Attribute Quality Management for Dynamic Identity and Access Management ☆

Michael Kunz[a,*], Alexander Puchta[a], Sebastian Groll[a], Ludwig Fuchs[b], Günther Pernul[a]

[a] *University of Regensburg, Universitätsstr. 31, 93053 Regensburg*
[b] *Nexis GmbH, Franz-Mayer-Str. 1, 93053 Regensburg*

Abstract

Identity and access management (IAM) has become one main challenge for companies over the last decade. Most of the medium-sized and large organizations operate standardized IAM infrastructures in order to comply with regulations and improve the level of IAM automation. A recent trend is the application of attribute-based access control (ABAC) for automatically assigning permissions to employees. The success of ABAC, however, heavily relies on the availability of high-quality attribute definitions and values. Up to now, no structured attribute quality management approach for IAM environments exists. Within this paper, we propose TAQM, a comprehensive approach building on a tool-supported structured process for measuring and improvement of IAM data quality. During the evaluation of three real-life use cases within large industrial companies we underline the applicability of TAQM for the identification and cleansing of attribute errors by IT and non-IT experts as well as the general introduction of quality management processes for IAM.

Keywords: Identity Management, Identity and Access Management, Access Management, Attribute Quality, Quality Management, Attribute-based Access Control, ABAC

1. Introduction

In order to provide secure and compliant access to IT resources, centralized identity and access management (IAM) has become one of the main challenges for companies. The successful fulfillment of existing compliance requirements is one of the core drivers when implementing IAM infrastructures and processes.

☆This work was partially supported by the German Federal Ministry of Education and Research (BMBF) within the collaborate research project DINGfest

*Corresponding author

Email addresses: michael.kunz@ur.de (Michael Kunz), alexander.puchta@ur.de (Alexander Puchta), sebastian.groll@ur.de (Sebastian Groll), ludwig.fuchs@nexis-secure.com (Ludwig Fuchs), guenther.pernul@ur.de (Günther Pernul)

While at the beginning only a small number of compliance standards and regulations had to be met (e.g. SOX [1], Basel II, Basel III [2]), nowadays, governments and organizations are more and more imposing compliance requirements that can only be governed by standardized IAM processes, guidelines, and technologies [3]. Furthermore, while initially only basic automation features were relevant, benefits related to process acceleration by facilitating efficient IAM measures play an increasing role for modern companies [3].

Today's IAM solutions already allow for a successful automation of most user administration workflows, offer dedicated functionality for security analyses, and at the same time deliver federation services and cloud-integration. They mostly employ role-based access control (RBAC) [4] as their underlying access control model. RBAC allows for a reduction of complexity by bundling permissions and employees into roles [5]. However, at the same time this can lead to steadily increasing role numbers and role administration efforts. In order to overcome these limitations, recent research as well as practical implementations are following the notion of attribute-based access control (ABAC) [6]. In contrast to RBAC, ABAC is more flexible and allows for the depiction of both, fine-granular and coarse-grained access rules [7]. ABAC checks the values of subjects', objects' or environmental attributes against pre-defined rules and allows or denies access based upon the fulfillment of these. Correctly maintained attributes (such as employees' business functions) do not only simplify entitlement provisioning but also support the discovery of violations of the least privilege principle. IAM workflows such as on-/off-boarding/movers of employees are easier covered by policies based on attributes rather than using static roles.

The success of any ABAC implementation, however, heavily relies on the underlying processes for a structured management of attribute definitions and attribute values. This task has not been receiving much attention within the research community up to now. Only few authors have pointed out that successful attribute management is a mandatory requirement for dynamic systems relying on attributes (e.g. [8]). Erroneously assigned attribute values can lead to unwanted access, effectively representing security risks and ultimately allowing intentional or unintentional abuse by insiders. Research offers several general data quality frameworks. However, these approaches do not offer the guidelines nor the fine-grained implementation details needed (e.g. quality metrics) for improving attribute quality in the context of IAM. They typically give high-level, non-IAM related guidance regarding the structure of quality improvement processes and do not provide specific recommendations or tool-support. At the same time, existing data quality metrics only provide generic support for attribute quality measurement and are not embedded into a process-oriented attribute quality concept. None of the existing approaches, for instance, offers a consistent overview of existing attributes within the various application systems connected to a centralized IAM infrastructure. They do not provide information about which attributes are incorporated into access policies, which access policies should be re-engineered, and which attribute values need to be investigated due to low data quality.

We argue, that due to the severe IT security risks imposed by low attribute quality, a structured and applicable approach to attribute quality management for IAM is needed. We thus introduce TAQM (**T**otal **A**ttribute **Q**uality **M**anagement), a dedicated attribute quality improvement approach tailored to the characteristics of IAM and attribute-ttribute-based access control in the remainder. In order to do so, we firstly introduce a generic, conceptual IAM model as the foundation of our research activities and to shape the scope of TAQM. Secondly, we analyze existing data quality management approaches regarding their suitability for IAM environments. Based on the results, we propose our novel approach TAQM for assessing, maintaining, and improving IAM attribute quality (Section 4).

The remainder of this paper is structured as follows. In Section 2, we outline our addressed problem and related work is provided. Section 3 describes our used research methodology as well as our contribution to the field. Afterwards, we present the IAM model (Section 4) which supports in discovering evaluation criteria for the analysis of existing data quality management frameworks in Section 5. Consequently, a comparison and selection of those frameworks is carried out in the same chapter. After this step, our attribute management approach is presented in Section 6 by integrating core elements from the chosen quality management frameworks with automation tools and procedures developed based on our IAM experience from industry projects. Within Section 7, the prototypical implementation of selected measures and optimization tools as part of the existing IAM analytics & cleansing platform *Nexis Controle*[1] together with a feasibility analysis within a real-world use case evaluation is provided. We conclude with known limitations of our approach and provide an outlook for further research in Section 8.

2. Problem and Related Work

Extensive research considering IAM processes, IAM policies and their implementation, as well as the underlying access control models has been carried out in the past [9]. RBAC, for instance, has evolved as the de facto standard for managing the access of thousands of employees to IT resources in many companies [10]. Following this concept, permissions are bundled into roles which are subsequently assigned to employees. This reduces administrative efforts but at the same time can lead to a steadily growing number of roles [11] while offering only a limited flexibility regarding contextual changes (e.g. departmental changes of employees) [12]. Furthermore, studies have shown that RBAC implementation costs an average of 2,410,000$ for a company of 10,000 employees [13]. As a result, ABAC has gained attention in both, research and practical application over the last years. ABAC leverages attribute definitions in order

[1] `https://www.nexis-secure.com`

to model dynamic access management policies based on attribute values of entities like employees or permissions. Initial ABAC approaches were introduced by Priebe et al. [14] and Yuan et al. [15]. A more comprehensive view on ABAC has been given by Hu et al. [6]. Research already provides approaches for initially developing or re-designing policies for IAM in a time-efficient and complexity-reducing manner (cf. [16, 3]). However, ABAC models heavily rely on the completeness and correctness of the underlying attribute values used within those policies. Consequently, a structured approach for maintaining attribute data quality is needed by organizations utilizing ABAC. Consider the following simple example[2] clarifying the addressed problem of attribute quality within ABAC environments:

Table 1 deals with different data quality problems relevant for IAM. It shows an excerpt of identities within an IAM system having the attribute "location" (working place of the employee) and "cost center" (used for internal accounting). Consider an additional ABAC policy granting access to the relevant file storage if and only if the employees' "location" equals *Munich*. One can identify two typical data quality problems within the table, as one "location" is shortened to *MUC* for employee #2 while #3 is completely missing a "cost center". According to the existing ABAC policy this heavily restricts the access to relevant resources for employee #2 as he does not fulfill the policy. Errors like these can for example arise if entries are inserted manually (e.g. by HR staff entering wrong identity information). Additionally such attribute data often does not get revised as it is seen as an unnecessary or too extensive task leading to a declining attribute quality. Thus ABAC can not be applied efficiently and identities are hindered in executing their work or circumvent such policies through direct assignment of permissions which can negatively affect security. In order to solve such problems this paper applies an attribute data quality management approach specifically tailored to existing IAM requirements.

Table 1: Example for IAM data quality problems

ID #	First Name	Last Name	Location	Cost Center
1	Yasmin	Olivid	Munich	Cost Center 1
2	Henry	Zellers	MUC	Cost Center 1
3	Charles	Ellsworth	Munich	

Over the last decades, a large body of work has been conducted in the field of quality management in general. Various notions of quality can be differentiated, e.g. *quality management* in general, *data quality management*, and *attribute quality management* [17, 18]. General quality management focuses on the quality of physical products while data quality management is specifically

[2]In the remainder of this article we mainly exemplify attribute quality for identities (e.g. employees) as this can be understood quite intuitively. However, all relevant elements within IAM may have respective attributes and are concerned with attribute quality (cf. Section 4)

dealing with managing the quality of structured data [19]. Additionally, within the research community the term information quality management is used for bundling activities related to the quality management of unstructured data [19]. However, as IAM in general handles structured data (e.g. employees and their master data, permission master data and assignments), information quality is of minor importance for our research.

Especially within the field of general quality management, extensive research has been published [20, 21]. Pioneering in the area of total quality management, Deming laid the foundation for modern quality management [17]. The defined principles are still incorporated into several quality management systems like the Baldrige Performance Excellence Program[3]. In addition, other approaches like lean management or six sigma were introduced [22, 23]. They added further dimensions to Deming's total quality management. However, none of these approaches provides a comprehensive integration of the notion of data quality [18]. As a result, dedicated data quality approaches try to cover these aspects by focusing on digital data quality. Batini et al. provide a comprehensive overview of existing data quality approaches [19] which provides the baseline for our overview of approaches in Section 5. Beside the existing data quality frameworks, researchers also have dealt with various data quality metrics over the last years. Using statistical or mathematical concepts, the quality of information and data sets can, for instance, be determined and compared [24]. There is a set of metrics defined for quantitative measurement of data quality. Examples would be metrics for consistency or timeliness [25, 26, 24] or techniques for detecting duplicate data entries [27]. They are also partly considered in the data quality frameworks mentioned above. Of course there exist many more different metrics yet we just wanted to give a hint on this topic.

