Towards a Contextualized Visual Analysis of Heterogeneous Manufacturing Data

Mario Aehnelt[1], Hans-Jörg Schulz[2], and Bodo Urban[1]

[1] Fraunhofer IGD Rostock, Germany
{mario.aehnelt,bodo.urban}@igd-r.fraunhofer.de,
[2] University of Rostock, Germany
hjschulz@informatik.uni-rostock.de

Abstract. Visual analysis spanning multiple data sources usually requires the integration of multiple specialized applications to handle their heterogeneity. This is also true in manufacturing, where data about orders, personnel, workloads, maintenance, etc. must be analyzed together to make well-founded management decisions. Yet, the orchestration of multiple data sources and applications poses challenges to the software infrastructure and to the analyst. We present a three-tiered approach to cope with these challenges. In a first step, we establish a domain-dependent workflow as the mental model of the analyst. Based on the novel concept of contextualization, we then align the different applications with this model for their meaningful integration. In a third step, we incorporate the data according to its use in the aligned applications by means of a service-based architecture. By starting the integration on the user level, we are able to pragmatically target and streamline the required integration to a degree that is technically achievable and interactively manageable. We exemplify our approach with the Plant@Hand system for integrating manufacturing data and applications.

1 Introduction

Heterogeneous data, meaning data that stems from various data sources and that comes in a multiplicity of data formats, is far from a purely academic research challenge. Instead, it is an equally exciting and frustrating reality in many application domains. The excitement about heterogeneous data stems from the fact that bringing together diverse data sources for an integrated analysis and decision making yields more reliable and comprehensive insights than investigating individual data sources alone. Whereas the frustration stems from the realization that it is extremely hard to actually do just that – to "bring together" diverse data sources. One of the reasons for this dilemma is that there exist highly specialized domain-dependent visual and non-visual analysis tools that allow users to perform a certain set of operations for a particular kind of data. The domain of manufacturing is no exception to that, as for example, *enterprise resource planning* systems (ERP) are used to manage information about orders and personnel, while *manufacturing execution systems* (MES) are employed to collect and evaluate data about the production process. For an integrated analysis, not only the different data sources must be used in combination, but also the various applications that are used to access them. This results not only in a technical challenge, but also in a challenge to the human analyst who must interactively manage this diversity of data and tools to pursue a cross-dataset/cross-application analysis.

Thus, our work aims to combine multiple such applications in a way that allows for a fluid integrated analysis across applications and thus across the different data sources they operate on. To achieve this with reasonable effort on the technical level and on the user level, we concentrate the integration on only the necessary aspects by following a three-step approach:

1. We investigate the domain context and establish a **general workflow**, which can be assumed as a mental model of an analyst in this field.
2. We link the applications to the workflow through **contextualization**, which anchors the tools in the spot where they are used – e.g., the spreadsheet with machine data right on top of the CAD drawing of the machine itself.
3. We perform the **data integration** using *enterprise service bus* (ESB) technologies according to the contextualization of the applications, which gives us the knowledge about how they feed data into each other.

The remainder of this paper details the related work in Section 2. Based on that, it introduces our 3-step approach of workflow generation, application contextualization, and data integration in Section 3. This approach is then illustrated by the *Plant@Hand* system, a domain-specific implementation of our concepts for the concrete case of manufacturing in Section 4. Section 5 concludes this paper and states our ideas for future work.

2 Related work

For the challenge of integrating heterogeneous data, various solutions already exist. They focus on different aspects, addressing either the user level, the application level, or the data level.

