions and future work.

2 Foundations

In this section we revise two key-concepts for understanding our approach: annotations and analytical provenance.

2.1 Annotations

The first studies on annotations were focused on paper textbooks and then transposed to electronic documents [14]. Annotating allows the interaction between distinct pieces of information to serve different purposes: description (placing data in a context, add sources of information, etc.), evaluation (reporting quality issues, questions or concerns, etc.) or a combination of both.

According to the W3C Recommendation, an annotation is defined as a set of connected resources featuring a body and a target that are interrelated. This approach supports interoperability as it allows annotations to be shared across different systems. The Figure 1 shows the W3C’s annotation model, where the target corresponds to the element we want to annotate, and the content of the body would usually regard the target. The annotation might contain meta-data that contextualizes the body’s contents. Annotations can assume many forms (e.g., text, sketching, highlighting, etc.) and be attached to different artefacts (e.g., documents, images, data sets, etc.) [7], which are not covered by the W3C specification and must be extended according to the application domain.

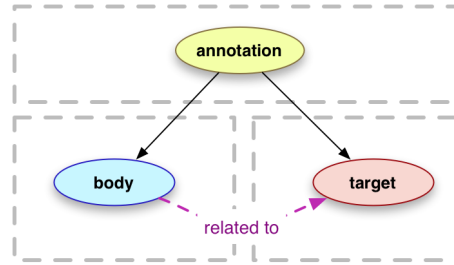


Fig. 1. The annotation model proposed by the W3C Recommendation.

2.2 Analytical provenance

Analytical provenance is a means to understand users’ reasoning processes while exploring data through visualization [18]. It studies users’ interactions while using a visualization system to extract exploration patterns that can explain how users explore data visually. The outcomes of analytical provenance can assist the evaluation of systems and algorithms, building adaptive systems, model steering, replicating, reporting and storytelling [23].

In this work, we are particularly interested in provenance data of visualization, describing the history of graphical views and visualization states [19], which can assist users on recreating their analytical reasoning process, while supporting verification, replication, reapplication, and sharing of exploration paths. In visualization systems, these data are often represented through history trees as in VisTrails [3] which allows users to create, change, and compare visualizations by exploring the graphical representation of the exploration flow.

Typically, provenance data is visually encoded as a sequence of actions indicating users' interactions [23]. The data is then encoded as a graph where nodes are entities (or concepts) that change state during the exploration process connected through line segments that indicates their previous state. Graph encoding allows the user to navigate in the exploration path, while interacting directly with the history graph to generate a story of their analysis [6]. Tools such as MGExplorer [15] and GraphTrail [5] represent the workflow through segment lines connecting different views of data, allowing users to understand the actions performed from one view to another.

3 Related Work

Annotations could be used on different types of data, be it to illustrate a drawing in order to provide an explanation, or using engineering annotations to support the design process of interactive systems. They could also support decision-making processes by being used as decision cards [12]. As shown in [16], there are multiple tools available to annotate documents, images or textual data. In this work, we focus on annotating result sets from queries, expressed and explored through visualizations. Furthermore, we should be able to support recording of users' findings during an exploration process, as well as sharing of annotations with fellow data analysts to support collaborative data analysis. Hereafter, we present some of the existing annotation tools in the literature.

The SenseMap [17] tool supports browser-based online sensemaking through analytic provenance. The tool provide data curation to indicate relevance of nodes (views in the graph) through an interactive annotation process: : if node is completely irrelevant, the user can remove it ; if a node is not quite relevant but the user wants to keep it to have a look at some point, they can minimize it ; if a node is very relevant, the user can favor it.

The Glozz [22] environment for annotation and exploration of a corpus is based on a generic model that can conform (by instantiation) to any annotation paradigm. The tool allows the manual annotation of texts that may have been previously annotated, as well as the annotation and visualization of simple or complex structures (units, relations and patterns (or clusters)). It also allows the exploration of annotations.

The UAM corpus tool [12] is a software for corpus annotation and exploration that allows to annotate text and images.

The GATE (General Architecture for Text Engineering) [8] is an open-source infrastructure that provides a set of language engineering tools for collaborative corpus annotation.

The ANALEC [10] tool combines corpus annotation, visualization and query management. It allows ergonomic annotation via the concept of view and the use of elaborate filtering of the available information, frequency calculations, search for correlations, and generation of tables, figures and diagrams.