However, attribute quality management has not been researched in the past to a sufficient extent. A recent identification of areas of research within ABAC does not even list attribute quality management as an individual category [28]. Other authors are aware that data quality is connected with attributes within IAM but do not provide any valid solutions [29, 30]. None of the provided approaches satisfies the requirements of IAM environments: While quality metrics rather aim at providing isolated mathematical means instead of dealing with structured data quality management processes, existing quality management or ABAC approaches fail to provide concepts for the evaluation, strategic management, and optimization of attribute definitions and attribute values. Consequently, organizations are missing a comprehensive approach for measuring, maintaining, as well as improving their IAM attribute quality.

In the following we aim at closing this research gap by defining a general attribute quality management approach for IAM environments. It does not only provide a generic high-level methodology but also integrates specific metrics for attribute quality handling. By doing so, it satisfies the demand for providing metrics suitable for attribute quality evaluation within ABAC while at the same

[3] `https://www.nist.gov/baldrige`

time offering an integrated process-oriented approach that can be applied to large-scale IAM scenarios.

3. Methodology

The underlying research methodology is displayed in Figure 1 and based on the principles of Hevner et al. [31]. Our facilitated knowledge base covers the foundations and methodologies from the two fields of IAM and data quality management. Business needs for increasing attribute quality within IAM systems are the baseline for our environment. On this basis we design a novel approach for evaluating, maintaining, and optimizing IAM attribute quality. We experimentally evaluate a prototype implementation of our solution consequently.

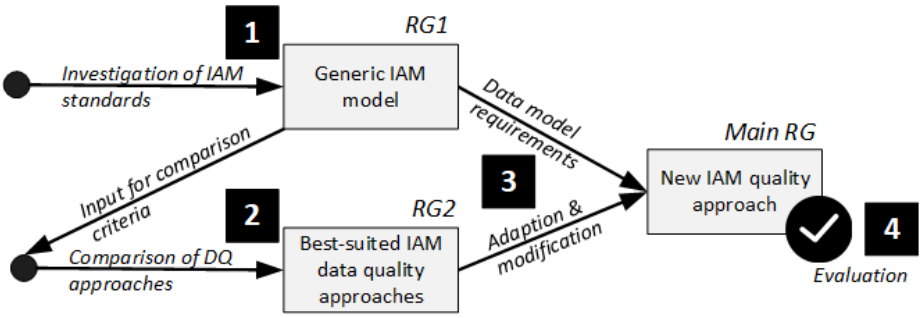


Figure 1: Applied research methodology throughout the paper

In order to achieve our research goals (RG), we firstly derive a generic conceptual IAM model (1) based on existing literature as well as project experience in order to get a comprehensive picture of IAM-relevant entities. The IAM model (RG 1) serves as input for establishing specific IAM quality requirements which can be used for the evaluation and comparison of existing data quality management approaches (2) regarding their suitability for IAM environments. The results (RG 2) then are inspected in detail and complemented with IAM-specific requirements, measures and optimization efforts (3). By combining elements of the approaches and specific requirements of our IAM model we generate a novel IAM attribute quality approach and a prototypical implementation (Main RG) which in turn is evaluated using real world data in order to demonstrate its feasibility and effectiveness (4).

4. Conceptual IAM Model

In the following, we compose a basic IAM model including all main entities relevant within the context of IAM. Note that in this paper, the term entity refers to any object whose attributes or master data are managed or used by an IAM system. The proposed model represents a minimal approach of mandatory

elements integrated in modern ABAC-based IAM environments and thus acts as the foundation of our novel IAM data quality management approach. Additionally, it represents the scope that needs to be addressed by the new approach. In order to come up with such a conceptual IAM model, we investigated the most relevant IAM standards and technologies: LDAP [32], SAML/Shibboleth [33], OAuth [34], SPML [35] and XACML [36] (for an overview cf. [37]). In order to reflect its practical relevance, SCIM [38], a relatively new industry standard already adopted by existing IAM products, was additionally included (even though it has not been listed in the aforementioned survey). For each standard, we extracted the covered entities relevant for IAM environments together with their mutual relationships and listed them in Table 2 within the third column. During a manual pre-selection process, standards which do not include any entities relevant for a comprehensive conceptual IAM model due to a different application focus (i.e. WS-Federation, CoSign, OZ, CAS, OIDC and Kerberos) have been excluded from further analyses. Kerberos, e.g., focuses on a client-server communication protocol rather than describing IAM entities. Furthermore some of these excluded standards are based on other standards (e.g. OIDC is based on OAuth) and would only include already identified IAM entities. In addition, several sub-models of standards like RBAC have been developed. They introduce additional concepts such as task-inclusion [39] or dedication to an organizational-structure [40] into the original RBAC model [9]. However, the goal of our conceptual model is to present the basic entities relevant for IAM in order to act as starting point for general quality management criteria - and not to act as generic model covering each IAM application scenario.

Based on all the input within Table 2 we crafted a more generalized model with the help of "derived IAM entities" and defined mandatory (digital identity, account, permission) and optional (role, policy, context, attribute) entities (cf. Figure 2 for the overall model). We made the transition from relevant entities to derived IAM Entities by listing all named entities. We then grouped the entities according to the following elements based on the respective standard description and our experience in IAM. For instance, LDAP speaks of persons and organizational persons. We therefore created the entity "(Digital) Identity" and matched it with the LDAP-specific terms and continued this for all standards:

(Digital) Identity. In organization-wide IAM scenarios, digital identities are the representation of human users, e.g. within the personnel management system (cf. [41, 42]). Some organizations operate multi-identity models in which one real-world employee is represented by several digital identities (sub-identities).

Account. Digital Identities are in turn represented within the target systems by user accounts (following the concepts from LDAP or SCIM). Note that several existing standards (e.g. SAML, SPML, XACML) do not differentiate identities and accounts because of their limited application-specific focus. Within IAM,

Acronym	Description	Relevant Entities	Derived IAM Entities
LDAP	Directory protocol for storage of user accounts and group memberships [32]	person, user, inetOrgPerson, organizationalPerson, group, groupOfNames, groupofuniquenames	Digital Identity, Account, Permission
SAML / Shibboleth	Exchange format for authentication and authorization information. Also attribute information can be exchanged. [33]	Subject, Attribute, AttributeQuery	Identity, Account, Attribute, Policy
SPML	Request and response protocol for provisioning information of accounts and resources [35]	Requestor, Provisioning Service Object, QueryClauseType	Identity, Permission, Role, Policy
OAuth	Standard for access delegation, typically for website access [34]	Resource	Role, Permission
SCIM	Standard for exchanging user and employee information within the cloud [38]	Identity, User, Singular Attribute, Multi-valued Attribute, Simple Attribute, Complex Attribute, Group	Digital Identity, Account, Attribute, Permission, Role
XACML	Standard for a policy language, that allows for a fine-granular and attribute-based expression of policies. Also offers an architecture and protocol for the interpretation of the language [36]	Attribute, Context, Policy, Policy set, Rule, Subject, Resource	Attribute, Context, Policy, Digital Identity, Account, Permission, Role
RBAC	Access control model that is currently de facto standard. It groups identities/accounts and resources/permissions together to roles [4]	User, Permission, Session, Role	Digital Identity, Account, Permission, Context, Role, Permission
ABAC	Access control model that founds authorization decisions on rules based on attributes [8]	Subject, Resource, Attribute, Environment Condition, Policy, Rule	Digital Identity, Account, Attribute, Permission, Role

Table 2: Table with investigated standards for the creation of a basic conceptual IAM model

an account represents the entity with which an employee logs on to a specific application system (e.g. LDAP Directory, SAP, etc.).

Permission. Accounts are typically assigned to permissions in order to grant access to certain IT resources. A permission is considered as any access right within a specific application system, irrespective of its granularity. It consists of an authorization statement and an object this authorization is granted for (e.g. following the original RBAC model [4]) and can be hierarchically aligned. For instance, a file share "Marketing Campaign 2018" can be typically accessed via read, modify or update rights. In this case, "Marketing Campaign 2018 - read" is an example for a permission. This permission can optionally be nested into another more generic permission "All Marketing Campaigns - read".

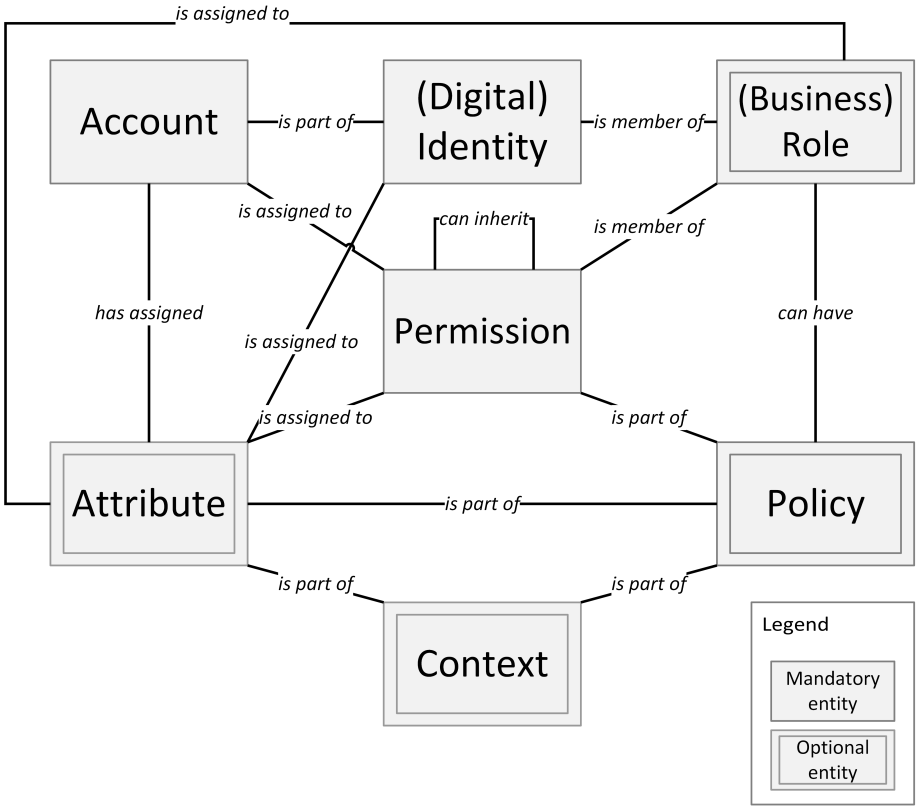


Figure 2: Conceptual IAM model based on IAM standards

Business Role. RBAC as the currently predominant access control model defines the concept of roles, which are interpreted as business roles in the IAM environment. Business Roles are used to bundle permissions from target systems (e.g. SAP roles, Microsoft Active Directory groups, or mainframe profiles) into

a business-relevant object (e.g. business function of an employee). In contrast to permissions, they are typically instantiated within IAM systems in order to allow an system-independent view on a meta-level.