On the **user level**, a fundamental problem of the multitude of data sources is the *information overload* challenging the user's cognitive capabilities. This problem aggravates the more data sources and thus the more applications handling them need to be integrated. Only recently, Landesberger et al. [1] identified the lack of *consistency* as one of the reasons for this, making it a crucial design problem for integrated analysis applications. On the one hand, this problem can be addressed by using the same basic visualization and interaction means, e.g., from a standard library, if the heterogeneity of data allows for it. On the other hand, consistency is also a question of the user's perception. Here we learn from research in *computer-supported cooperative work* (CSCW) that the user's mental model plays a growing role in designing interactive applications [2]. Each user has an individual small-scale model of certain aspects of reality, which influences his interaction with an application. This concept of *mental models* reaches back to 1943, when Kenneth Craik discussed the influence of human thinking on the perception of reality [3]. A common way to align multiple analysis applications and datasets with a mental model is to use the analysis workflow to be performed on the various datasets with the different analysis tools. For example, Streit et al. [4] align their visual representation of data sources with such a workflow to clearly convey what the users are looking at and how it relates to other data.

On the **application level**, bringing together different data sources by combining the visualizations and Visual Analytics (VA) applications, which are used to show and analyze them, is so far an unsolved engineering problem. The most prominent endeavor in this direction is the *Obvious* framework [5], which aims to define an interface standard for VA applications. Other approaches range from highly centralized architectures to highly flexible ones. An example for the former is the *Universal Visualization Platform* [6] that provides one core framework to which all applications are hooked via a plug-in mechanism. An example for the latter is the *Metadata Mapper* [7] that realizes a loose coupling of applications through a common communication bus. Furthermore, it is also possible to integrate applications based on their user interfaces, rather than their data or metadata. Various such approaches exist, such as, *virtualization* described by Besacier and Vernier [8], the *mash-ups* by Aehnelt [9], or *customized graphical user interfaces* introduced by Lee et al. [10].

On the **data level**, the integration of various data sources is a longstanding challenge in database research. Over the years, many different solutions have been developed to achieve such integration. Depending on the degree of integration, the developed solutions range from *federated databases* [11] for a rather loose linkage of independent data sources to *information fusion* [12] and *data warehouses* [13]. In addition, there exist also a variety of ways to perform the actual linking of data – e.g., based on *conceptual schemas* [14], based on logic [15], or based on ontologies [16].

It is noteworthy, that none of these approaches targets the integration of data sources on all three levels at the same time. Yet, this would be required to achieve a fluent integration that is easily understood and managed by a human analyst through the various applications he uses. This is where our solution comes in, which considers each of these levels in one of its steps, as it is described in the following section.

3 Integrating heterogeneous data and analysis tools

As opposed to other approaches for integrating heterogeneous data, we do not start on a technical data or application level, but on the user level to gain an understanding of the actually required integration before performing it. This permits us to anticipate and preserve domain-specific relations between data sources, while omitting others that are not relevant for the case at hand. Being confronted only with the amount of integration complexity that is necessary makes the integration manageable and understandable by the analyst and at the same time also technically achievable from an architectural standpoint. On top of that, the firm grounding of the application and data integration in the domain creates an extra level of application consistency, as data and applications behave according to the analyst's own understanding of his work reality. The three steps of our integration approach are detailed in the following.

3.1 Understanding the general workflow of the application domain

Initially, we have to understand the domain-specific application context and based on this the likely mental model of a data analyst. One possible context to establish is the analysis procedure that the analyst follows. Although, individual procedures can deviate in detail, a generalized workflow model can be used to capture it abstractly. Figure 1 shows such a generalized workflow that aims to model the data analysis process in manufacturing. In essence, the monitoring and control of manufacturing processes are mainly a periodical review of planned and reported figures. Deviations are to be further investigated in order to identify upcoming problems and implement adequate solutions promptly. On management level, *key performance indicators* (e.g., overall equipment effectiveness, cycle-time-ratio, delivery delay) give an overall impression of manufacturing performance. On operational level, the detailed production or assembly status in relation to production and work orders as well as resources are crucial for further decision making.

