Contributing to Current Challenges in Identity and Access Management with Visual Analytics

Alexander Puchta, Fabian Böhm, Günther Pernul

Contributing to Current Challenges in Identity and Access Management with Visual Analytics

Alexander Puchta[1], Fabian Böhm[2][0000−0002−0023−6051], and Günther Pernul[2]

[1] Nexis GmbH, Franz-Mayer-Str. 1, 93053 Regensburg, Germany
alexander.puchta@nexis-secure.com
[2] University of Regensburg, Universitätsstr. 31, 93053 Regensburg, Germany
fabian.boehm | guenther.pernul@ur.de

Abstract. Enterprises have embraced identity and access management (IAM) systems as central point to manage digital identities and to grant or remove access to information. However, as IAM systems continue to grow, technical and organizational challenges arise. Domain experts have an incomparable amount of knowledge about an organization's specific settings and issues. Thus, especially for organizational IAM challenges to be solved, leveraging the knowledge of internal and external experts is a promising path. Applying Visual Analytics (VA) as an interactive tool set to utilize the expert knowledge can help to solve upcoming challenges. Within this work, the central IAM challenges with need for expert integration are identified by conducting a literature review of academic publications and analyzing the practitioners' point of view. Based on this, we propose an architecture for combining IAM and VA. A prototypical implementation of this architecture showcases the increased understanding and ways of solving the identified IAM challenges.

Keywords: Identity and Access Management · Identity Management · Visual Analytics

1 Introduction

Identity and access management (IAM) has become a vital component of modern companies as it enables the management of identities and grants access to necessary resources. IAM also assures compliance with regulations like SOX [41] or Basel III [3]. To achieve this, IAM systems consist of manifold policies, processes and technical solutions [13]. The core of IAM are identities like employees and their access rights to resources maintained within the system. Besides human identities, new technologies like the Internet of Things (IoT) require the integration of technical identities (e.g. sensors and machines) into IAM [27]. Thus, the number of elements maintained in the system is constantly rising. This will ultimately lead to an identity explosion where a vast amount of heterogeneous identities has to be managed in a single system. This results in numerous problems to be addressed in the next years to ensure IAM systems remain an effective part of companies' IT landscapes.

A solution for those problems needs an effective way to manage and analyze the huge quantity of information with often thousands of identities and hundreds of thousands of entitlements. To decide whether information about an identity is wrong or redundant access rights are assigned to it, the knowledge of domain experts with experience and deep understanding of an enterprise's individual IAM landscape is needed. In this work we investigate how this domain knowledge can be integrated into an IAM landscape by leveraging Visual Analytics (VA) as VA is one of the central methods to include domain experts' knowledge and utilize their feedback [11]. In order to reach this goal, this work investigates three research questions:

- **RQ-1:** What are current and upcoming key challenges within IAM to be solved by integrating domain knowledge?
- **RQ-2:** How can VA be integrated into an existing IAM architecture and which steps are necessary?
- **RQ-3:** What could an exemplary VA solution for IAM look like and which challenges could be solved?

By answering these research questions our work focuses on two main contributions. We provide a list of challenges for current and future IAM. This list is an outcome of a structured analysis taking both academic and practice viewpoints into consideration. We also demonstrate how VA can be applied helping to integrate domain knowledge in tasks to identify IAM anomalies and possible erroneous configurations (e.g. over-authorization or wrong identity attributes). Therefore, we develop a prototypical visualization designed in cooperation with experienced IAM practitioners.

The remainder of this work is structured as follows. Section 2 introduces some background on IAM systems as well as related work regarding the integration of VA into IAM. Next, Section 3 follows a structured, two-fold approach to identify current challenges for IAM system as seen from academia and practice to answer *RQ-1*. An architecture to integrate VA into IAM (*RQ-2*) as well as a corresponding proof-of-concept visualization (*RQ-3*) are presented in Section 4. The benefits of this prototype regarding the identified challenges are highlighted with exemplary use cases in Section 5. Section 6 concludes our work and highlights possible future research directions as well as current limitations.

2 Background & Related Work

In this chapter we define key concepts of IAM and introduce related work regarding the integration of VA into IAM.

2.1 Background

IAM consists of two main fields which are managing identities and granting them access to resources. According to Pfitzmann and Hansen [33] an identity is a subset of attributes uniquely identifying a person. An identity is either real or

exists as a digital identity like profiles in social media. Real and digital identities are often linked, and a real identity may own multiple digital personas. However, in the following we assume each entity to have exactly one digital identity as the scope of this work is limited to a single company's context. Currently, IAM regards employees, contractors or customers as identity because they all need to have access to certain resources [45]. In addition to humans having digital identities, technical equipment like machines or sensors are entities which need access to resources, too. Thus, these technical identities also are relevant for maintaining them within an IAM [12].

Digital identities in an IAM are managed from their creation to their deletion when not needed anymore. During this life cycle, access control is used to provide access to applications, data or other information [35]. Enterprises often employ role-based access control (RBAC) in order to grant access [37]. In RBAC, roles are utilized to bundle single access rights and consequently assigned to identities. On the contrary, attribute-based access control (ABAC) leverages identities' attributes and predefined access policies for dynamic access management [16].

To maintain landscapes with thousands of identities, enterprises employ IAM systems which are able to support the identity life cycle and provide identities with the correct entitlements. Besides that, modern IAM systems offer a variety of other functionalities (e.g. Single Sign-on) which are not detailed any further in this work.