Policy. A policy represents a rule or a set of rules that defines whether a role or permission is assigned to an account or identity. This follows the definitions known from ABAC, SPML and XACML.

Context. Context represents the relevant scenario a policy can be part of. In terms of ABAC, for instance, environmental conditions such as the time zone or geologic location during a log-in represent a certain context.

Attribute. Attributes represent the meta-data related to accounts, identities, roles, policies, and permissions and are in turn assigned to a specific context. SCIM, XACML, and ABAC are, for instance, heavily depending on attributes. Table 3 shows typical examples of attributes for the entities defined in our model. Note that companies usually define additional custom attributes to fit their needs.

Table 3: Typical examples for Attributes in the IAM Context

Entity Type	Attribute Example
(Digital) Identity	Department, Job Title, Location
Account	Account Type (e.g. Admin, User), Target System
Permission	Criticality, Target System
(Business) Role	Criticality, Business Function

Contribution to the attribute quality approach. With the help of the IAM data model we clarify the entities and the respective attributes which have to be considered for an attribute quality management approach. Using this model one can easily see which entities of IAM have to be included into a comprehensive attribute quality management approach and thus defines its scope. Otherwise important aspects (e.g. context information) do not get included in such approaches when it comes to practical implementation. The model itself also serves as an important requirement for the approach and is used as evaluation criterion for the existing data quality management approaches in Section 5. Finally we ensure that the underlying data model of our approach is correct and includes all relevant objects used in relevant IAM standards.

5. Selection & Evaluation of Data Quality Management Approaches

In the following we investigate existing data quality management approaches in respect to their suitability to serve as foundation for structured IAM quality management. The evaluation is then executed on the basis of criteria derived from both, our IAM data model from Section 4 and general IAM conditions.

5.1. Selection of Data Quality Management Approaches

During a first step, we collocate a list of potentially suitable approaches known from literature. Batini and Scannapieco provide a comprehensive overview of traditional approaches for data quality management [19]. Note, that two approaches mentioned in their publication were excluded of our analysis: CIHI is an approach [43] purely focusing on administrative databases within the Canadian health care sector. Similarly, ISTAT [44] is solely used for Italian public administrations and the improvement of address data of citizens and businesses. Due to their specific application scenario, both approaches are not suitable for our purpose. Furthermore, we extended the list of [19] with QIAM, to the best of our knowledge, the only approach specifically focusing on data quality within IAM so far [45]. The full list of investigated data quality management approaches can be found in table 4.

Table 4: Overview of data quality management approaches

Data Quality Management Approach	Reference	Focus
AIMQ	[46]	Data Quality Assessment
AMEQ	[47]	Data Quality for Mechanical Products
CDQ	[48]	Generic Data Quality Framework
COLDQ	[49]	Costs of Low Data Quality
DaQuinCIS	[50]	Data Quality for Cooperative Information Systems
DQA	[51]	Data Quality Assessment
DWQ	[52]	Data Quality for Data Warehouses
IQM	[53]	Data Quality for Web Data
QAFD	[54]	Data Quality for Financial Data
QIAM	[45]	Data Quality for IAM
TDQM	[18]	Generic Data Quality Framework
TIQM	[55]	Generic Data Quality Framework

5.2. Description of Data Quality Management Approaches

In the following the aforementioned approaches are briefly introduced. *AIMQ* introduces a questionnaire to collect and analyze data. Based on these results, activities for improvement are identified. Within *AMEQ* an activity-based approach is used to measure data quality for mechanical products. *CDQ* consists of a comprehensive data quality framework based on business processes. Within this approach, the effects of low data quality on business processes are analyzed. *COLDQ* investigates the cost of low data quality for organizations by means of a scorecard based approach. *DaQuinCIS* focuses on cooperative information systems with an e-government context. Their D^2Q model is used to define data sets and quality properties. The *DQA* framework focuses on the identification

and comparison of subjective perceptions of individuals and objective measurements of data quality. Based on this, differences regarding the quality can be deduced resulting in tasks for improvement. The *DWQ* approach is one of the first methods related to data quality management. It is centered around data quality for data warehouses and leverages queries to determine the quality of a data warehouse. On the contrary *IQM* concentrates on web-specific data quality. This approach combines different existing tools for websites to determine the data quality. *QAFD* is used as a framework for data quality regarding financial data. It introduces initial quality measures for financial application scenarios. *QIAM* is the first approach to introduce structured data quality management in IAM. However, it is specifically focused on data quality for role-based access control systems, essentially limiting its view and applicability in ABAC environments. The *TDQM* approach is the first comprehensive data quality framework to be introduced. It is based on a process-driven cycle analogously to the Total Quality Management by [17]. Another generic data quality framework is represented by *TIQM*. In addition to commonly used processes within data quality management, *TIQM* establishes a specific process for culture transformation.

5.3. Definition of Comparison Criteria

Within this and the following sections the previously introduced quality management approaches are compared regarding their applicability in IAM environments (see Table 6).

Our evaluation is based on criteria specifically relevant within attribute-based IAM environments. These criteria can be understood as IAM-specific requirements for attribute quality and have been derived from both, research publications and experience from various real-life IAM projects as well as from our IAM data model (cf. Section 4). Note that our goal was not to define an exhaustive list of criteria, but rather to come up with a selection of requirements of major importance for IAM. We argue that only quality management approaches that cover those basics are suitable for application. Table 5 lists the selected comparison criteria, followed by a short description and discussion.

Focus on Attribute Quality. Approaches suitable for IAM environments need to be centered around data quality and its improvement. Several existing approaches only deal with the flow of information and do not explicitly focus on the quality of (data) attributes (e.g. [49]). This, however, is a core requirement within data-driven IAM environments in which attributes are critical for the interpretation of attribute-based access policies (e.g. correctness of employee attributes or entitlement attributes).

Governance. Governance is of high importance in today's IAM infrastructures for organizing and structuring an IAM's performance. Therefore, a data quality approach for IAM needs to be capable of integrating organizational responsibilities and tasks during quality measurement and improvement. For instance, actions regarding quality management need to be audited in a secure manner (e.g. who decided which employee was assigned to a certain attribute value).

Table 5: Used criteria

Criteria	Description
Focus on Attribute Quality	Data-centered perspective on attribute quality instead of e.g. concentration on data flows
Governance	Integration of long-term management processes such as responsibilities for entities and related tasks
Iterative Approach	Ongoing process-oriented approach with repetitive phases for incremental improvement of data quality
Granularity	Level of detail of quality measures of the respective approaches
IAM Content	Inclusion of IAM relevant topics and requirements within the framework
IAM Completeness	Capability to integrate all entities known from the conceptual IAM model

Thus the notion of responsibilities by IT- as well as non-IT staff for actions and tasks needs to be considered.

Iterative Approach. Data quality management in IAM environments needs to be based on a strategic and iterative methodology as entity-related and organizational structures constantly change during the development of an organization. It has to be capable of integrating new data elements (e.g. new IT systems getting on-boarded) throughout an evolving attribute quality management cycle. As a result, every suitable approach has to provide a structured and iterative process for data quality (e.g. a PDCA (Plan-Do-Check-Act) cycle).

Granularity. Granularity describes the level of detail a certain quality management approach incorporates. While many deliver a generic process model (e.g. [49, 53, 51]), specific details of actual technical measures are often missing. For ABAC, implementation guidelines with a deep level of technical context already exist [56, 57]. We argue that for successful adoption, combining both worlds is required. While high-level processes can guide strategic development, technical implementation details offer guidance and ease real-life implementation.

IAM Content. Within this criterion, we assess the extent to which an existing approach already covers IAM-specific topics. There is already a wide range of different requirements relevant for IAM (e.g. Identity Life Cycle Management) that can be leverage for this criterion [58, 59]. In case an existing approach covers or addresses certain IAM functionalities out of the box, it might be adopted easily to form a basis for structured and generic quality management in IAM environments. Please note that this criterion is aggregating a range

of different topics to avoid defining a single criterion for each specific IAM functionality.

IAM Completeness. Existing approaches typically cover general data quality requirements or data quality dimensions [18, 46]. In respect to IAM, they need to be able to cover all entities and relationships included within the previously presented conceptual IAM model. The entities should either be explicitly mentioned or the approach should at least be flexible enough to consider all entities. The same holds for all relationships between the entities (e.g. the binding of attributes to permissions). As the introduced conceptual IAM model represents a minimal approach of mandatory elements, we argue an existing approach needs to be capable to deal with those entities in order to be suitable for structured long-term quality management.

5.4. Evaluation Summary

In order to evaluate the presented attribute quality management approaches we applied a Likert Scale from 0 to 4 points (equally from ○ to ●) [60]. For each criterion we rated the approaches relatively to each other and selected the top-rated three approaches (i.e. the top quartile) for further usage during our research. The final analysis is displayed in Table 6 with the top quartile approaches marked in green. They are selected to act as baseline for our new quality management approach presented in the following section.

Table 6: Comparison of data quality management approaches

Approach	Focus on Attribute Quality	Governance	Iterative Approach	Granularity	IAM Content	IAM Completeness
AIMQ	◑	◔	○	◔	◔	◑
AMEQ	◕	○	◕	◕	○	○
CDQ	◕	◔	◕	◑	◑	◑
COLDQ	◔	◑	◑	○	○	○
DaQuinCIS	●	○	●	◔	◔	◔
DQA	●	◔	○	◔	◔	◑
DWQ	◕	○	○	◑	○	◔
IQM	◑	○	●	○	○	○
QAFD	◑	◔	◕	◑	○	○
QIAM	◔	◕	●	◔	●	◕
TDQM	●	◕	●	◔	◑	◕
TIQM	◕	◕	●	◑	◑	◕

While generic requirements like *Focus on Attribute Quality* and *Iterative Approach* are rated quite well on average, IAM-specific requirements (e.g. *Governance* or *IAM Content*) are not addressed sufficiently.