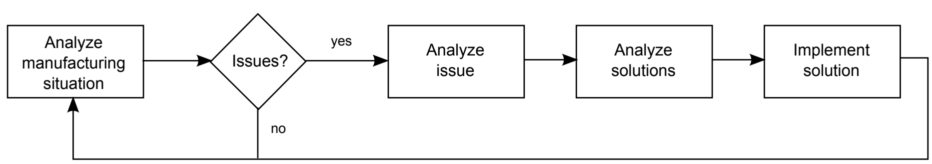

Fig. 1. Workflow model of data analysis and decision making in manufacturing monitoring and control.

The production supervisor – being the analyst of various manufacturing data – is responsible for a timely and quality-conscious delivery of products. In each analysis step, he works with different data and varying analysis tools. When analyzing the manufacturing situation, he compares planned figures against reported ones, e.g., the sequence of *production orders* (times, delays, outputs), the allocation of *resources* (material, machines, staff), or the *workload* on production work places (efficiency of machines and staff).

The data and their visualizations are in this case provided by ERP systems and MES. They are rarely connected, which complicates the joint analysis of data from both systems. When it comes to unplanned interruptions, e.g., machine breakdowns, missing part orders, or industrial injuries, again other tools are needed to view and analyze this particular data. Here, a growing amount of unstructured data (sales order details, supplier part catalogs, part specifications, machine manuals) has to be reviewed in close combination with structured data (contact details, production orders).

When using domain-specific analysis tools, users are used to familiar products and their representation of data. The manufacturing domain works with standard business applications that form the user's understanding and interpretation of data, e.g., *SAP Business One PPS*, *HYDRA-MES*, and *product data management* (PDM) products, such as *Autodesk AutoCAD* or *Dassault SolidWorks*. This often leads to a dual encoding of information that influences the user's own mental model of data, e.g., *parts* have a recognizable material number in ERP and MES, and at the same time a visual representative in form of a geometric CAD model. Thus, the production supervisor has both visual and abstract representatives of data in mind when analyzing bills of material. But, he only recognizes that the same part has been used multiple times in a single product based on its spatial location in the model and not by its serial number. This context knowledge facilitates the integration of the different applications to seamlessly perform the outlined general analysis workflow across application boundaries.

3.2 Contextualizing visualization and VA applications

From reviewing the application domain, we learned about the goals and structure of analysis work steps, as well as the data sources and the applications through which they are accessed and analyzed. Using a process called *contextualization*, this diverse information is then combined on the application level to form an ensemble of tools that reflects the assumed mental map of the analyst. Contextualization is a novel approach in Visual Analytics that enriches purely technical entities, such as software applications and data sources, with the context knowledge about their relations with each other, as they are derived from the application domain in the first step. In its simplest form, it can be used to capture the identity relations of the aforementioned dual encodings and thus to link their respective applications.

Yet, context knowledge can also be understood as a hierarchical construct which nests smaller sub-contexts (e.g., a particular machine) in a broader super-context (e.g., a factory building). The more confined the current context is (e.g., analysis of a single machine instead of the entire factory), the more specific is our knowledge and thus the easier it gets to derive a suitable application ensemble. Yet, as the context is reduced, the fewer data sources and data types still relate to that context. For example, work progress data and workforce data cannot be linked contextually to an individual machine, even though a drop in work progress may have a causal relation to a single broken machine. The same issue occurs for abstract data with no relation to the current sub-context at all. In these cases, we have to look on higher contextual levels for relations between such data. Thus, for example, workforce data (e.g., age or skills) refer to manufacturing teams responsible for specific tasks on a construction part. The abstract data of one sub-context can then be visualized close to the visualization of related data of another sub-context utilizing the relationship between both on a higher contextual level.

