# A Reference Model of Integrative Identity Management for IUBH

# – Proposal –

Version: 0.02

# History of Modifications:

| Date | Editor | Version | Modifications | Note |
|---|---|---|---|---|
| 03.01.2017 | H. Kampf | 0.01 | Beginning of creation | |
| 24.01.2017 | H. Kampf | 0.02 | Release for review | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |

# 1 Introduction

An integrative identity management (IdM) means here the administration of entities (persons and objects) with their rights determined in roles for specific contexts. For this purpose, each entity is described by sets of attributes (attribute frames) assigned to a unique identifier. The IdM considered here mainly represents the technical design of processes under the aspects of efficiency, flexibility as well as security and less a sociological aim e. g. the right of informational self-determination of persons (acting under pseudonyms), without keeping out of mind the data protection regulations.

The basic purpose of managing personal identities is the creation and deletion one and only on (unique) digital identities as well as their distribution (provisioning), thus securing the access to IT resources (e-learning applications and their modules, IT systems and network infrastructure etc.); e. g. erasing specific and associated rights to a person for accessing to IT resources caused by leaving the company as an employee or by exmatriculation as a student the corresponding rights immediately will be revoked and the access to the specific IT resources is blocked.

The IdM should be designed also to provide mechanisms such as single-sign-on (SSO) and / or password synchronization so that after a single login or after a password change the identity of a user is available immediately for a particular period of time within an application process running on possibly different dedicated systems to enable the user to access of all resources driven by its role and authorization. The entry into the systems may be qualified by means of a corresponding authentication procedure by means of verification methods which check the authorship of the login to identify the real person.

In this overall scenario, the term "provisioning" is applied also which describe the comprehensive process to automate all steps required to synchronize, deploy, and manage user or system access rights based on roll-based access control concepts for heterogeneous target systems.

The concerted action of several organizations in the dissent on their autonomies suggests that the underlying system will have a federated structure, but with the help of replication, referral and forwarding mechanisms (caching, shadowing, referrals, chaining, multicast, provisioning, etc.) it can be consider in a macroscopic manner as a central system with its corresponding advantages. As infrastructure basic technology, systems like (meta) directories and their associated services (e. g. connectors) seem to be accepted.

# 2 Identity Management

In the IT environment, processes are continuously taking place with their focus at

- persons and
- organizations as well as
- roles and
- authorizations

Persons get profiles which can change during time because of given different status and membership to one or more organizations resp. Organizational units (university, faculty, department, institute, lecture, project, etc.). With help of these profiles, roles can be defined so that dedicated accesses to (IT) resources are granted to persons via these roles within their scope of work processes. At the same time, the structures of organizations can also change by establishing new one and dissolving other, as well as merging or dividing several organizational units, whereby the profiles as well as the assignment to the corresponding roles change, but the role-rights relation may remain unchanged.

Be able to efficiently deal with these organizational processes of changes at IT level, an appropriate administration, identity management, is required. The basic goal of identity management is to create and distribute digital identities as well as to secure IT resources.

## 2.1 The Term “Identity Management”

At present time, a consolidated holistic definition of identity management[1] (IdM) and what it contains, as well as a firm nomenclature, does not seem to be in sight in view of the author, so that this concept and its deducted associations are tried to outline comprehensively by their phenomens from a point of view driven by this specific task here.

At present state of the art, IdM can be understand as an ecology consists of processes, services, applications and IT infrastructure as well as compliances, agreements and policies in order to define primarily digital identities of persons, but also more extensively where required those of organizational units and (IT) resources. Using these digital identities in well-defined contexts of interaction in a dispositive and operative manner the target is to be able to access to dedicated IT resources in a secure way in order to fulfill specific tasks efficiently and in compliance to norms.

For the sake of clarity and in stringency to the assigned task, the focus for the model of an interdisciplinary integrative identity management here is on identities which reflect natural persons with corresponding to management processes with regard to

- existence (generation, maintenance, deletion and distribution),
- certificates (entitlement certificates),
- profiles, roles and authorization mechanisms as well as
- access processes, including the involved
- authentication and authorization problems.

In analogy, the following can be applied relatively simply to other object classes, such as rooms, events, etc., and thus expand the IdM.