The results reveal three existing approaches with a similar performance, dominating the other quality management approaches: TDQM, TIQM (each 17 points), and QIAM (16 points). Therefore we did not decide to take only one approach and improve it with attribute quality techniques as we did not want to resign the benefits of the other two well rated approaches. However, none of those approaches cover all IAM-related requirements to a sufficient extent. QIAM representing an IAM related approach for example lacks the focus on attribute quality as it was mainly developed for a role based environment (e.g. for SAP roles). Furthermore it does not offer an existing toolset of metrics like other approaches do. Regarding TDQM and TIQM it is exactly vice versa as they lack sufficient IAM content and granularity.

6. The TAQM Approach

The comparison and evaluation of existing frameworks above has shown the inability to cover core requirements needed for structured attribute quality management in IAM (cf. Section 5). To overcome this gap we present TAQM (**T**otal **A**ttribute **Q**uality **M**anagement), a novel attribute quality management approach for IAM environments (see Figure 3). Its main characteristics are:

- Cyclic execution in order to cope with the dynamic nature of IAM data
- (Semi-)automated tool support (e.g. during discovery of attribute errors)
- Integration of human experts to foster existing organizational knowledge
- Data-centricity that focuses on correctness of data in order to support the improvement of quality management and policies
- High-level structure and low-level guidelines for a fast and easy deployment

TAQM is aiming at supporting both, the technical and organizational nature of IAM. It follows the core concepts for IAM data quality presented by Fuchs and Pernul (QIAM, [45]) as well as those of the TDQM and TIQM frameworks by Wang et al. [18] and English et al. [55]. Following Wang et al.'s well-respected method of defining, measuring, analyzing, and improving quality, TAQM incorporates four cyclic phases (see Table 7). Additionally, it employs the process orientation known from TIQM in order to structure each execution phase and introduce automation support. Furthermore, we complement TAQM with IAM-specific activities and supportive automation tools in order to increase adaptability. Each of its four main phases thus provides low-level implementation guidance in order to overcome the limitations of already existing purely high-level approaches.

Note that the automation techniques presented are not exhaustive and might be extended in future work. However, we argue that they already cover the basic required tasks and thus present a valid baseline for TAQM implementation. The evaluation in Section 7 underlines this assumption in real-life scenarios.

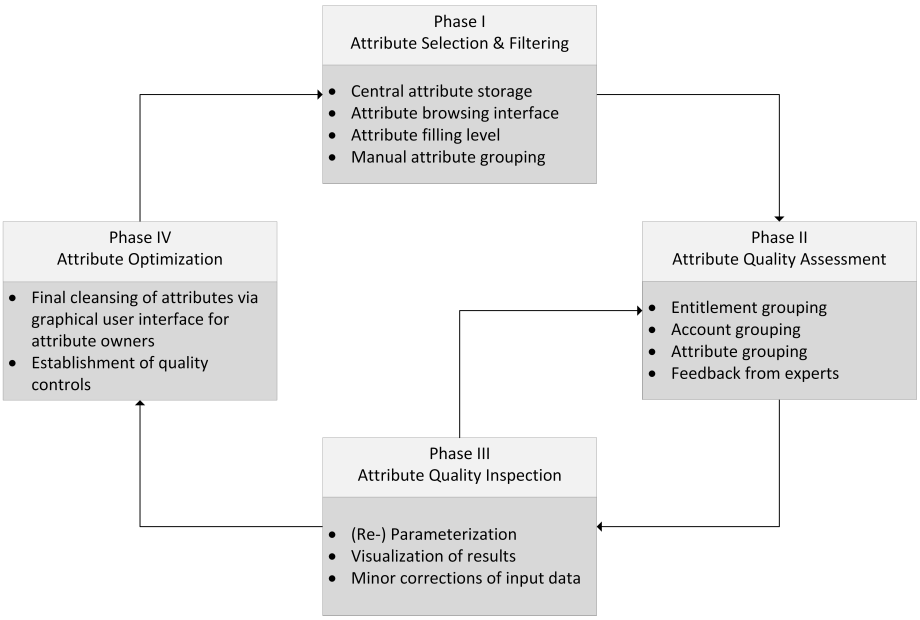


Figure 3: Proposed TAQM approach

Table 7: Phases of TAQM

TAQM Phase	Corresponding Phase in [18]
Attribute Selection & Filtering	Define Phase
Attribute Quality Assessment	Measure Phase
Attribute Quality Inspection	Analyze Phase
Attribute Optimization	Improve Phase

6.1. Attribute Selection & Filtering

During this initial phase, human experts select the set of attributes and IAM entities relevant for a specific company (cf. Figure 4). For instance, attribute-based access policies often rely on employees' attributes such as the employee type or work location while permissions usually carry ownership or information like their risk-rating. In order to allow for an efficient selection process, TAQM suggests the following automation tools during Phase I:

- Central attribute storage based on the proposed conceptual IAM model(cf. Figure 2)
- Attribute browsing and selection interface
- Attribute filling level analysis
- Manual attribute grouping functionality

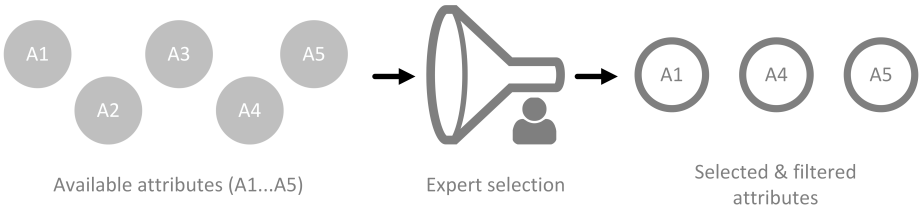
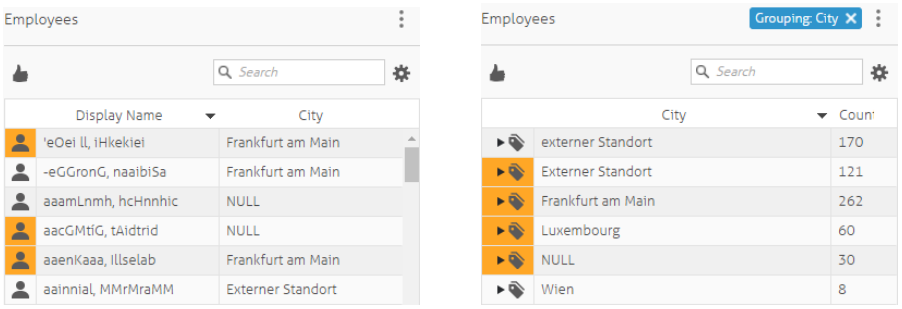


Figure 4: Simplified overview of the *Attribute Selection & Filtering* phase

First of all, relevant attributes need to be loaded into a centralized attribute storage (cf. Section 4). This storage is responsible for handling the connection between entities and their attributes and acts as basis for data analysis measures. Note that existing IAM systems already offer a centralized database for managing the execution of operational IAM processes like the onboarding or offboarding of employees. They, however, lack the functionality of structured attribute management including data browsing, analysis, and dedicated attribute-related processes. We thus argue existing IAM implementations need to be extended with a dedicated centralized attribute storage for this purpose. To overcome this problem the attribute storage can be used as the master system regarding all attributes while IAM systems just need to manage the attribute values accordingly (e.g. the attribute storage defines a certain range of valid values. This range could be queried by the IAM system and the values can be saved respectively). Additionally traditional IAM databases can not be changed so easily compared to a dedicated and separated attribute storage. We recommend to design this storage based on our proposed conceptual IAM model (cf. Figure 2). The model provides a consistent foundation for executing the steps of TAQM and contains the required entities and relations. From a technical point of view it ensures that the proposed visualizations and algorithms can be implemented and executed. Furthermore people applying TAQM should be familiar with the model, as it is based on various IAM standards. Therefore visualizations (grouping, aggregation, etc.) based on the model are easy to understand for human experts. This is especially important as experts have to make various decisions based on such visualizations while executing the steps of TAQM.

Secondly, data browsing interfaces are required to improve the attribute analysis and selection process by human experts. In real-life implementations with a high number of attribute definitions and values combined with several thousand instances (e.g. employees or permissions) such support is mandatory. The data browser should further be enriched with automated attribute filling level analyses (following the metric published by [61]) in order to detect major syntactic attribute quality issues like NULL-values. In many scenarios, empty attribute values are unwanted and hint at potential process flaws which prevent access control policies from correct interpretation. For extended discussions on such NULL-values and their handling, see Heinrich et. al [61].



(a) Employee attribute value list (b) Grouped employee attribute value list

Figure 5: Attribute NULL-value analysis (anonymized employee names)

Thirdly, data grouping functionality is required to support human experts when analyzing the distribution of attribute values and gain insight about possible source data quality issues (see [62]). It reveals syntactic data errors (i.e. typos, inconsistencies etc.) and allows for a semantic analysis of attribute value distributions. Our experience in practical projects revealed that many organizations are not aware of current attribute definitions and used attribute values. This typically stems from a long history of decentralized attribute management processes carried out for systems individually. Figure 5 visualizes a single-view (a) and grouped-view (b) example of employees' work location which could be integrated in a supportive tool for TAQM execution.

Finally, the centralized attribute storage should offer a graphical attribute selection interface. According to Wang et al. this supports human experts choosing the desired set of attributes relevant for further improvement [18]. During the first execution cycle of TAQM, we recommend to only select a basic set of the most important attributes. These attributes can subsequently be evaluated by applying different constraints like a maximum number of characters for a string attribute. As IAM is heavily relying on organizational anchoring [41], we argue that an ownership concept for IAM entities should be introduced within this phase (similarly to role owners as shown in [10]). An attribute owner acts as the primary contact person for an attribute and its values and is responsible for its maintenance. A similar definition can be also found in [18].