Our approach differs from previous works by allowing an annotation to be linked to the data represented in a view.

4 Our Approach

Our ultimate goal is to allow users to create annotations for recording decisions during the visual exploration process of data. We assume that annotations should formalize the relationship between the users' intentions (ex. insights, questions, etc.) and actual artefacts of the visualization (those being annotated), such as a particular visual representation of the data. In order to accomplish this goal, we followed four steps:

1. **Create an annotation model** describes the concepts covered by an annotation and the procedure to use the model together with visualization tools. Subsection 4.1 presents our proposed model.
2. **Define the dependencies** between the annotations and the visualizations being annotated. As described in subsection 4.2, the dependencies might refer to the scope and the diverse elements in the display (e.g., views, queries, set of itemsets, etc.).
3. **Identify the attributes of provenance data** that allows to store and restore annotations in the context, as shown in subsection 4.3
4. **Development of tool support** that implement annotations in the context of visualization. Our implementation of the annotation model is presented in section 5 along with a set of use case scenarios to demonstrate the usage and feasibility of our tool.

4.1 Extending the conceptual model of annotations

Visualization techniques can be generalized as complex artefacts comprising a **query**, which serve to fetch data from the data set, a **dashboard** that can host one or more **views** (each view featuring a visualization technique) that are composed of a subset of **objects** making reference to a particular itemset in the data set. This general definition can be applied to any visualization tool. It defines a scope where annotations created by the users can be connected to visual representation of data. The section 5 shows different use case scenarios where annotations require a connection to views, objects, dashboard, and queries.

We also assume that an annotation might be a follow-up comment to a preexisting **annotation**, so an annotation can be annotated. It might also be

the case that an annotation is not connected to anything in particular, or be left over to be connected to an target later on by the user.

In order to accommodate all these scenarios, we have extended the W3C annotation's model, as shown in Fig. 2. First of all, we create a new class called *Artefact* that is used to refine and explicitly describe any idiosyncrasies of what is annotated, i.e. an **information visualization**. Such a class allows us to differentiate annotations created on other artefacts (ex. documents, drawings, etc). The other extensions inherit from the class *Target* and refer to the type of **selector** that can be used to annotate the different elements.

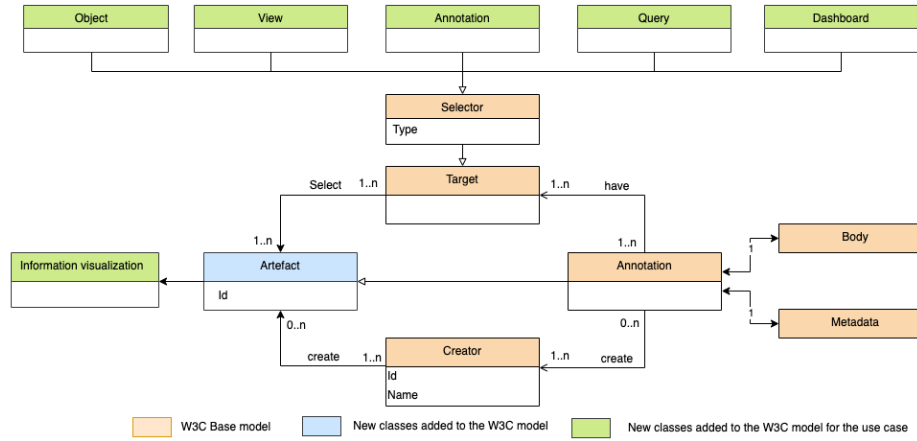


Fig. 2. Extended W3C annotation's model

4.2 Dependencies between annotation and visualizations

The annotation model described in the previous sections allow to describe annotations as singleton class *mge-annotation*, as shown in Fig. 3. The class *mge-annotation* is not connected to any other element, which grants independence of implementation of the approach from a particular visualization tool. The other classes at Fig. 3 refer to the model used to describe the components of the tool MGExplorer that is used to illustrate the approach. Fig. 4 shows the mapping between the singleton class *mge-annotation* and the other classes of MGExplorer.

It is worthy of note the connections between the class *view-annotation* and the sub-set of data a view might contain, handled by the class *object-annotaiton*, and the class *InfoVis technique* that refers to the actual visualization technique in the display. Further illustrations on how these connections are made at the tool level are illustrated in the section 5.