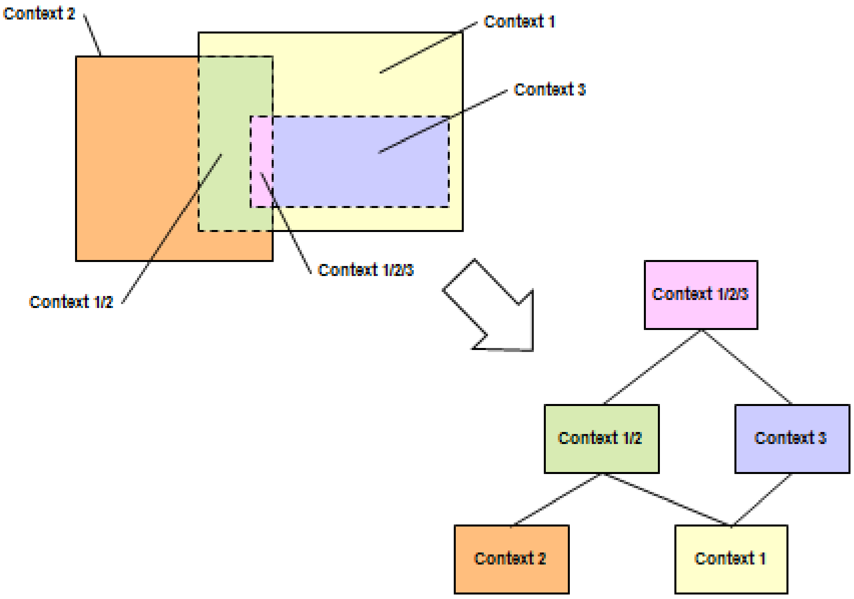
## 2.2 Contexts

If intersections and / or sub-sets are extracted and defined as logical higher contexts, the entire contexts can be arranged like a tree, and so it is formed a hierarchy of inheritance with regard to the attributes. Also, some inherited attributes can be supplemented and / or overridden as a result of context-sensitive variants.

The digital identity provides an accessible representation of a natural person in totality of its attributes as well as in well-defined parts of the totality. A unique identifier in all involved contexts serves hereby for an unambiguous recognition and enables the correct identification in the relationship of communication between single contexts.

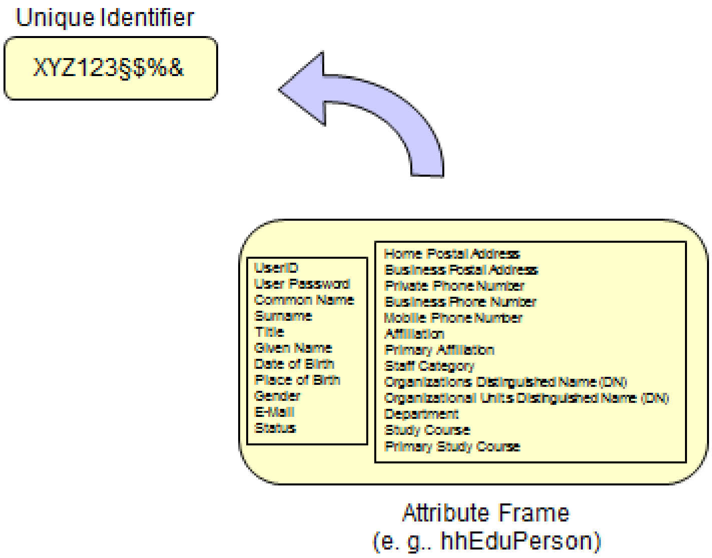
Thus, for example, the attribute set of a situational context after instantiation reflects a digital image of a natural person with its action-specific role in a particular organizational unit without having to handle all attributes of the whole contexts.

[1] The term "identity and access management" (IAM) is often used as a synonym in this context




## 2.3 Unique Identifier

Attributes which describe a person who is natural in existence are not always sufficient in availability of quantity and characteristic of values as a criterion for the distinction. Also a part of this quantity of attributes can be changed during time, which could result in collisions, which would have to be solved with considerable effort to recuperate an unambiguous identification. In addition the effort to maintain consistency overall contexts considerably increases. Caused by these reasons a surrogate, i.e. an artificially generated, system-immanent attribute is used as a unique identifier, which separates persons clearly from each other. Depending on the context, different attribute frames may be assigned to this identifier.