2.2 Related Work

There are some existing publications applying visual representations for IAM problems. The earliest integration of VA to the best of our knowledge is the "role graph model" by Nyanchama and Osborn [32]. It is based on RBAC and is used to optimize existing roles for a company. In addition to that, several authors propose a matrix-based approach to visualize users and their entitlements [5,28]. Based on that, VA can be applied to identify suitable roles or outliers with extensive entitlement assignments. Recently, Morisset and Sanchez introduced a tool to visualize ABAC policies [30].

These approaches are focusing mostly on interactive visual techniques for Access Control. To the best of our knowledge there is no existing work taking Identity Management into consideration to build a more cohesive visual solution. Therefore, we try to fill this gap by identifying general IAM challenges where domain knowledge of experts is needed to solve them. After identifying those challenges, we build a prototypical visual approach to demonstrate how domain experts can be integrated.

3 IAM Challenges

This section defines current or future IAM challenges where domain knowledge of human experts may play a vital role. They can serve as a starting point to deduce requirements for any type of solution trying to tie experts and IAM systems

closer together. In Section 4 we introduce a proof-of-concept visual solution to tackle some of the herein defined challenges.

For identifying the challenges, existing academic literature as well as practitioners experience within the field of IAM are taken into consideration. We are aware that there are far more challenges than the five proposed by us. However, based on the results of our structured analysis and the domain knowledge of practitioners, we chose the most relevant ones with respect to the necessity to integrate domain experts. An *IAM challenge* in the context of this work is a current or future problem with the need to be solved for IAM. Challenges already being tackled or focusing only on parts of an IAM system (e.g. access control) are not considered in this work. Neither do we consider problems where the inclusion of domain expert knowledge is not vital. To identify challenges, we follow a structured approach introduced in the Section 3.1.

3.1 Approach for Identifying Challenges

We derive current challenges following a structured approach depicted in Figure 1. We ensure to include both the scientific and the practitioners' view as IAM is an active research field as well as a highly relevant topic in enterprises.

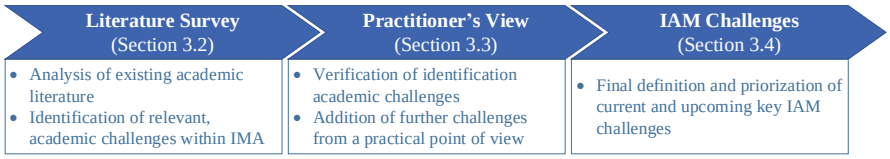

Fig. 1: Approach for defining the key IAM challenges.

During the literature survey we analyze existing academic literature published in the last ten years regarding IAM and respective challenges or problems. This scientific viewpoint allows us to derive a first set of IAM challenges. However, IAM is highly business-driven and there are numerous practical approaches outside the academic world. Thus, we also include the perspective of practitioners.

The goal of this second analysis is twofold. We verify the academic IAM challenges but also identify further challenges not yet considered by scientific literature. Three different sources of information are leveraged in order to minimize subjectivity of different business opinions:

1. Analyst reports and surveys from the IAM industry
2. Interviews with IAM consultants with 3 to 15 years of experience
3. Interviews with companies applying IAM solutions

In a last step, we integrate all inputs from the analysis into five IAM challenges. The list of identified challenges is not exhaustive for the IAM field of research. Our work is focusing only on current challenges that can strongly benefit from integrating experts' knowledge.

3.2 Literature Survey

In order to identify relevant literature, we follow a structured approach by defining keywords to review relevant IAM literature. As we are defining challenges for the entire IAM system we only take resources into consideration which are dealing either with *"identity and access management"* or specific problems and challenges within *"identity management"* or *"access management"*. We transform these phrases into suitable search terms[2,3] and applied them to the dblp computer science bibliography[1]. Dblp is a service indexing relevant academic journals and proceedings of peer-reviewed conferences from computer science. Searching dblp results in a feasible number of results with a high suitability. Therefore, we can ensure to get only relevant academic publications. Dblp serves as a quality gate for our scientific analysis as it returns a manageable amount of entries compared to other engines like Google Scholar with nearly 10.000 results for the second search term. We manually filter the results based on title, abstract, and key sections to remove findings not mentioning any challenges or problems.

We apply a second, more unstructured search to identify additional relevant entries. In this step we include further academic databases (IEEE XPlore, ACM, Google Scholar) to find additional literature not listed within dblp. This results in a total of 19 academic publications mentioning or clearly defining relevant challenges for IAM. We group the identified problems and define the first four challenges (cf. *C1* to *C4* in Section 3.4).

3.3 Practitioner's View

We now conduct a business analysis to include the practitioners' point of view. Information received in this process step is often hard to generalize as it reflects subjective opinions. However, by including various sources of information we try to overcome this deficit. In the business analysis we look at reports from specialized IAM analysts namely KuppingerCole[4]. This company is focused on IAM and technologies around that sector and thus has accumulated valuable knowledge in this area [26,40,43]. Additional input is generated by Gartner [9], Forrester [7] and IDG Research Services [20].

Furthermore, we conduct interviews with three different IAM consultants with several years of practical experience in the field of IAM projects. Besides that, four companies already applying IAM solutions are inquired regarding

[2] *identity/access management challenge/problem*

[3] *identity-and-access-management*

[1] `https://dblp.uni-trier.de/`

[4] `https://www.kuppingercole.com`

possible challenges. While the four previously defined challenges are verified throughout the interview, a fifth one (*C5*) arises as a current problem of IAM from a business viewpoint.

Table 1: Results of literature survey on ten years of academic work.