6.2. Attribute Quality Assessment

After attribute selection, an initial semi-automated assessment of the current attribute quality needs to take place in order to foster later human analysis during Phase III (depicted in Figure 6). The output are possible anomalies included within the analysis results. Note that TAQM also allows for a subjective manual quality estimation by attribute owners or responsible staff. Using this method known from [51] allows us to identify deviations from subjective and objective perceptions of attribute quality. However, in general, assessment automation typically is required in IAM scenarios with a very large number of

attribute values and assignments. As general data quality metrics already were

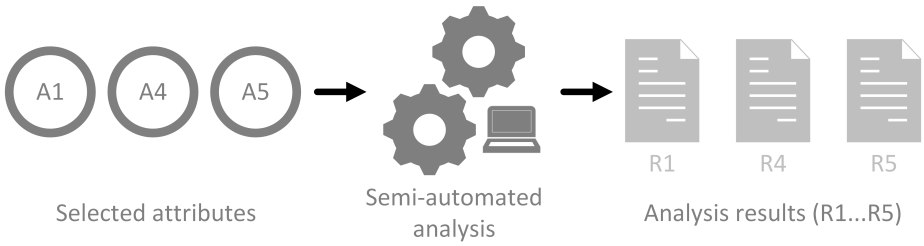


Figure 6: Simplified overview of the *Attribute Quality Assessment* phase

subject to profound research in the past [18, 46, 51], we are focusing on IAM-specific attribute assessment procedures in the following. In contrast to TDQM [18], for instance, we work with a predefined but yet extensible set of quality metrics specifically suited for IAM data. We rely on core concepts known from the field of role development [63], where clustering of employees according to access data is executed in order to identify suitable role candidates. However, in contrast to the limited role-oriented view in [63], we allow for the automated analysis of any conceptual IAM model entity. Attribute Quality Assessment is executed throughout three steps:

- Data grouping
- Data validity check
- Assessment execution

At first relevant data is grouped and correlation matrices are generated. These matrices are in turn validated for their applicability before a data quality assessment can take place.

Data Grouping. Firstly, all relevant entities are statistically grouped based on attribute values using two-dimensional matrices. The underlying assumption is that certain entity attribute values typically are well-managed due to their company relevance (e.g. "cost center" assignments of employees). Starting from such high-quality attributes, the highlighting of quality issues for other attributes can take place. Table 8 shows a practical example in which the number within each cell represents the number of employees having the same attribute values for both attributes. Imagine a "location" attribute of employees which has not been managed in a structured manner and data errors are expected. Following the example, employees might be grouped according to their "cost center" attribute which in turn is related to a second attribute dimension (e.g. the "location" attribute). Note that also missing attribute values can be handled, by collecting those NULL-values in a special group (see Figure 5). During the later assessment, outlier detection mechanisms can then highlight potential

Table 8: Simplified example for data grouping & validity check

		Attribute **Cost Center**	
		Cost Center 1	Cost Center 2
Attribute **Location**	Location A	98	100
	Location B	2	100
	Location C	0	50

data errors for the location attribute. Accordingly, other entities like entitlements, roles, and accounts can be grouped in order to detect quality issues using classification techniques known from [63].

Data Validity Check. After two-dimensional matrices for all relevant entities and their attributes have been created, their suitability for further analysis has to be confirmed. Two main issues limit the meaningfulness of a given matrix: i) Similar distribution of attribute values and ii) too few group memberships. In case of small groups (e.g. only two employees are assigned to a certain location; see location B in Table 8 row 2), further outlier analysis does not make sense. The same holds in case the attribute value distribution is not meaningful for a certain row or column in the matrix. Imagine the "Cost Center 1" consisting of 100 employees which all are assigned to different locations to the same extent. As a result, outlier detection is not able to determine a predominant location for this Cost Center. In order to automate the validity check, we apply threshold-based ratings related to group size and group distribution of each matrix row and column in order to define whether it is used during the subsequent assessment execution. More precisely, we validate if 50% of the most frequent attribute values cover 80% of the entities. Note that those thresholds can be configured according to the given scenario.

Assessment Execution. In the final assessment phase, TAQM identifies outliers and potential data quality issues. One automation technique, for instance, highlights all entities with attribute values in groups, where the overall distribution of the value is below a certain threshold (e.g. 5%). Following our previous example, Cost Center 1 consists of 100 employees out of which 98 are assigned to the location "A" while only two employees are assigned to "B" and none to "C". Those two outliers could hint at suspicious attribute values (e.g. an employee has a wrong location attribute). Contrarily, "Cost Center 2" would not indicate any anomalies. Similarly, imagine a number of financial-related permissions within a SAP system. A matrix calculated based on classification techniques reveals the various departments those permissions are used in. Based on that, it might be revealed that employees in the "Cost Center 2" department are wrongly assigned to those finance permissions.

6.3. Attribute Quality Inspection

After the assessment execution, the identified analysis results (containing possible quality issues) need to be reviewed by human experts in order to decide

if they indeed represent attribute quality errors or false positive alerts (see Figure 7). This can be cumbersome as a potentially high number of attribute analyses (each built on a different similarity matrix) needs to be conducted.

Figure 7: Simplified overview of the *Attribute Quality Inspection* phase

Typically, organizations attach a variety of different attributes to each managed entity within their IAM system. A large number of outliers detected within the numerous matrices might be the result. We hence argue that tool support is required for verification of outliers and suggest two techniques in order to achieve this:

- Visualizing coloured outlier matrices:
  Each of the matrices that resulted in suspicious attribute values can be visualized in a human-understandable form using outlier colouring. For example, orange or red highlighting (depending on the confidentiality level for a certain outlier), might hint at high-likely errors within the source data while green colouring shows standard attribute values. Group sizes and confidentiality ranges can further support human interpretation.

- An interactive grid visualization:
  Grid-based visualization techniques are able to display employee permission assignments within a two-dimensional matrix (cf. Figure 11). They have mainly been used for role development so far [64]. However, by highlighting specific attribute values of any entity, they allow for a contextual result analysis by offering data grouping, data coloring, or data filtering. For instance, grouping algorithms might discover results for a departmental head that has a single attribute value no one else is assigned to in his department. While this attribute value assignment might be conspicuous in terms of outlier detection, an interactive grid visualization can easily allow a human expert to identify this identity as departmental head (e.g. by displaying the department attribute of every identity).

Note that in comparison to TDQM [18], we allow a loop-back to Phase II at this stage. This enables an expert to quickly re-adjust data quality assessment mechanisms in case the outlier detection mechanisms have not been configured appropriately (e.g. include less attribute dimensions, change threshold values, or ask attribute owners to provide further semantic information about attribute values).

6.4. Attribute Optimization

After the identification of potential attribute quality problems and a first result validation by a human expert, the last phase of TAQM serves three main goals: On the one hand, it aims at cleansing identified errors (cf. Figure 8). On the other hand, it suggests the introduction of data quality standards as well as the set-up of strategic measures to maintain these quality standards.

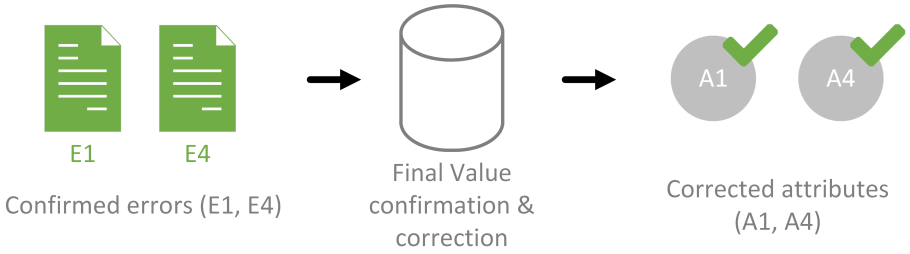


Figure 8: Simplified overview of the *Attribute Optimization* phase

Data Cleansing. Regarding data cleansing, human experts need to rate the outliers and propose correct attribute values or mappings for identified data errors. This could cover correcting current attribute values, assignments of user accounts to identities, or the clean-up of excessive permission assignments. Imagine a company where applications' user accounts have not been mapped to existing identities, i.e. employee master data from the HR system. During a first data cleansing cycle, automated analyses can identify which user account belongs to which employee based on attribute correlation regarding the different accounts, their assigned permissions, and the employee master data. After mappings have been automatically proposed (a user account with finance permissions in an SAP system could, for instance, be mapped to an employee within the finance department) and a human expert reviewed the results, a second execution cycle of TAQM might lead to further outliers which could not have been detected without this initial identity mapping (e.g. the attribute data of the user account is erroneous or the assigned permissions contain risks violating the principle of the least privilege).

Data Quality Standards. Regarding the establishment of and adherence to data quality standards, the knowledge of human experts can be fostered in order to gather semantic knowledge about the data. They can identify whether potential errors have technical or organizational reasons and make recommendations for data quality standards. Manually and decentralized attribute management, for instance, is very likely subject to a higher error rate than automatically derived attributes by a centralized department. As a result, not only the one-time clean-up of data errors, but a change in organizational or technical processes is required. In case organizational errors are the reason for user account attribute

issues, for instance, an IAM manager might define improved attribute management processes which include the definition of minimum data quality standards such as to prohibit NULL-values for IAM-relevant attributes.

Data Quality Maintenance. Besides such organizational change, technical measures can support the strategic maintenance of data attribute quality. Structured data reviews by attribute owners, departmental managers, or permission and role owners might be introduced. In the field of IAM, this typically is referred to as data re-certification and covers the human inspection of attribute values and assignments between entities from the conceptual IAM model. However, full data re-certification of attribute values as well as authorization assignments typically results in a significant organizational effort and thus costs. In order to minimize both, we propose a risk-based review of entities. For attributes with a large number of detected errors and a high impact on access policies, on the one hand, a full review might be reasonable. On the other hand, uncritical attributes or assignments might only partially be reviewed. Only suspicious data values might require periodic human evaluation in this case.

After completion of Phase IV the cycle restarts in Phase I. The now cleansed attributes together with previously gathered re-certification decisions can serve as an input for further optimizing other attributes or re-evaluating the quality improvement. Note, that TAQM by design allows for an on-demand parallel execution. Consider an organization that already cleansed a number of personnel master data attributes but is now forced by regulations to connect all locally-managed IT applications to their IAM system. Before this can take place, a comprehensive cleansing should be executed for all attributes that later might be included in attribute-based access policies. At the same time, another TAQM cycle might re-evaluate the previously cleansed personnel master data attributes in parallel. Even more, both execution cycles might involve different experts or be managed by different staff within the organization.