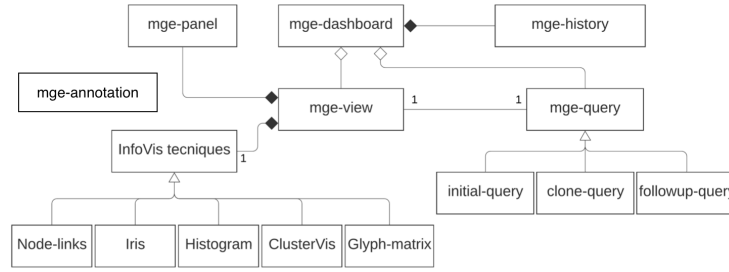


Fig. 3. MGExplorer model

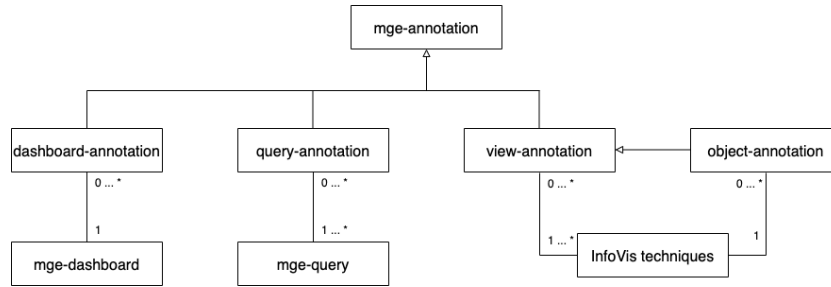


Fig. 4. Annotation targets available in MGExplorer

4.3 Attributes in data provenance

An annotation has a body and it also might have a meta-data. Whilst some of the body's contents is given by the users (ex. comments), other body attributes can be automatically collected from the context. Tab. 1 provides an exhaustive view of body attributes that can be captured when connecting different types of the target. These attributes are available through the classes in the mapping model shown (but not detailed) in Fig. 4. For example, when the user created a free annotation (i.e. target **None** in Tab. 1), we can only capture its *id*, *body* provided by the user, and the *timestamp*. However, when a user decides to connect the annotation to a target such as *Object*, it is possible to include other attributes automatically such as *views' title* (which describe the name of the window holding the data item), the *view's visualization technique* (which refers to the particular visualization being used to show data in the view), the *object* or the set of *views' subset* (corresponding to data items in the view). Such a combination of different attributes allows to determine the full context for data in the display and hence the data provenance.

5 Use Case Scenarios and Tool Support

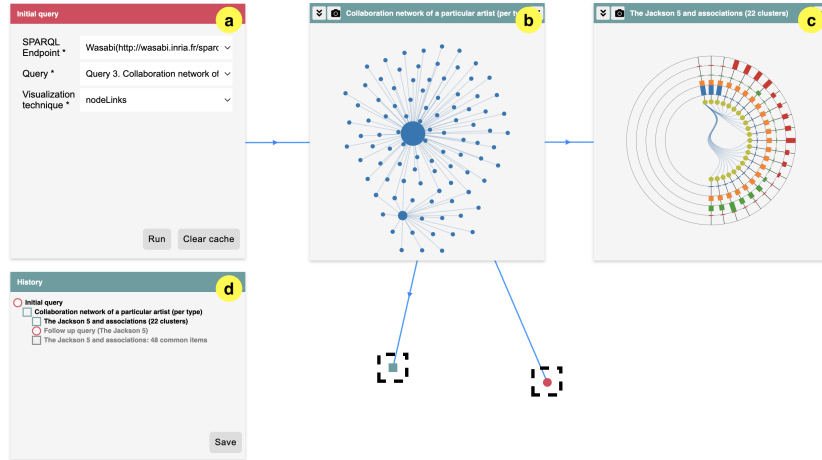
The use case scenarios presented hereafter, demonstrate the use of annotations when exploring the WASABI [2] data set, a SPARQL endpoint which consists

Table 1. Annotation attributes

Body attributes	Targets				
	Object	View	Query	Dashboard	None
Id	x	x	x	x	x
View's title	x	x	x		
View's visualisation technique	x	x			
Object	x				
Object's type	x				
View's subset	x	x	x	x	
Body	x	x	x	x	x
Timestamp	x	x	x	x	x
Path sequence				x	

of more than 2 million commercial songs retrieved from multiple sources on the web. In our scenarios we employ two queries. The first query retrieves data of a collaboration network of a particular artist described by type (producer, writer, performer) of collaboration. The second query retrieves data describing a network of artists by the genre of their productions.