Source	**Year**	**C1**	**C2**	**C3**	**C4**	**C5**
Hovav and Berger [15]	2009	x	x			
Mahalle et al. [27]	2010	x	x	x		
Bandyopadhyay and Sen [2]	2011	x	x	x		
Jensen [22]	2012		x	x	x	
Kanuparthi et al. [23]	2013	x	x		x	
Fremantle et al. [12]	2014	x		x		
Xiong et al. [46]	2014		x			
Hummer et al. [18]	2015	x		x	x	
Kunz et al. [24]	2015				x	x
Hummer et al. [19]	2016	x		x	x	x
Moghaddam et al. [29]	2017		x			
Servos and Osborn [38]	2017		x			
Asghar et al. [1]	2018		x			
Damon et al. [8]	2018	x		x		
Hummer et al. [17]	2018	x		x	x	
Indu et al. [21]	2018	x	x	x		
Nuss et al. [31]	2018	x		x		
Povilionis et al. [34]	2018		x		x	
Kunz et al. [25]	2019			x	x	x

3.4 IAM Challenges

Within this section the identified IAM challenges are described in detail. A mapping of all relevant academic publication to the challenges is provided in Table 1. Table 2 maps the results of our analysis with practitioners to the challenges.

Table 2: Analysis results from practitioners' view.

Source	**Year**	**C1**	**C2**	**C3**	**C4**	**C5**
IDG Research Services [20]	2017	x		x		
KuppingerCole and CXP Group [26]	2017	x				
Tolbert [43]	2017	x		x		
Diodati et al. [9]	2018	x		x		
Small [40]	2018	x		x	x	
Cser and Maxim [7]	2018	x	x	x		
Interviews (IAM consultants)	2019	x	x	x	x	x
Interviews (Companies applying IAM)	2019	x		x	x	x

Challenge 1 - Identification of all Relevant Identities (C1): For current and future IAM systems the identification of all relevant identities may sound like a simple task. However, especially in practical application it is not. One of the major reasons for this is the integration of various types of identities into IAM. Currently, mainly employee and contractor identities are maintained in an IAM system. A recent trend, customer IAM or shortly CIAM, strives to add customer identities into these systems as well [7]. Additionally, the Internet of Things requires integrating even more identities, mostly technical ones [31]. Furthermore, numerous IT systems are not even connected to IAM. Nevertheless, such systems also contain various identities with the need to be identified for IAM in order to prevent identities not being centrally manageable. These trends hinder IAM to establish a central view of all relevant identities. However, this view is vital for any further analysis to be done within IAM (e.g. identification of unnecessary accounts or entitlements).

Challenge 2 - Privacy within IAM (C2): As modern IAM systems offer a centralized view on nearly all employees, contractors and even costumers including their attributes the need for privacy arises. Especially business solution power users like IAM administrators can easily retrieve personal information from the identities. Based on our practical experience this could be a simple mail address but may also uncover more sensitive information like wage brackets or entitlement usage information. In order to protect this information in compliance with regulations, privacy mechanisms are needed to grant access to such information only when necessary and for authorized users. This challenge is mainly focused by scientific research and not by practitioners at the moment. However, as the European General Data Protection Regulation (GDPR) came into effect in 2018, it certainly will have an impact on the business sector of IAM. Please note that this challenge is limited to the application of privacy mechanisms on IAM systems and does not include the application of IAM systems for enhancing GDPR compliance within companies.

Challenge 3 - Heterogeneity of Various Identities (C3): As there are various identities within an IAM system, they are not identical. In fact, they differ quite a bit as identities consist of various attributes (e.g. first name, department). Considering *C1*, it gets clear that not all identities have the same kind of attributes. Technical and human identities are likely to have a completely different set of attributes. For example, technical devices do not have a first name, but instead have an attribute indicating their software version. This, on the one hand, rises a technical challenge to integrate this variability of identities into one underlying data set for IAM. In addition, IAM mechanisms like provisioning of entitlements still need to be working for all of these identities. On the other hand, it also hinders the analytic part of IAM as domain experts need to browse through an enormously large, heterogeneous database. By applying VA, domain experts could be supported as various attributes can be displayed in a more accessible way than in currently deployed table-based reports.

Challenge 4 - Data Quality & Data Management (C4): When it comes to attributes and other data existing in IAM, data quality and the underlying data management in IAM system needs to be considered. Attributes are often manually entered by different people; thus, wrong or inconsistent values are very likely to occur. For example, the current business location of an employee may be added by HR employees. If the employee moves to another department of an enterprise, the location also needs to be changed. Manual processes for attribute modifications exacerbate data quality issues as one can forget to adjust the location attribute. Therefore, IAM mechanisms like provisioning of entitlements based on the attribute *location* might fail. Additionally, wrong attribute values limit the possibilities of IAM analytics. Although an approach to improve attribute quality management was lately introduced [25], algorithms can only detect anomalies but can neither confirm nor reject whether it is a real data error. To do so, domain experts are needed, and VA can be highly beneficial to support related decisions by integrating domain expert feedback.

Challenge 5 - Transformation from Role-based IAM to Attribute-based IAM (C5): Challenge 5 was identified during the interviews with IAM consultants as it is not explicitly defined as an upcoming challenge in academic literature. It comprises the enterprise IAM transformation from a role-based approach to an attribute-based one. As mentioned before, enterprises mainly depend on an RBAC approach. However, this can lead to an increasing number of existing roles and requires increasing effort regarding role management [10]. In order to overcome these limitations, ABAC can be applied [16]. However, as this is a fundamental change of approach for IAM companies have to consider various factors (e.g. processes, technologies and policies [13]). Changes needed for this transformation are therefore not limited to access control, but existing research is mainly focused on the transformation of the access control model [36,47]. To the best of our knowledge there is no overarching approach how an enterprise IAM can be transformed from a role-based approach to an attribute-based one.