6.5. Fulfillment of Evaluation Criteria by TAQM Approach

After we outlined the main features of TAQM, the following section discusses them with respect to the previously introduced requirements (cf. Section 5 for data quality management approaches for IAM environments).

Focus on Attribute Quality. We purely focus on attribute quality (i.e. data quality of IAM attributes) as this is the foundation for IAM based on ABAC. However, one can argue that the TAQM approach could benefit from an integration of information flow components. We see this as a possible extension whereas we want to answer the initial question of how attributes for an IAM using ABAC can be improved.

Governance. Governance is a key element within IAM but it is not really integrated into data quality management approaches so far. Thus we strongly tie tasks and responsibilities within our approach to the organization and its individuals. The concept of defining attribute owners and to delegate tasks based

on structured re-certification processes (cf. Phase IV of our approach), supports the requirement of governed processes.

Iterative Approach:. TAQM is a cyclic approach, consisting of four different phases which are applied subsequently (including a possible quality assessment loop). By design, it fosters the iterative attribute quality refinement and thereby reflects the fact of IAM being an ongoing process with changing environmental conditions. Carve-ins or carve-outs of companies, movers (i.e. employees changing their position within the organization), or newly introduced IT systems result in a highly dynamic nature of attributes.

Granularity. Most of the existing data quality management approaches fail to deliver fine-grained guidance of how to improve data quality. TAQM does not only consist of a generic high-level phase model but also offers low-level automation metrics and data analysis tools for improving attribute quality within IAM. The previously introduced metrics and analysis techniques present a basic tool set that can be further complemented with individualized implementations.

IAM Content. TAQM is, to the best of our knowledge, the first comprehensive approach to manage attribute quality within an IAM environment and to deal with its specific characteristics. In order to achieve this, we combined existing data quality management approaches with IAM-specific requirements and included all relevant entities based on a structured data model. Additionally, TAQM focuses on the functionalities of IAM that are affected by poor attribute quality such as access regulating attributes.

IAM Completeness. The initially proposed conceptual IAM model serves as the baseline for the TAQM. While TAQM aims at increasing the IAM attribute quality in general, all available IAM entities are fully integrated. Attributes for all relevant and selected entities can be investigated and improved.

7. Applying TAQM

In the following, we evaluate our novel approach according to the design science research evaluation framework of [65]. We build our efforts on a naturalistic ex-post evaluation for rating the effectiveness of our socio-technical artifact using organizational access from our IAM project experience. We describe the application of TAQM throughout three real-life use cases from different companies. The required input datasets have been extracted from the companies' IAM systems and contain employee master data, organizational structure, user accounts, and entitlements from IT applications together with various attributes for all entities (see Table 9). The data was imported into a tool prototype automating the correlation analyses, data review process, as well as data cleansing during TAQM Phase II, III, and IV. Note that the central attribute storage of the prototype as well as most visualizations (e.g. Figure 11) for the experts are based on the proposed conceptual IAM model (cf. Figure 2). As stated in

Dataset	Pseudonym	Employees	Accounts	Entitlements	Applications	Attr. supplied	Date
#1	FactComp	19,829	17,308	7,018	1	8	08/2017
#2	FinComp	5,865	64,429	214,586	387	8	08/2017
#3	AutoCorp	11,386	17,698	75,274	1	15	09/2017

Table 9: Investigated IAM data sets

Section 6 this facilitates the technical implementation and improves the comprehensibility for the human experts. We facilitated the data analysis platform *Nexis Controle* provided by Nexis GmbH[4], a German IAM company and integrated our TAQM prototype functionality. Extending an existing software allowed us to facilitate available data import as well as workflow functionality (e.g. used during the review of identified outliers).

Note that the use cases for company 1 and 2 focus on the application of the first three TAQM phases (Attribute Selection, Quality Assessment, and Quality Inspection) as both companies executed a first TAQM cycle covering those phases in the year 2017. They are planning to execute phase IV subsequently. Use case 3 demonstrates the applicability of TAQM within an industrial company which is currently in the process entering Phase IV of our approach.

7.1. FactComp

The first use case covers a globally-operating manufacturing company with more than 12.000 internal and 4.000 external employees managed using a centralized IAM system which is connected to the main IT applications (Active Directory, SAP ERP, SAP HCM, amongst others). The company is currently improving security and user management efficiency by modeling attribute-based access rules in order to automate joiner, mover, and leaver processes. TAQM was employed to execute an initial attribute quality assessment for employee attributes which later are included in access management policies (e.g. every employee in the IT department is assigned to certain privileges automatically based on the "Department" attribute value).

Phase I. : At first, we imported the available HR master data together with user account and permission data stemming from the company-wide SAP ERP system (see dataset # 1 in Table 9). In total we received eight attributes out of which four were attached to the employee entity (employee group, managerial responsibility, employee type, and IT domain) and four to the permission objects. We were asked to only include permission data stemming from the

[4] `https://www.nexis-secure.com`

company-wide SAP ERP system for permission analyses as the company just recently completed a permission clean-up within this system and thus was able to deliver high-quality permission data. Due to space restrictions in this paper we only focus on the "IT domain" attribute which expresses the company area an employee is assigned to. An employee's IT domain is one of the characteristic attributes deciding about required access privileges. However, it is currently still manually maintained and thus error prone. In summary, 59 distinct attribute values assigned to each of the 19,829 checked identities were provided. Manual data review by IT experts confirmed that all attribute values were syntactically correct (e.g. no typos or spelling mistakes were detected).

Phase II. : During Phase II, we firstly analyzed all other employee attributes regarding their suitability for a correlation analysis regarding the "IT domain" attribute following our proposed approach from Subsection 6.2. The attribute "Employee Group" (revealing the management level of an employee), for instance, is already managed in a semi-automated manner by HR personnel due to its importance for payroll processes and thus suitable. We then automatically created all related classification matrices and attribute value groups for correlating the "IT domain" attribute of employees with their "Employee Group". This assessment led to a set of 184 conspicuous attribute value groups. Besides correlating employees' master data attributes, we also correlated the "IT domain" attribute with SAP ERP permissions assignments of the employees. This way, we were able to highlight employees which, according to their SAP ERP permissions, are likely member of a different attribute value group (e.g. an employee with access rights typical for the "IT domain" *Marketing* who is assigned to the "IT domain" *Sales*). This step resulted in a total of 58 possibly erroneously assigned value groups.

Phase III. : Together with IT experts, we reviewed the identified outliers for six "IT domain" areas. The example matrix for the Sales area is displayed in Figure 9, highlighting outliers using orange and red coloring depending on the level of significance (in this case 5%). The first column ("Count") lists the total count of members of a value group while the second column ("Name") shows the value of the respective attribute for the "Employee Group" attribute. The results show that within the Sales area there is only a small number of attribute values for the IT domain shown in the remaining columns (e.g. *PATZ*, *ExecutiveBoard*, ...). Note that the values stated represent the percentage of all value group members assigned to a certain IT domain. IT experts, for instance, confirmed during data cleansing that the attribute value *Africa* should typically be appearing only within the domain *International Operations*.

Another example presented in Figure 10 shows the distribution of attribute values of a certain employee group having similar entitlements within the domain *International Operations*. While additional analyses show, that the orange value (domain *PATZ*) is occurring within this employee group rather normally (0.03% within this group vs 0.01% overall, the overall values are not displayed within the result matrix), only 1 out of 88 employees are attributed to the value *CFS*.

Count	Name	SA_SalesAndMarketing	PATZ	ExecutiveBoard	Breweries	IOSAfrica
499	B1 (499)	0.998				0.002
330	12 (330)	0.99	0.01			
151	06 (151)	0.96	0.03	0.01	0.01	
54	0E (54)	1				
49	04 (49)	1				

Figure 9: Entry of the result matrix with highlighting for the area Sales (screenshot is distorted)

Count	Name	CFS	IOSCentralLCS	ECM	IOSChina	PATZ
88	30 (88)	0.01(1)	0.88(77)	0.1(9)		0.01(1)

Figure 10: Entry of the result matrix with highlighting for the area International Operations (screenshot is distorted)

After the result interpretation and the exclusion of false positive results by IT experts, we arrived at a final set of 100 suspicious value assignments for the "IT domain" attribute out of the initially identified 184 outliers by the employee attribute correlation and 44 suspicious value assignments out of the initially identified 58 by the permission correlation. In total, 308 employees have been affected by these findings. We argue that manually discovering these errors would not have been possible. Only using automated correlation analysis provided by a tool-based prototype, a focused review by experts and the timely execution of the first three TAQM phases (and hence a structured attribute quality management for the IAM system) for FactComp has been made possible.

7.2. FinComp

In our second use case, we supported a company which operates in a highly-regulated environment during the improvement of their already existing IAM system. The organization manages 5.864 employees and 387 IT applications using a centralized IAM tool. Most of the application permissions are still managed manually, resulting in over-privileged employees. The access to building- and location-specific information managed via Microsoft Active Directory group memberships in specific was identified as error-prone. Reasons are, amongst others, the usage of complex group hierarchies within the Active Directory for handling the distribution of building-related information. As a result, we were asked to analyze the attribute quality of the employee attribute "building" which displays the building an employee is working in. Note, that in the following we only briefly describe each TAQM phase with a focus on highlighting differences and side effects that have not yet been discovered within our previously described use case. Table 10 shows the two important attributes for this use case. Within each phase different techniques to discover wrong attribute values for the "building" attribute are employed.

Phase I. : Personnel and organizational information together with account and permission data from 387 different applications was provided and imported in

Table 10: Relevant Attributes

Attribute	Attribute Description
City	Defines the city the employee is currently working at. Known to be correct.
Building	Defines the building the employee is currently located. As one building is assigned to exactly one city these two attributes should correlate for each account. Otherwise it would indicate a wrong attribute value. Known that attribute errors may occur.

our tool prototype. An initial attribute completeness check for the "building" attribute showed that only 92% of all employees are assigned to an attribute value, leaving 456 employees without a valid attribute value. These data errors (NULL-values) have been directly handed over for data cleansing and are excluded in the remainder.