The approach is implemented as a plugin for the tool MGExplorer [15]. This tool uses chained views to assist the exploration of multidimensional and multivariate graphs, which allows to depict analytical provenance via a sequence of views connected through line segments to represent their dependency, while supporting one or more visualization techniques applied to one or more datasets. Figure 5 depicts the exploration process as follows:

**Fig. 5.** Visualization tool

1. From a query panel, the user selects a SPARQL endpoint and executes a predefined query that retrieves a network data set describing a particular phenomena (co-authorship, co-occurrence, etc) (Figure 5a).
2. The resulting data set is visualized through a node-link diagram featuring a network (Figure 5b). The nodes are interactive allowing to create new views (five techniques are available: network, clusters, pairwise relationships, temporal distribution, and listing of items) from data subsets (Figure 5c). This subsetting operation is available in all visualization techniques, allowing the user to further explore the data set through different perspectives.
3. Upon each subsetting operation, a new view is created and linked to the previous view through a line segment. The provenance data regarding this operation is automatically included in a history panel (Figure 5d).
4. The views can be moved around, allowing the user to rearrange the visualization space in meaningful ways. Further, users can hide any of the currently displayed views (Figure 5e), which they may revisit later using the history panel, thus cleaning the display area in a way that helps them to focus on what is relevant to the task at hand.
5. The user can import new data by adding query panels, which resulting data can be explored seamlessly as the initial data set.

For the sake of coherence and to support analytical provenance, the annotation technique was implemented as a view that can be instantiated anytime and then connected to a view, an object, or the dashboard through line segments, or yet represented as a singleton when the annotation is free of context.

5.1 Overview of annotations

While visually exploring the collaboration network of the artist *Michael Jackson* (Figure 6a), we noticed a collaboration between him and *Justin Timberlake*. We decided then to further investigate it by displaying the list of songs that define this collaboration (Figure 6b). We observe that the song on which these artists collaborated was released in 2014, which collaboration could not be possible since *Michael Jackson* died in 2009. To report this issue, we create a free annotation (Figure 6c), where we indicate that “*Michael Jackson died in 2009*”.

We continue inspecting other collaborations of this same artist. We notice in the node-link diagram, a collaboration between Michael Jackson and Pitbull. The list of (Figure 6d) shows that they worked together on the song “Bad” released in 2007. However, this cannot be true; Pitbull made a remix of the song “Bad” in 2012 but he never collaborated with Michael Jackson. To report the problem, we will annotate the whole exploration process by creating an annotation on the dashboard (Figure 6e) indicating a “*Problem with Michael Jackson’s collaborations in the dataset. Pitbull did a remix of “Bad” in 2012*”. This annotation is not linked to any particular element of the dashboard, but it rather retrieves the history (provenance data) of exploration.