Arranging the application ensemble in this way aids the analyst's recognition and anticipation of known correlations, patterns, and representatives of reality. Basically, the data is shown in its associated applications where it is expected (e.g., a machine data visualization on top of the CAD drawing of that machine) and where it is needed according to the workflow. For this, Aehnelt et al. [2] already identified *context objects*, *metaphors and analogies*, as well as *interaction scripts* as suitable means to encode information and interact with it according to the user's mental model. This allows the user a clear interpretation and improves the efficiency of data analysis. For example, when the production supervisor is analyzing the planned and reported figures, he aims to identify deviations in time, quantity, and quality. In this example, the analysis application needs to visualize and highlight critical deviations using a visual encoding that matches the user's mental model. To also make the underlying context accessible and visible, the initial ensemble of overlaid views can be interactively rearranged by the user at any time during the analysis. In order to not overburden the application ensemble with views, *semantic and visual linking* [17] can be employed to shift loosely related applications to the periphery while at the same time maintaining and communicating their relations.

This approach of contextualization allows us to visually combine the familiar ERP, MES, and CAD representations embedded into a consistent user interface which adopts the virtualization or composition of analysis tools within a more generic framework. This framework applies the contextualization rules automatically, thus influencing the presentation and interaction with data according to the given domain knowledge, such as the currently pursued workflow and the available data sources.

3.3 Data integration with enterprise service bus architectures

Having established the application ensemble, we now know which data sources must be integrated to facilitate the analysis across application boundaries. In particular, data that is connected via identity relations (dual encodings) and data that is connected by referring to the same context (i.e., the same machine) are likely to be used in concert and should thus be linked on data level as well. To this end, we use ESB technologies that allow a flexible configuration and transfer of data between heterogeneous interfaces, data models, and applications. Within the ESB, an integrated data model derived from our workflow model (Step 1) maps the domain specific data types onto an abstract inter-application model (Step 2) of data. *Contextualized network graphs* [18] help us to automatically find relations and fill the integrated model with heterogeneous data. Thus, having streamlined the data integration to the concrete case at hand, we are able to concentrate on linking only the necessary data: work orders and resource allocations from ERP, time reports from MES, and 2D/3D product models from PDM with unstructured data from document sources.

4 Plant@Hand

Our approach of contextualizing the visual analysis of manufacturing data was implemented in the *Plant@Hand* system. It specifically addresses the monitoring and control of manufacturing and assembly works in shipbuilding industries. The system integrates different data analysis functionalities as they are usually provided by ERP, MES, and PDM systems individually within a consistent multi-touch user interface. Plant@Hand supports multiple users working collaboratively with different contextualized data representations. The three steps of our approach are realized in this system as follows.

4.1 Analysis workflow in Plant@Hand

The monitoring and control workflow (see Figure 1) addresses a supervisor's procedure of analyzing progress and figures at a shipbuilding site. Thus, we adopt his own mental model to design our integrated analysis application accordingly.

The analysis application Plant@Hand starts with an overview of the manufacturing and assembly situation combining the data from all required sub-contexts in representative views. Issues are immediately highlighted to simplify the first workflow step. The supervisor and his team can explore all available data for each issue individually using data specific views, e.g., work reports, part models, time schedules. Through interacting with these views, the next workflow step of analyzing adequate solutions is supported. By this means, the team can modify workplan data and get instant feedback on possible resource gaps or time conflicts. Once the supervisor has decided for a solution, it is implemented, meaning that data modifications are passed on to the responsible manufacturing software (ERP, MES, PDM) and its solution-dependent business logic.

4.2 Contextualized data representation and interaction

The main contextualization is based on the monitoring and control workflow as well. All required data is provided in separated views using familiar and established representatives for data.

The application is built on top of a visual representation of the main construction plan containing a 2D drawing of ship sections and installation details. Each user is then visually represented by an individual task view containing his own work orders and details. Further views visualize planned work orders and schedules, reported work results and issues, 3D construction models, and assembly tutorials. In our application, we additionally use a *spatial* and *temporal* contextualization of all available information to give any data a visual relation to product models and time schedules. Thus, analyzing a specific work order highlights its visual representatives in other views, e.g., the corresponding geometric section in the construction plan, the work orders' time schedule and staff allocation, or related work reports and issues (see Figure 2). This linking references the user's mental model of where the work is located, when it is due, which people are assigned to it, and which results were already reported.