These set of well-formed unique identifiers in their totality represents a so-called controlled vocabulary which, due to the "1:1" relationship to natural persons, has neither homonyms nor synonyms. In its semantics, this vocabulary does not allow any conclusions to be drawn from the represented natural persons, which can contribute to anonymization within certain contexts according to privacy policy.

For example, generating world-wide unique identifiers can be supported by means of a Private Enterprise Number (PEN) created and maintained by the Internet Assigned Number Authority (IANA) in a public registry.

## 2.4 Fundamentals of Identity Management Systems

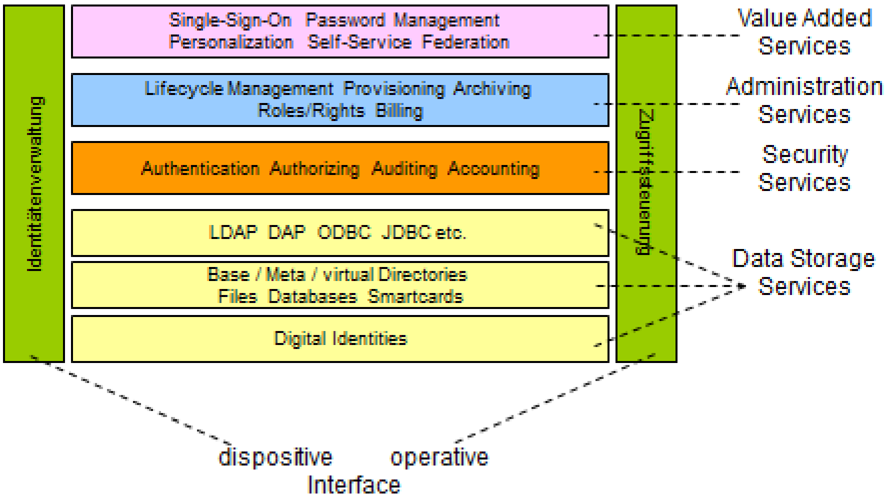
A person who has access to an IT platform with differently provided services respectivly applications, or general resources, must be clearly identifiable within a particular context by means of certain characteristics via a possible identity-detection method. In other words, the personal data stored in a technical system represent, in correspondence with a natural person, the digital identity by means of which the access and processing authorizations are determined and the access is released correspondingly.

For managing stored digital identities as well as access rights to (IT) resources identity management systems (IdMS) come into sight, which act in a singular or in a collaborative manner. Through their interfaces to perform

- identity management and
- access control

certain services at different levels for the dispositive resp. operational access be available.

The Access to each IdMS is subject to the same requirements as for any (IT) resource; that means that an IdMS itself needs an internal identity management system with the basic functionalities not only for the internal administration of identities but also with the inclusion of services with self-service and self-care to regulate the control of accesses, only with the difference that their identities are not provided with regard to other systems.




### 2.4.1 Data Storage Services

Many applications have internal data maintenance services with their own data storage in the form of common database systems with a corresponding access protocol, possibly also with a (L) DAP interface, in which identities (user data) and access rights application oriented be stored and maintained or there exists an associated directory service on separated base directory. Such repositories often form source systems from which, for example, meta directories are filled, as well as target systems that are filled with corresponding data.

In order to keep a specific portion of the identity data consistent across contexts, meta directories are used in conjunction with the LDAP protocol, which serves the resource-oriented IdMS as both a target and a source system, that means that their task is to merge data from different sources, to synchronize them, and to dissolve contradictions in information, for example in case of same user names. For these parts of data, the decentralized source systems are defined as the "leading" system, so that these systems always contain the data that is generally valid. The meta directory service itself is an automated copy process, which takes the individual data from certain leading systems (source systems) according to certain rules, stores them redundantly in the internal directory and distributes them to all other systems (target systems) as required.

Virtual directories can be used for cooperative purposes. They integrate information from different directories dynamically at runtime and represent it as a directory, i.e. the information is collected from several directories and presented in the desired structure. The virtual directory itself includes only mapping information of connections, so, in case of a request, the directory service provides only a reference where it can be picked up the corresponding information, or the corresponding information is delivered from the other directory using the stored reference.