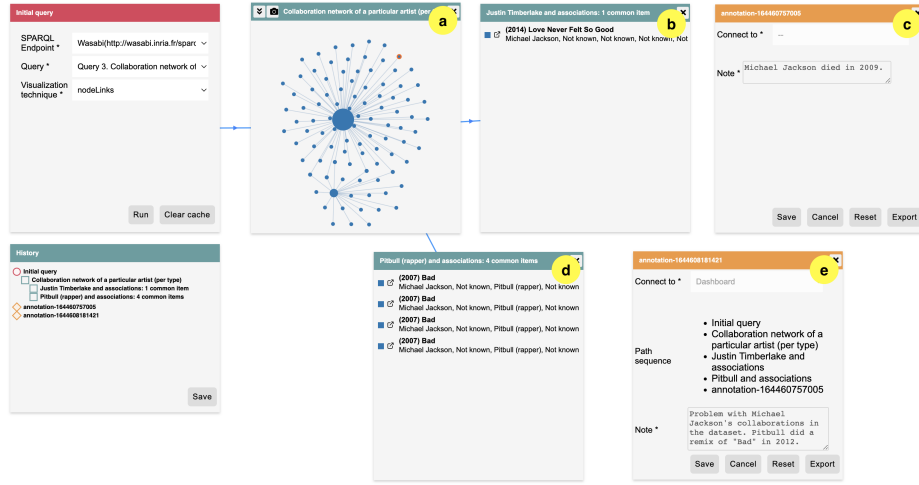


Fig. 6. Overview of annotations: choosing annotation’s targets

5.2 Managing multiple annotations

When annotations are not connected to a particular item, we have to go through the whole data set to find the actual item causing issues. Alternatively, we can connect the annotation to the view causing the issue; as in Figure 7.a featuring a node-edge connected to the annotation “*Error: Pitbull didn’t collaborate with Michael Jackson in 2007. He performed in a remix of the song ”Bad” in 2012.*”. This issue refers to two views in the display and we can connect both views to the annotation to reinforce our concerns. For that, by selecting those two views in the annotation window as shown at Figure 7b, we create an annotation that connects to the two views through two line segments. Then we create a new annotation to report a duplicated song in the list of songs. Along the exploration process, the dashboard might become full of views. Thus, to reduce visual cluttering, we close this last annotation (Figure 7e) which nonetheless remains in the history panel allowing us to retrace the exploration path and reopen the view if needed.

To analyze the evolution of members in the group “The Jackson 5”, we try a new query. We try to launch a new query from the node-link diagram (Figure 7d) but we notice that query does not exist. Since this analysis is important for us, we create an annotation “*I need a query to explore the temporal evolution of the members of a group*” connecting the query view (Fig. 7e).

5.3 Annotating objects in a view

Our approach also allows the user to annotate a specific item on a view. In this scenario, we demonstrate how to retrieve a data item causing issues with respect to the two collaborations of Michael Jackson seen in the previous scenarios.

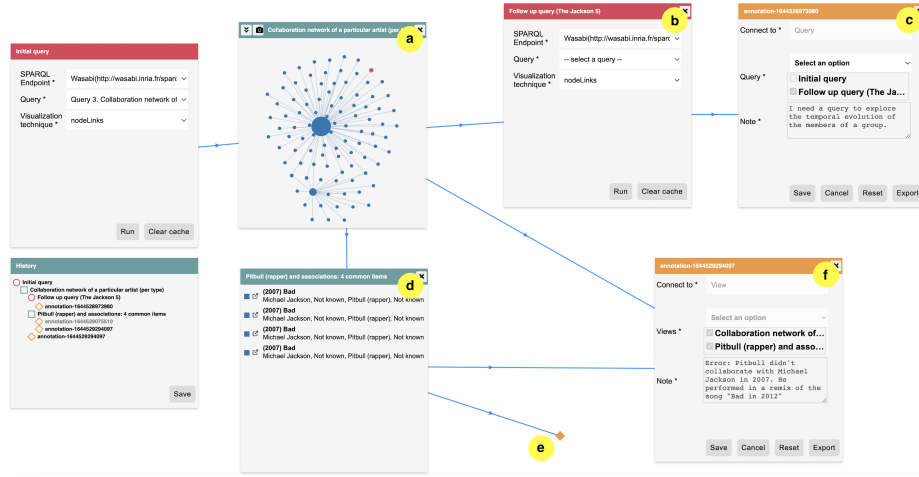


Fig. 7. Managing multiple annotations

When exploring the node-link diagram, we first create an annotation and connect it to both links indicating the collaborations of Michael Jackson with Justin Timberlake and with the rapper Pitbull. As shown by Figure 8a, the annotation view displays a description of the objects and it is linked to the view where the objects appear through a line segment. This approach works in every visualization technique. For instance, we further explore the collaboration between Michael Jackson and Pitbull using a different visualization technique, the ClusterViz. Here, we can create an annotation and link to the objects representing this collaboration (Figure 8b).

5.4 Exporting annotations

To export the data describing the annotations, we can use the button at the bottom of the annotation view (Figure 9b). This action produces a *json* file (Figure 9a), containing the attributes of the annotation, as shown in Table 1 and a link that allows the user to reopen the graphical annotation view.