Tables 1 and 2 compare the results and show that *C1*, *C3*, and *C4* are found in both worlds and can easily be identified as relevant IAM challenges. Privacy in IAM and therefore, *C2*, is mainly embraced by academic literature and not explicitly mentioned in the business sector. *C5* is not described explicitly in academic literature but only mentioned very shortly by 3 articles. We identified this challenge by conducting interviews with IAM consultants and companies.

4 Applying Visual Analytics to IAM

Any of the previously identified challenges can benefit from including domain experts and their knowledge. VA has proven its capabilities to help integrate domain expert knowledge in complex and data-intensive cyber security tasks throughout the last years [44,6]. Additionally, decision makers can be supported with VA by making highly technical data sources more accessible. Therefore, we

argue that leveraging concepts from VA to solve the identified challenges in IAM is a reasonable approach. As described in Section 2, there is some existing work that has shown the feasibility and utility of VA in the context of IAM. However, none of the challenges identified in Section 3 has been explicitly tackled with visual approaches yet. We try to fill this gap as we describe the architecture and design of our new visualization approach. The visualization design cannot support all the identified challenges as they are far too different in requirements. However, our approach shows how heterogeneous information about human and technical identities can be integrated into a single visual representation. The resulting view allows identifying existing identities (c.f. C1) and their structures (c.f. C3) as well as users can detect problems regarding data quality (c.f. C4).

The visual representation is designed and implemented in close cooperation with IAM practitioners which were also part of our interviews during the challenge identification. By including them in development, we ensure that the representation that is helpful for practical use. The participating experts are IAM consultants working for numerous clients and with years of experience in practical work with IAM projects. While the current visual tool is at a proof-of-concept stage, we are planning to continue our fruitful cooperation with these experts to develop a solution that can be used in their day-to-day work. Our cooperation also allowed for the development of the prototype based on the adaption of anonymized real-world identity data from a medium-sized company in the manufacturing sector with around 1.200 employees.

The underlying architecture for our prototypical application is depicted in Figure 2 and its main components - *Data Sources*, *Data Preparation*, and *Data Visualization* - are described in more detail throughout the following sections. This architectural design is based on the Information Visualization Pipeline [4] which is a widely accepted structural design concept for any interactive visualization approach. The applied architecture shows how identity-related information can be collected and integrated from different sources and how the information needs to be prepared for VA concepts supporting domain experts. The identification of different data sources and their integration into a single, displayable data set are a starting point for any visual representation of identity data. Therefore, the main part of our architectural design, the *Data Preparation*, demonstrates how visual representations in general can be integrated into an existing IAM structure. The operations executed during the *Data Integration Engine* and the *Data Transformation* step need only small adjustments for varying *Data Sources*. The last part of the architectural design, the *Data Visualization*, demonstrates how VA can contribute to the focused challenges by introducing an exemplary visualization of identity information. Interaction in this step is crucial as it ensures that experts can adjust the view for their personal needs and explore the data based on their own preferences to gain insight.

4.1 Data Sources

Our current proof-of-concept tool collects information about identities from three main data layers. Although the company representing the use case has a cen-

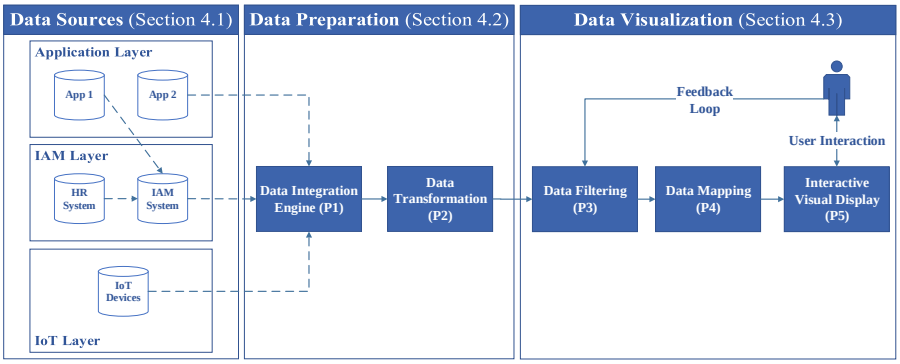

Fig. 2: Architecture for IAM Visual Analytics.

tral IAM system with a role-based access control mechanism, not all information about the existing identities is fed into it. Only partial information from the *Application Layer* is integrated into IAM, while other applications, like the company's Active Directory (AD) to manage windows accounts, are not connected to it. Therefore, information from these systems needs to be collected separately. Technical identities representing IoT devices are currently not integrated into IAM but rather maintained separately (IoT-layer). The wide variety of different data sources storing information about the companies' identities and the missing integration of this information into IAM are the main reasons for the challenges we focus in this prototype. It becomes increasingly hard for any company to keep track of its identities when the information about them is so spread out. Additionally, the different information systems store the available data in different formats or data models. Furthermore, it is very important for any company to keep the quality of their identity information at a high level. Spread out data makes this very difficult, especially when data is maintained redundantly in different repositories.

Additional data sources can be plugged easily into our architecture via the *Data Integration Engine*. Our proof-of-concept system works with one source from each layer. This number of data sources is already enough to demonstrate how VA helps to leverage experts' domain knowledge in the context of the aforementioned challenges as is demonstrated in Section 5.