### 2.4.2 Security Services

The process of authentication here serves to verify a registered identity, stored in a data base (database, directory, smartcard), on base of certain features, in order to determine that a person associated with this identity is that which it pretends to be. Verifying this assertion the person submits certain characteristics such as

- a password issued by an authorized authority,
- a signature or
- biometric features, etc.

trying to authenticate itself. With help of these features, an attempt is made to find and identify this person via matching methods with the stored digital identities – authentication is here equivalent to the process of identification because of the unambiguous proof of identity. After verification, a token, the so-called authenticator, can be generated, allowing others to notice that a successful check has taken place.

After successful authentication, the authorization, which controls the access in a rule-based and / or roll-based manner, may be carried out; that is, for example, that the access to relevant resources by the appropriate access rights based to the assigned role to an identity is granted.

While accounting records logs the use of services for statistics, billing, revision, detection of security breaches and offering evidence, auditing investigates and evaluates these records to determine whether security violations have actually occurred, who has triggered them, and what resources were affected.

### 2.4.3 Administration Services

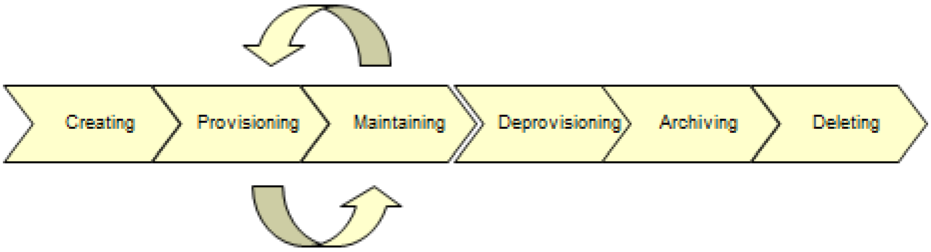
In addition to the pure lifecycle management (creating, maintaining deleting), the administrative process for digital identities supports three other services:

- provisioning
- archiving
- deprovisioning

Provisioning[2] and deprovisioning includes the transfer process, which is designed by means of push or pull mechanism, which automates, after creating, maintaining or deleting an identity, all steps which are needed for synchronization and availability of digital identities as well as revocation of rights in the direction of more or less heterogeneous target systems (application systems, but also other IdM systems) and if necessary rule-based, transforming and / or collision-resolving (treatment of duplicates).

Digital identities may be modified during their lifecycle, which must be recorded to be able to certify what access rights in which certain time frames someone have had, regardless of whether resources were accessed or not. Archiving of this identity history also includes the process of the so-called logical deletion. After a certain time, the archived identity-related data can be permanently deleted.

[2] Provisioning here is not equivalent to replication, which is used in the case of data volume and / or load distribution performed in order to be able to scale the availability accordingly.




Along with identity-related administration services, there is another service that enables the creation and administration of roles and their associated usage and access rights to specific resources, as well as the classification of roles in organizational structure, this in order to be able to perform an authorization check (authorization) within an application process. Rights are thus indirectly assigned to the identities via roles, whereby it is fundamental that an identity or the person respectively who embodies them may not have more rights than he needs to perform a specific task.

The billing service, among other things, designs the entire process of invoicing and includes the sub-processes:

- tariffing
- gathering of service data per user
- assignment of the usage data to the appropriate tariffs
- accounting and invoice delivery
- controlling and collection

### 2.4.4 Value Added Services

With self-service services, the owner of the identity that means the natural person represented by the digital identity is to be enabled to see specific attributes of his digital identity or to change and delete them as well.

The password management allows system users not only to synchronize usernames and passwords of the various application systems or contexts but also to change and reset them all more or less at the same time. In addition, the local identification and password policies of the various application systems and contexts are taken into account. At the same time it have to be ensured that for each creation of the combination of username and password as an equivalent to the unique identifier the condition of a "1:1"-relationship for the respective context or application system is fulfilled.

Single-sign-on services allow that after a single login the identity of a user is available in a secure way for a certain period of time, so, that the user can access to all resources in a accordance to his authorization without logging in again (e. g. Central Authentication Service – CAS). This entry into the system is qualified by means of a corresponding authentication method, which verifies the authenticity of the login by means of verification methods, thus identifying the real identity of the user and thereby identifying it.