5.5 Collaborative use of annotations

In this scenario, we demonstrate how annotations can be shared among teammates in order to support collaborative analysis of data sets. For that, we explore a data set describing the relationship network of artists based on the genre of their productions. We analyse the relationships using the node-link diagram (Figure 10a), using the search bar under the visualization technique to locate the artist “Madonna” by. We found that Madonna produced songs of the same genre as other four artists. We use the visualization technique called IRIS (Figure 10b) to identify what genres they have in common. The width of colored bars in IRIS

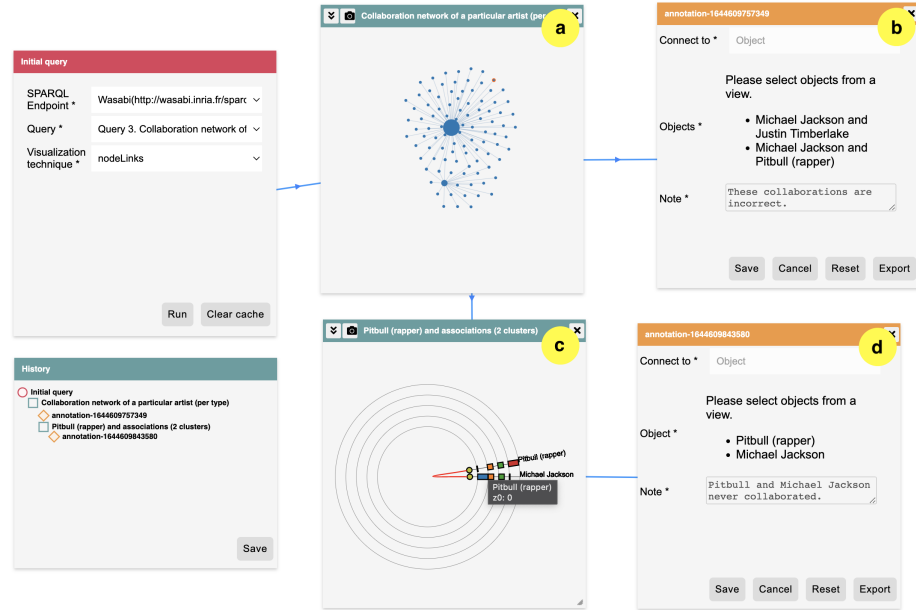


Fig. 8. Annotation of objects

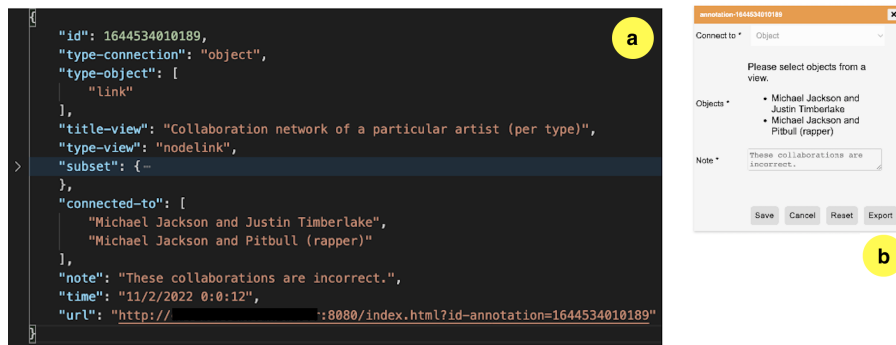


Fig. 9. Exporting annotations

represent the number of songs where Madonna have a genre in common with every other artist (in the periphery). By studying this chart, we notice an issue with the visualization: all bars have the same height and colors. However, as we hover over the bar between Madonna and one of the artists, the information box that appears shows that there is no song under the genre “dance” between those artists. We create an annotation on the dashboard to raise the issue (Figure 10d) stating that *“Genres with no data shouldn’t appear in the visualizations”*. Our colleague, *John Doe* could disambiguate or fix the issue. For that, our tool provides a shareable link of the dashboard, so that *John Doe* can open on their own browser and continue the exploration. When *John Doe* finds an answer to the issue, he replies by creating a new annotation *“It’s a problem related to the system and not the dataset. It’s fixed now.”* which it connected to our the initial annotation raising the question (Figure 10c).