Phase II. : During Phase II we firstly executed an account correlation with the two attributes above. We aimed at revealing correlations of specific application systems with certain buildings and created groups of employees according to the assignment of a user account within each of the imported IT applications. Note that we only considered applications that comprise between 50 and 750 user accounts in total for two reasons: Firstly, widely-used IT applications (such as the Active Directory) cannot be attributed to be only used within specific buildings. Secondly, applications with too few user accounts lead to a high number of false positive results. After a first execution and a subsequent reconfiguration of the applied thresholds (loop-back cycle of TAQM) we discovered 98 IT applications with conspicuous distributions of user accounts, affecting 587 employees' "building" attribute value.

Besides the account correlation we correlated the employees' "building" attribute with other existing master data attributes (analogous to Table 8 or Section 7.1). We, for instance, correlated the employee attribute "city" (describing the city the respective employee works in) with employees' "building" assignment. This additionally identified three suspicious value groups, affecting the "building" assignment of further three employees out of 5.864 which, looking at the data, work in a building which is not located within the city they work in.

Phase III. : During a detailed visual inspection regarding these results, we were able to discard a total of 409 suspicious employees together with IT experts of the company as false positive results. This resulted in remaining quality issues regarding 123 employees. We, for instance, employed the introduced grid visualization in order to confirm one finding of an employee in Singapore with

[5]Grid visualization is a component from the software *Nexis Controle*

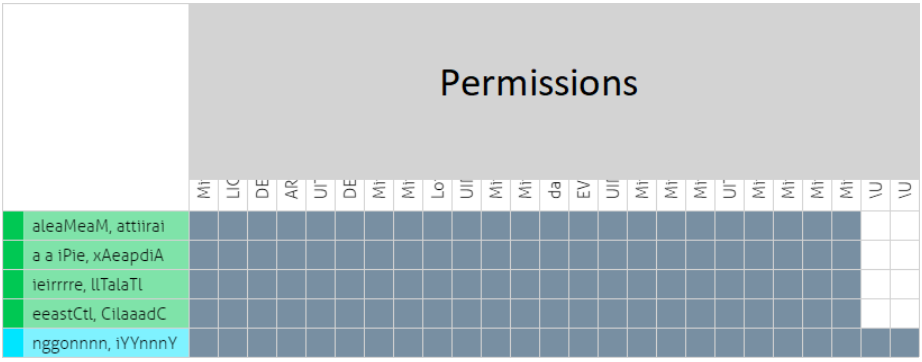


Figure 11: Grid visualization[5] filtered for a small department with colored employees based on their city attribute and having the building attribute value *Building_GER_2* (screenshot is distorted)

a wrong "building" attribute by coloring all employees based on their "city" attribute values.

The prototype displays all five employees having the pseudonymized building attribute value *Building_GER_2*. Note that all other employees not having this specific attribute have been excluded and are not shown. The remaining ones are coloured according to their "city" attribute (see Figure 11): The four green-coloured employees (city attribute value *Hamburg*) positively correlate with the mentioned building value. However, the only blue-colored employee with the city attribute *Singapore* but also with the building value *Building_GER_2* has two additional permissions, no other employee is assigned to (bottom right corner of the figure). The permissions' names are anonymized, however, in the original version they were clearly related to an Asian and Singapore region (e.g. calendar for region Asia). Therefore the "city" attribute is correct as the permission is indicating an Asian location while the value for "building" needs to be adjusted. One reason for such data quality errors might be inadequately enforced mover processes of employees who changed their work location and only received an update on their "city" attribute without adequately changing their "building" attribute.

7.3. AutoComp

The third use case covers a large, world-wide operating company in the automotive sector with more than 11,000 employees and more than 75,000 system entitlements within one application (see dataset #3 in Table 9). Within this use case we want to give an example how to exploit our findings by using the knowledge of experts within Phase IV to cleanse data quality issues. The company recently completed an HR-based project to introduce three new employee attributes ("job", "jobgroup", and "jobbox") for assigning access privileges according to employees' jobs within the organization. The project aimed at defining and assigning valid values for those attributes to all employees in a top-down manner, i.e. manually by experts for each department. During this

process, departmental managers were asked to provide their employees' job-, jobgroup-, and jobbox assignments based on a predefined list of valid values.

Phase I, II, and III. : After importing the provided data into our data storage we analyzed the attribute filling levels (job (96%), jobgroup (96%), and jobbox (71%)). Despite the fact that the attributes have a logical dependency (i.e. every job is assigned to a specific jobgroup and every jobgroup is assigned to a specific jobbox), they have been maintained in an independent manner. Consequently, we were able to find a total of 67 suspicious attribute value groups, covering 304 user accounts (out of 11,500 accounts) during Phase II. During Phase III, IT expert analysis confirmed these results to a large extent, for instance in a case where all except one employee with the same job value were in the same jobgroup.

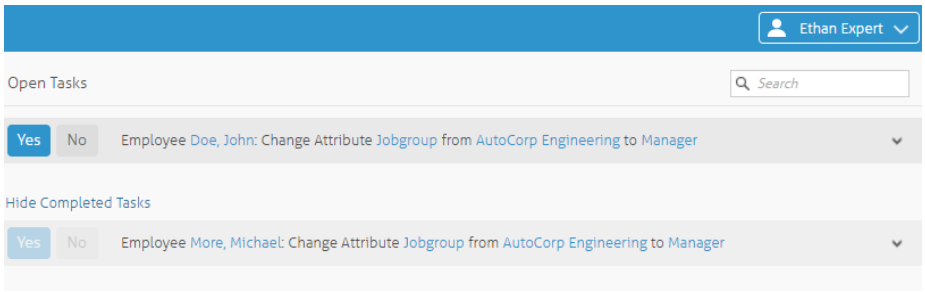


Figure 12: Verification[6]of anomalies via graphical user interface

Phase IV. : In order to cleanse the identified quality issues, our results were provided to responsible staff (e.g. the attribute owner) during Phase IV. We informed responsible IAM managers about the planned data cleansing process together with non-IT experts. In order to maximize user adoption, we used the existing graphical user interface for business experts of *Nexis Controle* (see Figure 12). The software's built-in delegation workflows already offered basic expert review processes which we extended in order to display TAQM results from Phase II and III to non-IT experts.

By using a simplified graphical user interface to review data quality issues, we foster the integration of organizational knowledge from non-IT staff. Potential stakeholders are departmental managers, attribute owners, entitlement owners, or other experienced employees. During TAQM execution we, for instance, automatically delegate findings from Phase III as review tasks to responsible experts. In the following, we present a simplified example of two exemplary tasks which have been delegated to an attribute owner. Via the graphical user interface he is asked to accept (e.g. cleanse) or decline (e.g. ignore) the attribute quality issues via a simple two-option button design. He can also optionally be

[6]Verification component is from the software *Nexis Controle*

allowed to browse the concerned entities for further information or review his previously completed tasks. In this example we asked "Ethan Expert" who is the owner of the attribute "jobgroup" to confirm two identified outliers and change the attribute value to *Manager* (i.e. a pre-calculated potentially correct value delivered by our employee attribute correlation during Phase II). The two employees were erroneously assigned the value *AutoComp Engineering* for their "jobgroup". Ethan Expert can, for instance, verify our findings by accepting both of the tasks to change the attribute values to the proposed new ones.

By using TAQM in combination with an expert-oriented simplified graphical user interface, companies are enabled to delegate different attribute quality tasks to non-IT experts. Additionally, expert decisions can be stored centrally for compliance or regulatory requirements. Note that, additionally to the already received positive feedback from the IAM managers, AutoComp currently is in the process of deciding about applying our approach during their next project phase.

8. Conclusion

The complexity and the number of challenges IAM has to tackle in modern companies is constantly rising. ABAC is one of the successors of RBAC and offers enough flexibility to overcome several access management challenges. However, the deployment of ABAC presents challenges itself that have not been addressed sufficiently yet by research. Essentially, a lack of attribute quality can lead to dysfunctional access control decisions and hence the existence of security vulnerabilities. Up to now there is no comprehensive attribute quality model ensuring a continuously high quality of attributes used within ABAC policies. To close this gap, we proposed TAQM, a structured approach for data quality management in IAM environments.

We initially derived a conceptual model for IAM, compared existing data quality approaches and analyzed their applicability for IAM. Subsequently, we defined TAQM as a process model and developed different tools to support the execution of each phase. In a last step we applied TAQM within three different IAM use cases. We were able to identify attribute value errors prior unnoticed and verify the suspicious quality issues in cooperation with company experts in real life projects.

After successful evaluation in real-world projects, we now plan to extend our research and monitor the long-term performance of TAQM. We want to analyze to which level the overall attribute quality increases over a longer period. Additionally, we aim at fine-tuning the various automation tools by integrating automated parameter configuration functionality for an easier application of TAQM. Lastly, we also want to investigate TAQM's extensibility towards other fields. One possibility would be the integration with identity behaviour analysis (e.g. answering questions like "are employees which are behaving similarly assigned to similar attribute values?"). Another option is applying previous expert decisions for an improved verification of anomalies. Past decisions could,

for instance, be used to verify identified outliers at run-time in order to increase the detection rate.

References

[1] SOX, Sarbanes-oxley act of 2002, pl 107-204, 116 stat 745 (2002).

[2] Basel Comittee on Banking Supervisions, Basel III: Int. framework for liquidity risk measurement, standards and monitoring.

[3] M. Hummer, M. Kunz, M. Netter, L. Fuchs, G. Pernul, Adaptive identity and access management—contextual data based policies, Journal on Information Security 2016 (1) (2016) 19.

[4] R. S. Sandhu, E. J. Coyne, H. L. Feinstein, C. E. Youman, Role-based access control models, Computer 29 (2) (1996) 38–47.

[5] L. Fuchs, G. Pernul, Hydro–hybrid development of roles, in: International Conference on Information Systems Security, Springer, 2008, pp. 287–302.

[6] V. C. Hu, D. F. Ferraiolo, D. R. Kuhn, A. Schnitzer, K. Sandlin, R. Miller, K. Scarfone, Guide to attribute based access control (abac) definition and considerations, NIST Special Publication (2014) .

[7] N. K. Sharma, A. Joshi, Representing attribute based access control policies in owl, in: 10th International Conference on Semantic Computing, IEEE, 2016, pp. 333–336.