In addition to single-sign-on services, the federation also allows the one-time log-in, but the difference here is that by means of an infrastructure of trust including associated protocols and procedures (such as Shibboleth, Liberty Alliance, PAPI, etc.) the responsibility of authentication and authorization is delegated to other parties participating in this federation on trust arrangements. The purpose of this federation is to bring together decentralized web services in order to provide them with a uniform and reliable authorization process for other organizational units.

## 2.5 Organization-wide Identity Management

The operational structure of each university consists analog to companies of business processes that connect persons and organizational units with services and applications, IT infrastructure components and other resources (rooms, etc.) in order to provide services and / or products for internal and external customers rather users for their core business.

Formally speaking, a business process consists of a chain of business value-creating activities with a specific start and end point. The flow of activities is managed at the top level by persons, the actors, who play a role according to their organizational integration and / or competences and provide a specific service to the customer using IT resources.

From the processes, IT services are performed to fulfill specific tasks, which provide individually and / or possibly via workflows relevant results for these processes. For this purpose, actors need access to relevant IT systems and also to its components (services, applications, networks, etc.). In order to make this possible, user accounts are usually set up and corresponding usage rights are defined for accessing, for example, to an e-learning-platform with its applications and the underlying middleware as well as operating software.

This view of the processes at least gives the first hint to the question of which identities from the circle of users and actors have to be managed at all. And also, this context makes clear that the question cannot be here of which is the leading technical system for the IdM, from which further data may be provided, but it must be based on the processes in which digital person identities are generated and used and may be deleted. If these processes are determined, then the source and target systems and the paths between the individual technical systems will be recognized.

Consolidating many data sources and sinks for digital identity information or keeping it at least synchronously is one of the cardinal tasks of identity management. Taking the degree of integration as the determining factor for the applied repositories[3] so, the approaches range from common directory for several applications or IT resources to federal solutions.

## 2.6 Centralized vs. distributed Directories

A common central base directory for all applications existing in the involved organizations and universities has the advantage that the identity information is only stored in one place and can be centrally administrated and maintained. This reduces the maintenance effort and allows the applications to access to consistent data at any time.

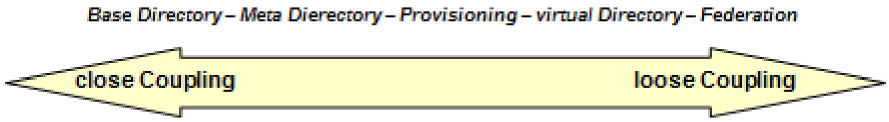
However, from the technical point of view, this approach includes the serious disadvantage that the volume and the complexity of the attribute frame increase with each integrated resource that stores its own application-relevant information (e. g. roles and rights) into this frame and after a certain number of application the reduction of maintenance efforts is counteracted. Furthermore, it is to be noted that each application does not support all types of directories and / or has proprietary data repository for identities, but possibly provides a standardized LDAP- interface to access identities.

[3] Since a directory is usually used as a repository, directories will be utilized further, even if a dedicated database and other could still provide identical functionality.

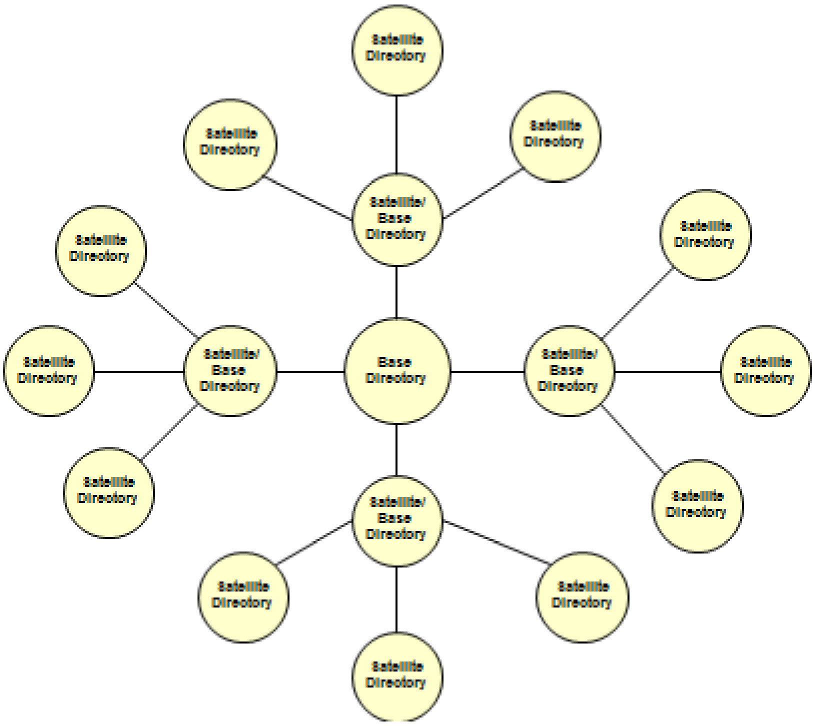
In terms of data protection aspects, there are also concerns regarding the necessity, data avoidance and minimization, as well as the principles of purpose (personal data should be collected and processed for specified and explicit purposes), since there is a close coupling between personal information.