4.2 Data Preparation

The main purpose of this part of the architectural design is to integrate and normalize the data from the sources into a single data model and format. Additional fields are added and calculated in this step. The resulting data is structured as a single table containing all relevant and necessary information about the identities. This step is essential for further visual display as it defines the level of detail available to the users. The operations applied to the data in this step are

dynamic and can be changed whenever different information is of interest or new data sources are plugged into the architecture.

Data Integration Engine (P1): This part of the architecture is responsible for collecting data relevant from the data source and integrating them into a single cohesive data set. Our proof-of-concept work extracts CSV data from all data sources. However, each CSV export contains a different set of attributes. To preserve the information about the source of a data set, we annotate the data with a flag depicting the source. Additionally, we add a field describing whether the identity is a human or a technical identity. In our conceptual setting, this identity type is mainly dependent on the data source. For example, identities extracted from IAM are automatically annotated to be *"Human"* as only employees or costumers are integrated into IAM. In the same way, identities extracted from the IoT layer are annotated to be *"Technical"* identities. The cohesive data set is built as a union of the three data sets depicted in Figure 2: $ApplicationLayer \cup IAMLayer \cup IoTLayer$.

Data Transformaion (P2): After integrating all available data sources into a single, high-dimensional table, this data set is structured as needed for the visualization in this phase. This part of the architecture applies a variety of transformations. These include splitting a single field into multiple fields, replacing values in a specific field, calculating additional fields based on existing information. The result of this step is a cohesive data set containing all relevant and necessary information.

4.3 Data Visualization

This last part of the architecture is responsible for creating the interactive visual representation with the subset of the data selected by the domain expert. The interactions available to the users enable exploratory work with the visualization and the identification of inconsistencies, miss-configurations as well as structures and dependencies in the set of identities.

Data Filtering (P3): The interactive filtering assures the efficiency of the following steps and guarantees the expressiveness of the resulting view for the user. The subsequent *Data Mapping (P4)* can be CPU-intensive for very large data sets and, therefore, the input for this step needs to be as small as possible. It only contains the fields (columns) the user wants to see. The interactive selection of fields relevant for the user and the early integration of this interaction into the architectural design ensures that only relevant data is passed to the subsequent components. Up to this point the proposed architecture is generalized and can be applied to various visualization approaches within IAM. However, the *Data Mapping (P4)* and the *Interactive Visual Display (P5)* are highly dependent on the visualization technique selected for a specific VA solution. Therefore, the following considerations are specific to our exemplary solution.

Data Mapping (P4): This phase in the architecture maps the filtered identity information into a dynamic hierarchical data structure which is necessary for the prototype to visualize the data correctly. We will not elaborate this data structure any further as it is specific for the proof-of-concept visualization and is prone to change for different visualization types.

Fig. 3: Screenshot of the prototype available under `http://bit.ly/iam-vis`. Please note that the current version of the tool is only working in Google's Chrome Browser.

Interactive Visual Display (P5): Before we are able to build a visual representation for the data at hand, it is necessary to choose a suitable visualization technique. This technique needs to be capable of displaying the dimensions and structure of the underlying data properly. For our prototypical visualization the technique must be able to represent multi-dimensional and hierarchical data. While there is a number of techniques (e.g. tree diagrams, circle packing, sunburst diagrams, or treemaps) which fulfill this requirement [39], each technique has its own advantages and disadvantages. It mainly comes down to the use case as well as the subjective preferences of the users which technique is most suitable. We used design sketches of the different suitable visualization techniques to interview IAM domain experts about their preferred visual representation. These interviews resulted in the *sunburst diagram* to be the most preferable technique to apply in the proof-of-concept application. However, any of the mentioned as well as a number of other techniques might be suitable, too.

The sunburst diagram displays a hierarchy using a series of concentric circles. Each ring itself corresponds to a level of the hierarchy. Therefore, underlying data structure is similar to a tree where the root node is depicted by the central ring and outermost circles represent the leaves of the tree-like structure. The sunburst's rings are sliced up and divided based on their hierarchical relationship to the parent slice. Therefore, the sunburst highlights structural, hierarchical relationships while being more scalable than other hierarchical visualization types. Figure 3 depicts the main view of our proof-of-concept application consisting of three main parts.

In (A) experts can drag-and-drop the boxes for the corresponding fields in the data set they want to be depicted in the sunburst diagram between two main lists. Each box represents an IAM employee attribute of the normalized data set. Exemplary fields which are contained in the proof-of-concept are the "Organisational Unit", "Data Source", or the number of entitlements of an identity. The upper list holds the currently active (i.e. displayed) fields and the lower one the inactive attributes. The first element in the list of active elements serves as the root element (innermost circle in the sunburst) when constructing the hierarchical data set. Accordingly, the second active attribute is displayed as the next outer circle. Logically, the last active element is included in the Sunburst diagram as the outermost circle.

The central part of the view is dedicated to the sunburst diagram (B). Each segment of the circles (attributes) depicts a single characteristic of an attribute in relative size to all existing identities. Hovering a segment brings up the number of identities depicted by this segment. The relative frequency with respect to the number of current root element (innermost circle) and the path to the hovered segment are displayed on top of the visual representation. Left-clicking a segment allows zooming in on this particular representation of an attribute. This improves the readability of specific hierarchy levels in very granular sunburst displays. When zoomed into a segment, clicking the white area in the middle of the sunburst diagram brings the zoom one hierarchy level upwards. Right-clicking a segment brings up a dialog. This dialog holds a table with identities in the rows and all attributes available for them in the columns. The identities displayed in the table are dependent on the clicked segment of the sunburst diagram as only the identities whose attributes fulfill the path to the segment are shown in the table. The table allows filtering and sorting of the currently displayed identities. In the dialog identities can also be reported for further analysis if necessary.