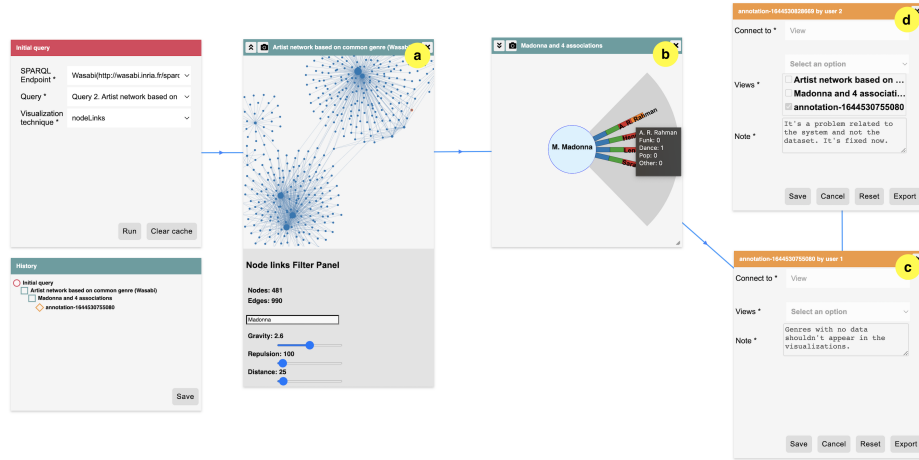


Fig. 10. Annotation sharing: collaborative use of annotations

5.6 Tracking provenance analysis

From the previous scenario, the user *John Doe* has an overview of the whole schema produced by us, including the data set, the query and the annotation created during the exploratory stages. This allows *John Doe* to track the path explored and retrieve the origin of the annotated data.

6 Discussion, conclusions and future work

Annotation is a concept familiar on paper but it is often hard to implement and seldom fully exploited, beyond tools for annotating digital document. At

the best of our knowledge, we could not find methods for annotating individual elements during the visual exploration of data sets. In this paper we discuss some of the problems for engineering annotations for supporting representation of provenance when exploring large data sets. On one hand, our approach allows to encode information about data provenance (i.e. sequence of data exploration) as part of the annotations, so when analyzing annotations it is possible to restore the full sequence of actions allowing users to go from queries to data in the display. On the other hand, annotations should help users to explain insights and decisions, which is necessary to perform provenance analysis on data. The validation of the approach is made by construction, which means that we demonstrate its feasibility by creating a tool and illustrating its use by a set of relevant scenarios.

Through the scenarios, we have shown how annotations can be used to provide information to complete the data set, point errors and mistakes, express user's needs, compare results, and share insights with other users. These examples are not exhaustive but yet representative for the use of annotations. We also have demonstrated how annotations can be extracted from the tools and yet include subsets of data and sequence of actions, thus become available to support analysis of provenance using other tools. It is interesting to notice that users might report issues with the data set using different combinations of annotations and views.

The originality of this paper concerns aspects for engineering annotation beyond digital document. As discussed here, we need to consider the specific context of the use of annotations (data sets and visualizations) for building the appropriate tools. We introduce a new topic and we want to level the discussion about the importance of having annotations to support the exploration process; for that we need innovative tools including the analysis of the provenance. Whilst the plug-in developed is specific to the visualization tool MGExplorer, we consider that the steps described in the approach are generic enough and can be adapted to support the development of other annotating tools.

We also demonstrated how to share annotations with other teammates to discuss findings. We want to advertise our tool to collect data from real users, not only for assessing the usability of the tool but also for collecting data enabling further investigations about analysis of provenance. In the future, we also consider to extend the possibility for including other annotation formats thus allowing to create annotations using highlighting of elements, support drawings, symbols, speech as replacement of text. Other improvements involve the synchronous edition and annotations.

References

1. Agosti, M., Ferro, N., Orio, N.: Annotations as a tool for disclosing hidden relationships between illuminated manuscripts. Springer-Verlag (2007)
2. Buff, M., al: The WASABI dataset: cultural, lyrics and audio analysis metadata about 2 million popular commercially released songs. In: ESWC 2021: The Semantic Web (2021)