Fig. 2. An assembly team working collaboratively with Plant@Hand on a multi-touch table (left). Spatial and temporal contextualization of work orders in Plant@Hand (right). An underlying CAD drawing is used to provide context to the individual views, which are positioned at the place they refer to, e.g., a particular machine.

Interaction with Plant@Hand bases on multi-touch gestures. Here we allow for data specific interactions that are based on the workflow. A re-planning of resources to solve work-related issues can be done visually by moving work orders with drag gestures between work order view and personalized task views. In a similar way the application supports an interactive visual data analysis for exploring work situations or reported problems. This requires a contextualized interaction with 2D and 3D construction models, as well as with work report documents. Geometric models can be moved using drag gestures, zoomed with pinch and spread gestures, or rotated with a two finger rotation.

To ensure consistency between views, all views are provided with additional controls and interaction mechanisms from a central library (cf. Section 2). These controls are dependent on the data type, e.g., volume and play controls for video/audio media, navigation shortcuts for time schedule view (see Figure 3), or model manipulation controls for 3D object models.

4.3 Integration of manufacturing data

Plant@Hand combines data from ERP, MES, and PDM systems. With an open ESB infrastructure as described in [17], we establish an integrated data layer between such applications and Plant@Hand which filters and connects required data for our own analysis purpose. Data modifications in Plant@Hand are then passed on to the ESB that translates them into application-specific data updates and requests for underlying ERP, MES, or PDM systems. The advantage of using an ESB as transparent data layer between analysis application and heterogeneous data sources becomes evident through its highly configurable infrastructure. It allows us to flexibly model the dependencies between data and data related information flows and thus to configure the data integration to closely resemble the data flow between the applications according to the workflow.

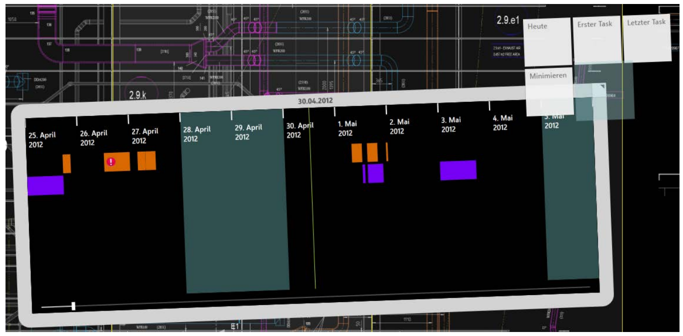

Fig. 3. Contextualized controls of time schedule view.

5 Conclusion and future work

A first qualitative evaluation of Plant@Hand under field conditions showed positive impact of using contextualization on both efficiency and usability. We brought our Plant@Hand on-site into a manufacturing company for cooling devices and evaluated it there with assembly staff. After a short learning phase of basic touch gestures, the assembly team worked autonomously with the application. With respect to efficiency the main benefit was seen in the information integration avoiding time consuming search in paper documents or opening different applications to collect required data. The evaluation also showed a fast adoption of our provided contextualized visualizations. After working for a while with the application, the team was requesting a construction plan manipulation on design level which shows a seamless transition from monitoring and controlling towards re-planning. However, evaluating the overall benefit and performance increase in data analysis will require further long-term studies on-site.