(C) holds the description of the sunburst as a dynamic list containing all currently visible circles (attributes) and the respective visible ring segments (attributes values). Clicking the magnifier for a list element zooms in into the segment representing this element. A click on the counter badge brings up the table with the identities included in corresponding node in the hierarchy. Within this table identities can be marked for further analysis by using a *"Report"*-button for each identity in the details table-view (e.g. after identification of an anomaly), thus, providing the possibility for integration of domain expert feedback into other applications. However, further functionality beyond this notification is out

of scope for this work and needs to be implemented in a following version of the prototype.

5 Exemplary Use Cases

The current prototypical implementation[3] of our visualization for IAM was developed in co-creation with experts as suggested by Staheli et al. [42]. We regularly conducted semi-structured interviews with the participating practitioners to ensure that the implementation fits their needs and requirements. This section explicitly highlights how the visual display can support domain experts. We therefore go through several problems and inconsistencies based on one use case and identified by IAM experts while exploring the data. These had not been noticed before applying the visualization.

As the different problems only become evident in the sunburst diagram with different actively visualized attributes, we added predefined scenarios of the sunburst to our publicly available version of the prototype. Using the drop-down menu in the top right corner, we provide a video showcasing each of the following subsections. We would recommend to look at the corresponding video for each subsection in order to grasp the connection between the IAM problem and the sunburst visualization for identification of the inconsistency.

The exemplary use cases are based on the data set from a manufacturing company with 1.200 employees mentioned in Section 4. The company recently introduced an IAM system and connected the HR system as well as some minor applications. However, the Active Directory (AD) is currently not under IAM control because of its complexity as it was one of the company's first IT system growing for two decades. Therefore, some employees are missing an AD account while some AD accounts from former contractors and employees are still active. These orphan accounts are not identified via the IAM system, but they are still active and can be used for malicious activities (cf. Section 5.1).

Furthermore, the company made some investments in automating specific process tasks. Thus, two assembly machines and some automated users were integrated within the AD and were provisioned by an AD administrator. However, there was no communication with the IAM department and, therefore, no access management or integration into the IAM system took place. This results in technical identities with excessive entitlements. As the company is not experienced with such technical identities, the risk for failures (e.g. deletion of data) resulting from misconfiguration is high (cf. Section 5.2).

During configuration and assignment of a location to the technical identities some flaws regarding the existing location attribute values were detected. As entitlements shall be assigned automatically within the new IAM system based on a policy, the identification and correction of these values is highly relevant. Otherwise, identities with an incorrect value for their location attribute are not assigned enough entitlements (cf. Section 5.3).

[3]The prototype is available under `http://bit.ly/iam-vis`. Please note that the current version of the tool is only working on Google's Chrome Browser.

5.1 Identities Not Managed within a Central IAM (C1, C4)

As stated before, some identities within the company are not integrated in the central IAM system. Identifying these is a hard task considering the spread-out information. Our approach integrates applications not connected to the IAM system. Taking a look at the *"Data Source"* attribute the sunburst shows in which layer the respective data originates. Identities in the *"IAM Layer"* segment of the diagram are collected directly from IAM. However, another 17 identities are not managed by the IAM system. Three of them are maintained in the *"AD Layer"* while 14 are gathered from the *"IoT Layer"*. Taking a look at the details view of those 14 identities brings out that they are technical devices. Adding another circle for the *"Identity Type"* to the Sunburst allows an analyst to see that none of the identities within the *"IAM Layer"* are technical devices. So obviously the company has not integrated its technical identities into the central IAM system.

5.2 Identities with an Unusual Number of Entitlements (C3)

Another use case needing the attention of domain experts are identities with anomalous high number of entitlements. The Sunburst Diagram facilitates the identification of relevant entities and a decision how to proceed. Displaying the *"Entitlements"*, *"Identity Type"*, and the *"Function"* a small set of identities becomes visible having more than 76 entitlements. This seems conspicuous as most of the entities in the company have 0 to 25 roles assigned to them. Zooming into the segment with 76 to 100 entitlements a technical device attributed with function *"Support"* becomes visible and an identity from the company's customer *"Brandmark"* has an anomalous number of entitlements. These findings do not indicate an error per se, but it might be necessary to carry out further analyses. By browsing through the Sunburst domain experts are enabled to find various of similar cases. Any identity which might be over-authorized has to be examined, if all the entitlements are still needed (e.g. via recertification). If not, this indicates a serious security breach as identities having excessive permissions are legitimately allowed to access classified resources.

5.3 Poor Data Quality in IAM Data (C4)

The Sunburst diagram allows for identifying data quality flaws via several means. A first possibility is to compare similar attribute fields which originate from different data sources. Exemplary for this used when comparing *"Location (AD)"* and *"Location (IAM and IoT)"*. Information about locations of identities is administered in both the application layer and the IAM layer. Usually the IAM layer which contains the HR should be the master system for attributes like the location. After analyzing the data, some quality issues included in this system become apparent. An example is visible by zooming to the value *"Berlin"* of the *"Location (AD)"* attribute. The IAM layer has in fact 3 different attributes for identities having this value in the AD system, namely the correct value *"Berlin"* but also *"BER"* and *"10249 Berlin"*. Presumably, this value was recorded manually by the HR employees resulting in inconsistent data.