[8] V. Hu, D. F. Ferraiolo, D. R. Kuhn, R. N. Kacker, Y. Lei, Implementing and managing policy rules in attribute based access control, in: 16th International Conference on Information Reuse and Integration, IEEE, 2015, pp. 518–525.

[9] L. Fuchs, G. Pernul, R. S. Sandhu, Roles in information security–a survey and classification of the research area, Computers & Security 30 (8) (2011) 748–769.

[10] L. Fuchs, M. Kunz, G. Pernul, Role model optimization for secure role-based identity management, in: 22nd European Conference on Information Systems, AISeL, 2014.

[11] A. Elliott, S. Knight, Role explosion: Acknowledging the problem, in: 8th International Conference on Software Engineering Research and Practice, 2010, pp. 349–355.

[12] R. S. Sandhu, The authorization leap from rights to attributes: Maturation or chaos?, in: 17th ACM symposium on Access Control Models and Technologies, ACM, 2012, pp. 69–70.

[13] A. C. O'Connor, R. J. Loomis, Economic analysis of role-based access control, NIST, Gaithersburg, MD 2010.

[14] T. Priebe, W. Dobmeier, B. Muschall, G. Pernul, Abac–ein referenzmodell für attributbasierte zugriffskontrolle, in: Sicherheit, 2005, pp. 285–296.

[15] E. Yuan, J. Tong, Attributed based access control (abac) for web services, in: 9th International Conference on Web Services, IEEE, 2005.

[16] Z. Xu, S. D. Stoller, Mining attribute-based access control policies from rbac policies, in: 10th International Conference and Expo on Emerging Technologies for a Smarter World, IEEE, 2013.

[17] W. E. Deming, Out of the crisis, Cambridge University Press, 1986.

[18] R. Y. Wang, A product perspective on total data quality management, Communications of the ACM 41 (2) (1998) 58–65.

[19] C. Batini, M. Scannapieco, Data and information quality: Dimensions, principles and techniques, Springer, 2016.

[20] P. B. Crosby, Quality is free: The art of making quality certain, Signet, 1980.

[21] B. Aquilani, C. Silvestri, A. Ruggieri, C. Gatti, A systematic literature review on total quality management critical success factors and the identification of new avenues of research, The TQM Journal 29 (1) (2017) 184–213.

[22] J. P. Womack, D. T. Jones, D. Roos, Machine that changed the world, Simon and Schuster, 1990.

[23] P. S. Pande, R. P. Neuman, R. R. Cavanagh, The six sigma way, McGraw-Hill, 2000.

[24] M. Kaiser, M. Klier, B. Heinrich, How to measure data quality?–a metric-based approach, in: 28th International Conference on Information Systems, AISeL, 2007.

[25] H. Hinrichs, Datenqualitätsmanagement in Data Warehouse-Systemen, Springer, 2002.

[26] D. Ballou, R. Y. Wang, H. Pazer, G. K. Tayi, Modeling information manufacturing systems to determine information product quality, Management Science 44 (4) (1998) 462–484.

[27] A. K. Elmagarmid, P. G. Ipeirotis, V. S. Verykios, Duplicate record detection: A survey, IEEE Transactions on Knowledge and Data Engineering 19 (1) (2007) 1–16.

[28] D. Servos, S. L. Osborn, Current research and open problems in attribute-based access control, ACM Computing Surveys 49 (4) (2017) 65.

[29] J. Werner, C. M. Westphall, C. B. Westphall, Cloud identity management: A survey on privacy strategies, Computer Networks 122 (2017) 29–42.

[30] V. Hu, D. F. Ferraiolo, D. R. Kuhn, R. N. Kacker, Y. Lei, Implementing and managing policy rules in attribute based access control, in: Information Reuse and Integration (IRI), 2015 IEEE International Conference on, IEEE, 2015, pp. 518–525.

[31] A. R. Hevner, S. T. March, J. Park, S. Ram, Design science in information systems research, MIS quarterly 28 (1) (2004) 75–105.

[32] J. Sermersheim, Lightweight directory access protocol (ldap): The protocol (2006).
URL https://tools.ietf.org/html/rfc4511

[33] R. Philpott, N. Ragouzis, T. Wisniewski, E. G. Whitehead, H. Hinton, C. P. Cahill, J. Bradley, J. Hodges, J. Brennan, et al., Assertions and protocols for the oasis security assertion markup language (saml) v2. 0 errata 05 (2012).
URL https://www.oasis-open.org/committees/download.php/56776/sstc-saml-core-errata-2.0-wd-07.pdf

[34] D. Hardt, Rfc 6749 - the oauth 2.0 authorization framework (2012).
URL https://tools.ietf.org/html/rfc6749

[35] D. Rolls, Service provisioning markup language (spml) version 1.0 (2003).
URL https://www.oasis-open.org/committees/download.php/4137/os-pstc-spml-core-1.0.pdf

[36] T. Moses, extensible access control markup language (xacml) version 2.0 (2005).
URL http://docs.oasis-open.org/xacml/2.0/access_control-xacml-2.0-core-spec-os.pdf

[37] N. Naik, P. Jenkins, A secure mobile cloud identity: Criteria for effective identity and access management standards, in: 4th International Conference on Mobile Cloud Computing, Services, and Engineering, IEEE, 2016, pp. 89–90.

[38] P. E. Hunt, K. Grizzle, E. Wahlstroem, M. C., Rfc 7643 - system for cross-domain identity management: Core schema (2015).
URL https://tools.ietf.org/html/rfc7643

[39] S. Oh, S. Park, Task–role-based access control model, Information systems 28 (6) (2003) 533–562.

[40] A. A. E. Kalam, R. E. Baida, P. Balbiani, S. Benferhat, F. Cuppens, Y. Deswarte, A. Miege, C. Saurel, G. Trouessin, Organization based access control, in: 4th International Workshop on Policies for Distributed Systems and Networks, IEEE, 2003, pp. 120–131.

[41] P. J. Windley, Digital identity: Unmasking identity management architecture (IMA), O'Reilly Media, Inc., 2005.

[42] A. Pfitzmann, M. Hansen, A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management, technical report (2010).

[43] J. A. Long, C. E. Seko, A cyclic-hierarchical method for database data-quality evaluation and improvement, Information quality 1 (2005) 52–66.

[44] P. Falorsi, S. Pallara, A. Pavone, A. Alessandroni, E. Massella, M. Scannapieco, Improving the quality of toponymic data in the italian public administration, in: International Conference on Database Theory, Vol. 3, 2003.

[45] L. Fuchs, G. Pernul, Qualitätssicherung im identity- und access management, HMD Praxis der Wirtschaftsinformatik 50 (1) (2013) 88–97.

[46] Y. W. Lee, D. M. Strong, B. K. Kahn, R. Y. Wang, Aimq: A methodology for information quality assessment, Information & management 40 (2) (2002) 133–146.

[47] Y. Su, Z. Jin, A methodology for information quality assessment in the designing and manufacturing process of mechanical products, in: Information Quality Management: Theory and Applications, Idea Group Publishing Hershey, 2006, pp. 190–220.

[48] C. Batini, M. Scannapieco, Data quality: Concepts, methodologies and techniques, Springer, 2006.

[49] D. Loshin, Enterprise knowledge management: The data quality approach, Morgan Kaufmann, 2001.

[50] M. Scannapieco, A. Virgillito, C. Marchetti, M. Mecella, R. Baldoni, The daquincis architecture: A platform for exchanging and improving data quality in cooperative information systems, Information systems 29 (7) (2004) 551–582.

[51] L. L. Pipino, Y. W. Lee, R. Y. Wang, Data quality assessment, Communications of the ACM 45 (4) (2002) 211–218.

[52] M. A. Jeusfeld, C. Quix, M. Jarke, Design and analysis of quality information for data warehouses, in: 17th International Conference on Conceptual Modeling, Springer, 1998, pp. 349–362.

[53] M. J. Eppler, P. Muenzenmayer, Measuring information quality in the web context: A survey of state-of-the-art instruments and an application methodology, in: 7th International Conference on Information Quality, 2002, pp. 187–196.

[54] F. De Amicis, C. Batini, A methodology for data quality assessment on financial data, Studies in Communication Sciences 4 (2) (2004) 115–137.

[55] L. P. English, Improving data warehouse and business information quality, J. Wiley & Sons, 1999.

[56] D. Brossard, G. Gebel, M. Berg, A systematic approach to implementing abac, in: Proceedings of the 2nd ACM Workshop on Attribute-Based Access Control, ACM, 2017, pp. 53–59.

[57] S. Bhatt, F. Patwa, R. Sandhu, Abac with group attributes and attribute hierarchies utilizing the policy machine, in: Proceedings of the 2nd ACM Workshop on Attribute-Based Access Control, ACM, 2017, pp. 17–28.

[58] I. Indu, P. R. Anand, V. Bhaskar, Identity and access management in cloud environment: Mechanisms and challenges, Engineering Science and Technology, an International Journal.

[59] U. Habiba, R. Masood, M. A. Shibli, M. A. Niazi, Cloud identity management security issues & solutions: a taxonomy, Complex Adaptive Systems Modeling 2 (1) (2014) 5.

[60] R. Likert, A technique for the measurement of attitudes, Archives of psychology 22 (140) (1932) .

[61] B. Heinrich, M. Kaiser, M. Klier, Does the eu insurance mediation directive help to improve data quality? - a metric-based analysis, in: 16th European Conference on Information Systems, AISeL, 2008.

[62] M. Kunz, L. Fuchs, M. Hummer, G. Pernul, Introducing dynamic identity and access management in organizations, in: 11th International Conference on Information Systems Security, 2015, pp. 139–158.

[63] L. Fuchs, Methodology for Hybrid Role Development, Vol. 69, Eul Verlag, 2010.

[64] S. Meier, L. Fuchs, G. Pernul, Managing the access grid - a process view to minimize insider misuse risks, in: 11th International Conference on Wirtschaftsinformatik, 2013, pp. 1051–1065.

[65] J. Venable, J. Pries-Heje, R. Baskerville, A comprehensive framework for evaluation in design science research, in: 7th International Conference on Design Science Research in Information Systems, Springer, 2012, pp. 423–438.