This preliminary evaluation indicates that contextualization is a suitable methodology to design visual data analysis applications close to the user's mental model in order to improve the usability and efficiency of working with heterogeneous data. With Plant@Hand, we have illustrated that our proposed approach can be adopted to any domain that has to deal with heterogeneous data and thus various applications to access and analyze them. In our concrete case of manufacturing, the resulting integrated application ensemble exceeds the available individual analysis tools as it is able to handle data from ERP, MES, and PDM sources through a single seamless analysis interface. While Plant@Hand is already actively promoted within the industrial context on fairs and trade shows, it is subject to further research. With its successor Plant@Hand3D, we are already experimenting with the value of more realistic data and context visualizations. This includes the possibility to move individual views from the ensemble onto smaller tablet devices with which users can move around freely. It allows them to pursue individual tasks on their own at the time and place these need to be pursued, and later reintegrate separately collected data back into the ensemble.

Acknowledgements

Work on this research has been funded in part through an individual grant by the German Research Foundation (DFG).

References

1. Landesberger, T., Schreck, T., Fellner, D.W., Kohlhammer, J.: Visual search and analysis in complex information spaces. In: Expanding the Frontiers of Visual Analytics and Visualization. Springer (2012) 45–67
2. Aehnelt, M., Peter, C., Müsebeck, P.: A discussion of using mental models in assistive environments. In: Proc. of PETRA'12. (2012)
3. Craik, K.J.W.: The nature of explanation. Cambridge University Press (1943)
4. Streit, M., Schulz, H.J., Lex, A., Schmalstieg, D., Schumann, H.: Model-driven design for the visual analysis of heterogeneous data. IEEE TVCG **18** (2012) 998–1010
5. Fekete, J., Hemery, P., Baudel, T., Wood, J.: Obvious: A meta-toolkit to encapsulate information visualization toolkits – one toolkit to bind them all. In: Proc. of VAST'11. (2011) 91–100
6. Gee, A.G., Li, H., Yu, M., Smrtic, M.B., Cvek, U., Goodell, H., Gupta, V., Lawrence, C., Zhou, J., Chiang, C.H., Grinstein, G.G.: Universal visualization platform. In: Proc. of VDA'05, SPIE (2005) 274–283
7. Rogowitz, B.E., Matasci, N.: Metadata mapper: A web service for mapping data between independent visual analysis components. In: Proc. of HVEI'11. (2011)
8. Besacier, G., Vernier, F.: Toward user interface virtualization: Legacy applications and innovative interaction systems. In: Proc. of EICS'09, ACM (2009) 157–166
9. Aehnelt, M.: Contextualized knowledge exchange with mash-up technologies. In: Proc. of eLBa'10, Fraunhofer IRB (2010) 143–153
10. Lee, L.C., Lutteroth, C., Weber, G.: Improving end-user GUI customization with transclusion. In: Proc. of ACSC'10, Australian Computer Society (2010) 163–172
11. Sheth, A.P., Larson, J.A.: Federated database systems for managing distributed, heterogeneous, and autonomous databases. ACM Computing Surveys **22** (1990) 183–236
12. Blasch, E., Bossé, E., Lambert, D.A., eds.: High-Level Information Fusion Management and Systems Design. Artech House (2012)
13. Chaudhuri, S., Dayal, U.: An overview of data warehousing and olap technology. SIGMOD Record **26** (1997) 65–74
14. Calì, A., Calvanese, D., Lenzerini, M.: Accessing data integration systems through conceptual schemas. In: Proc. of ER'01, Springer (2001) 270–284
15. Levy, A.Y.: Logic-based techniques in data integration. In Minker, J., ed.: Logic-Based Artificial Intelligence. Springer (2000) 575–595
16. Lenzerini, M.: Ontology-based data management. In: Proc. of AMW'12, CEUR Workshop Proceedings (2012) 12–15
17. Aehnelt, M., Bader, S., Ruscher, G., Krüger, F., Urban, B., Kirste, T.: Situation aware interaction with multi-modal business applications in smart environments. In: Proc. of HCII'13, Springer (2013)
18. Ceglowski, M., Coburn, A., Cuadrado, J.: Semantic search of unstructured data using contextual network graphs. Preliminary white paper. National Institute for Technology and Liberal Education, Middlebury College, Middlebury, Vermont, USA (2003)