6 Conclusion

The complexity of modern IAM systems is constantly rising (e.g. increasing number of identities, further IAM mechanisms). Therefore, new challenges emerge. Within this work we showed that VA can be integrated into IAM in order to solve some of them. To achieve this, we initially identified five central challenges through a review of academic literature and analysis of the experience of practitioners (*RQ-1*). Thereby, we discovered two challenges especially connected to the identification and management of identities. Furthermore, we expect more challenges within the topics *Privacy* and *Data Quality*. Besides that, there will be the future challenge to transform role-based IAM into an attribute-based architecture for enterprises. We do not claim that our list of IAM challenges is exhaustive. However, we focused especially on problems where the integration of domain expert knowledge is vital. We detected some additional challenges but excluded them as they are not in the scope of the paper (e.g. inclusion of trust management in IAM, identity as a service, compliance with regulations).

Based on these challenges we identified VA as a possible solution as it enables enterprises to integrate domain expert knowledge and utilize their feedback to solve upcoming IAM challenges. We proposed an architecture how IAM by leveraging concepts from VA in order to answer our previously defined *RQ-2*. Additionally, we implemented proof-of-concept visualization according to our architecture and based on real world data (*RQ-3*). By applying VA, we have shown that problems tightly connected to the defined IAM challenges can be identified. However, the implementation should be regarded as a first example how the architecture can be implemented and as proof that VA can support enterprises to solve central IAM challenges. Other visualization techniques might be applied to solve another subset of our identified challenges.

After proposing an architecture for integration of VA and a first proof-of-concept implementation we want to focus further on the process to choose a suitable visualization for the *Interactive Visual Display* component. Additionally, we want to introduce further VA implementations to solve the remaining IAM challenges. Afterwards, we can orchestrate the single implementations to an overarching coordinated view [14].

Acknowledgment. This research was supported by the Federal Ministry of Education and Research, Germany, as part of the BMBF DINGfest project (https://dingfest.ur.de).

References

1. Asghar, M., Backes, M., Simeonovski, M.: PRIMA: Privacy-preserving identity and access management at internet-scale. In: Proceedings of the 2018 IEEE International Conference on Communications. pp. 1–6. IEEE Computer Society (2018)
2. Bandyopadhyay, D., Sen, J.: Internet of things: Applications and challenges in technology and standardization. Wireless Personal Communications **58**(1), 49–69 (2011)

3. Basel Comittee on Banking Supervisions: Basel III: Int. framework for liquidity risk measurement, standards and monitoring (2010)
4. Card, S.K., Mackinlay, J.D., Shneiderman, B. (eds.): Readings in information visualization: Using vision to think. Morgan Kaufmann, Burlington (1999)
5. Colantonio, A., Di Pietro, R., Ocello, A., Verde, N.: Visual role mining: A picture is worth a thousand roles. IEEE Transactions on Knowledge and Data Engineering **24**(6), 1120–1133 (2012)
6. Crouser, R., Fukuday, E., Sridhar, S.: Retrospective on a decade of research in visualization for cybersecurity. In: Proceedings of the 2017 IEEE International Symposium on Technologies for Homeland Security. pp. 1–5. IEEE (2017)
7. Cser, A., Maxim, M.: Forrester - Top trends shaping IAM in 2018 (2018)
8. Damon, F., Coetzee, M.: The design of an identity and access management assurance dashboard model. In: Proceedings of the International Conference on Research and Practical Issues of Enterprise Information Systems. pp. 123–133. Springer (2018)
9. Diodati, M., Farahmand, H., Ruddy, M.: Gartner - 2019 planning guide for identity and access management (2018)
10. Elliott, A., Knight, S.: Role explosion: Acknowledging the problem. In: Proceedings of the 8th International Conference on Software Engineering Research and Practice. pp. 349–355 (2010)
11. Federico, P., Wagner, M., Rind, A., Amor-Amorós, A., Miksch, S., Aigner, W.: The role of explicit knowledge: A conceptual model of knowledge-assisted visual analytics. In: Proceedings of the 2017 IEEE Conference on Visual Analytics Science and Technology (2017)
12. Fremantle, P., Aziz, B., Kopecký, J., Scott, P.: Federated identity and access management for the internet of things. In: Proceedings of the 2014 International Workshop on Secure Internet of Things. pp. 10–17. IEEE Computer Society (2014)
13. Fuchs, L., Pernul, G.: Supporting compliant and secure user handling-a structured approach for in-house identity management. In: The Second International Conference on Availability, Reliability and Security (ARES'07). pp. 374–384. IEEE (2007)
14. Heer, J., Shneiderman, B.: Interactive dynamics for visual analysis. Queue **10**(2), 30 (2012)
15. Hovav, A., Berger, B.: Tutorial: Identity management systems and secured access control. Communications of the Association for Information Systems **25**(1), 1–42 (2009)
16. Hu, V.C., Ferraiolo, D.F., Kuhn, D.R., Schnitzer, A., Sandlin, K., Miller, R., Scarfone, K.: Guide to attribute based access control (ABAC) definition and considerations. NIST Special Publication (2014)
17. Hummer, M., Groll, S., Kunz, M., Fuchs, L., Pernul, G.: Measuring identity and access management performance - an expert survey on possible performance indicators. In: Proceedings of the 4th International Conference on Information Systems Security and Privacy. pp. 233–240 (2018)
18. Hummer, M., Kunz, M., Netter, M., Fuchs, L., Pernul, G.: Advanced identity and access policy management using contextual data. In: Proceedings of the IEEE International Conference on Availability, Reliability and Security. pp. 40–49. IEEE Computer Society (2015)
19. Hummer, M., Kunz, M., Netter, M., Fuchs, L., Pernul, G.: Adaptive identity and access management - contextual data based policies. EURASIP Journal on Information Security **2016**(1), 1–19 (2016)