3. Callahan, S.P., Freire, J., Santos, E., Scheidegger, C.E., Silva, C.T., Vo, H.T.: Vistrails: Visualization meets data management. In: Proceedings of the 2006 ACM SIGMOD International Conference on Management of Data. p. 745–747. SIGMOD '06, Association for Computing Machinery, New York, NY, USA (2006). <https://doi.org/10.1145/1142473.1142574>, <https://doi.org/10.1145/1142473.1142574>
4. Cernea, D.: User-Centered Collaborative Visualization. Ph.D. thesis, University of Kaiserslautern (2015), <http://kluedo.uni-kl.de/frontdoor/index/index/docId/4051>
5. Dunne, C., Henry Riche, N., Lee, B., Metoyer, R., Robertson, G.: Graphtrail: Analyzing large multivariate, heterogeneous networks while supporting exploration history. In: Proceedings of the SIGCHI conference on human factors in computing systems. pp. 1663–1672 (2012)
6. Gratzl, S., Lex, A., Gehlenborg, N., Cosgrove, N., Streit, M.: From visual exploration to storytelling and back again. *Comput. Graph. Forum* **35**(3), 491–500 (Jun 2016)
7. Hak, J., Winckler, M., Navarre, D.: PANDA: prototyping using annotation and decision analysis. In: Luyten, K., Palanque, P.A. (eds.) Proceedings of EICS 2016
8. H.Cunningham, V.Tablan, K.Bontcheva: Language engineering tools for collaborative corpus annotation. In: Proceedings of Corpus Linguistics (2003)
9. Kibana: (2022), <https://www.elastic.co/kibana>, (accessed January 2022)
10. Landragin, F., Poibeau, T., Victorri, B.: ANALEC: a new tool for the dynamic annotation of textual data. In: International Conference on Language Resources and Evaluation (LREC 2012) (2012)
11. Laranjeiro, N., Soydemir, S.N., Bernardino, J.: A Survey on Data Quality: Classifying Poor Data. In: Proceedings - 2015 IEEE 21st Pacific Rim International Symposium on Dependable Computing (2016)
12. Lopez, M.G., Rovelo, G., Haesen, M., Luyten, K., Coninx, K.: Capturing Design Decision Rationale with Decision Cards. In: IFIP Conference on Human-Computer Interaction (2017)
13. Madanagopal, K., Ragan, E.D., Benjamin, P.: Analytic provenance in practice: The role of provenance in real-world visualization and data analysis environments. *IEEE Computer Graphics and Applications* **39**(6), 30–45 (2019). <https://doi.org/10.1109/MCG.2019.2933419>
14. Marshall, C.C.: Annotation: From paper books to the digital library. In: Proceedings of the Second ACM International Conference on Digital Libraries. p. 131–140. DL '97, Association for Computing Machinery, New York, NY, USA (1997). <https://doi.org/10.1145/263690.263806>, <https://doi.org/10.1145/263690.263806>
15. Menin, A., Cava, R., Freitas, C.M.D.S., Corby, O., Winckler, M.: Towards a Visual Approach for Representing Analytical Provenance in Exploration Processes. In: 25th International Conference Information Visualisation (2021)
16. Neves, M., Ševa, J.: An extensive review of tools for manual annotation of documents. In: Briefings in bioinformatics 22.1 (2021)
17. Nguyen, P.H., Xu, K., Bardill, A., Salman, B., Herd, K., Wong, B.W.: Sensemap: Supporting browser-based online sensemaking through analytic provenance. In: 2016 IEEE Conference on Visual Analytics Science and Technology (VAST). pp. 91–100 (2016). <https://doi.org/10.1109/VAST.2016.7883515>
18. North, C., Chang, R., Endert, A., Dou, W., May, R., Pike, B., Fink, G.: Analytic provenance: process+ interaction+ insight. In: CHI Extended Abstracts Human Factors in Computing Systems, pp. 33–36 (2011)

19. Ragan, E.D., Endert, A., Sanyal, J., Chen, J.: Characterizing provenance in visualization and data analysis: an organizational framework of provenance types and purposes. *IEEE transactions on visualization and computer graphics* **22**(1), 31–40 (2015)
20. Tableau: (2022), <https://www.public.tableau.com/s>, (accessed January 2022)
21. Toniolo, A., Norman, T.J., Etuk, A., Cerutti, F., Ouyang, R.W., Srivastava, M., Oren, N., Dropps, T., Allen, J.A., Sullivan, P.: Supporting reasoning with different types of evidence in intelligence analysis. In: *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*. pp. 781–789 (2015)
22. Widlöcher, A., Mathet, Y.: The Glozz platform: a corpus annotation and mining tool. In: *DocEng '12: Proceedings of the 2012 ACM symposium on Document engineering* (2012)
23. Xu, K., Ottley, A., Walchshofer, C., Streit, M., Chang, R., Wenskovitch, J.: Survey on the analysis of user interactions and visualization provenance. *Computer Graphics Forum* **39**(3), 757–783 (2020)