20. IDG Research Services: Studie Identity- & Access-Management 2017 (2017)
21. Indu, I., Anand, P.R., Bhaskar, V.: Identity and access management in cloud environment: Mechanisms and challenges. Engineering Science and Technology, an International Journal (2018)
22. Jensen, J.: Federated identity management challenges. In: Proceedings of the 2012 IEEE International Conference on Availability, Reliability and Security. pp. 230–235. IEEE Computer Society (2012)
23. Kanuparthi, A., Karri, R., Addepalli, S.: Hardware and embedded security in the context of internet of things. In: Proceedings of the 2013 ACM workshop on Security, privacy & dependability for cyber vehicles. pp. 61–64. ACM (2013)
24. Kunz, M., Fuchs, L., Hummer, M., Pernul, G.: Introducing dynamic identity and access management in organizations. In: Proceedings of the 11th International Conference on Information Systems Security. pp. 139–158 (2015)
25. Kunz, M., Puchta, A., Groll, S., Fuchs, L., Pernul, G.: Attribute quality management for dynamic identity and access management. Journal of Information Security and Applications **44**, 64–79 (2019)
26. KuppingerCole, CXP Group: State of organizations - does their identity & access management meet their needs in the age of digital transformation? (2017)
27. Mahalle, P., Babar, S., Prasad, N., Prasad, R.: Identity management framework towards internet of things (IoT): Roadmap and key challenges. In: Proceedings of the 2010 International Conference on Network Security and Applications. pp. 430–439. Springer (2010)
28. Meier, S., Fuchs, L., Pernul, G.: Managing the access grid - a process view to minimize insider misuse risks. In: Proceedings of the 11th International Conference on Wirtschaftsinformatik. pp. 1051–1065 (2013)
29. Moghaddam, F., Wieder, P., Yahyapour, R.: A policy-based identity management schema for managing accesses in clouds. In: Proceedings of the 8th International Conference on the Network of the Future. pp. 91–98. IEEE Computer Society (2017)
30. Morisset, C., Sanchez, D.: Visabac: A tool for visualising ABAC policies. In: Proceedings of the 4th International Conference on Information Systems Security and Privacy. Newcastle University (2018)
31. Nuss, M., Puchta, A., Kunz, M.: Towards blockchain-based identity and access management for internet of things in enterprises. In: Proceedings of the International Conference on Trust and Privacy in Digital Business. pp. 167–181. Springer (2018)
32. Nyanchama, M., Osborn, S.: The role graph model and conflict of interest. ACM Transactions on Information and System Security (TISSEC) **2**(1), 3–33 (1999)
33. Pfitzmann, A., Hansen, M.: Anonymity, unobservability, pseudonymity, and identity management–a proposal for terminology. Lecture Notes in Computer Science **2009** (2004)
34. Povilionis, A., Arcieri, F., Talamo, M., Ananth, I., Schunck, C., Rosengren, P., Thestrup, J., Richter, J., Chiaravalottik, A., Schillaci, O., Gappa, H., Velasco, C.: Identity management, access control and privacy in integrated care platforms: The PICASO project. In: Proceedings of the 2018 International Carnahan Conference on Security Technology. pp. 1–5. IEEE Computer Society (2018)
35. Samarati, P., de Vimercati, S.C.: Access control: Policies, models, and mechanisms. In: International School on Foundations of Security Analysis and Design. pp. 137–196. Springer (2000)

36. Sandhu, R.S.: The authorization leap from rights to attributes: Maturation or chaos? In: Proceedings of the 17th ACM symposium on Access Control Models and Technologies. pp. 69–70. ACM (2012)
37. Sandhu, R.S., Coyne, E.J., Feinstein, H.L., Youman, C.E.: Role-based access control models. Computer **29**(2), 38–47 (1996)
38. Servos, D., Osborn, S.L.: Current research and open problems in attribute-based access control. ACM Computing Surveys **49**(4), 1–65 (2017)
39. Severino, R.: The data visualisation catalogue. `https://datavizcatalogue.com/index.html` (2019), accessed: 2019-02-21
40. Small, M.: Kuppingercole report - advisory note - big data security, governance, stewardship (2018)
41. SOX: Sarbanes-Oxley Act of 2002, pl 107-204, 116 stat 745 (2002)
42. Staheli, D., Yu, T., Crouser, R., Damodaran, S., Nam, K., O'Gwynn, D., McKenna, S., Harrison, L.: Visualization evaluation for cyber security. In: Proceedings of the 2014 IEEE Symposium on Visualization for Cyber Security. pp. 49–56. ACM (2014)
43. Tolbert, J.: Kuppingercole report - advisory note - identity in iot (2017)
44. Wagner, M., Rind, A., Thür, N., Aigner, W.: A knowledge-assisted visual malware analysis system: Design, validation, and reflection of kamas. Computers & Security **67**, 1–15 (2017)
45. Windley, P.J.: Digital identity: Unmasking identity management architecture (IMA). O'Reilly Media, Inc. (2005)
46. Xiong, J., Yao, Z., Ma, J., Liu, X., Li, Q., Ma, J.: PRIAM: Privacy preserving identity and access management scheme in cloud. KSII Transactions on Internet & Information Systems **8**(1), 282–304 (2014)
47. Xu, Z., Stoller, S.D.: Mining attribute-based access control policies from RBAC policies. In: Proceedings of the 10th International Conference and Expo on Emerging Technologies for a Smarter World. IEEE (2013)