This close coupling is also not advisable for systemic reasons, especially with regard to IT security, because it is necessary to protect sensitive data and business processes, as well as to exclude any reduction of business activities due to IT breakdowns.

This all leads to the strategy, which moves away from a single central directory as a connecting element of the universities and other organizations and tends to a forest of several trees of directories with varying characteristics.




By this holistic view the architecture would then consist of centrally oriented core directories around which decentralized distributed so-called satellite directories. The communicative connection between so complementary types of directories would be established by means of provision mechanisms. At the same time, this type of arrangement provided the advantage of system-immanent latent availability of load distribution, fallback and backup resp. replication functionality in most general sense.



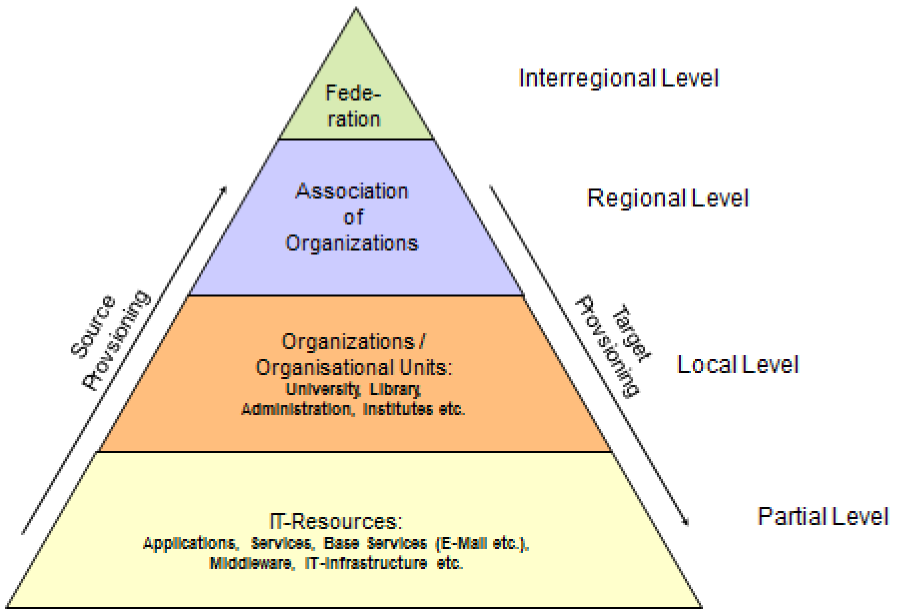

### 2.6.1 Layered Architecture

The concept of the IdM pyramid architecture of the universities and its associated institutions consists of 4 levels, in which IdM is motivated with different characteristics and goals. The levels are described as a model by means of their imaginary areas they will cover, whereby then a

- partial,
- local,
- regional and
- interregional

level results, which hierarchically arrange a logical layered architecture.

The individual IdM systems in the layers are communicatively linked to another in neighboring higher and / or lower level as well as possibly also within their own layer by means of provisioning mechanisms. This kind of architecture building up a coupled cascade of IdM systems and represents the so-called integrative identity management of the involved organizations. Here, one or more logically or physically separated source and target systems of a lower level, in which are produced and / or consumed identities, face to an IdM system that deals with this identities – this also applies within a level where similar structures exist.

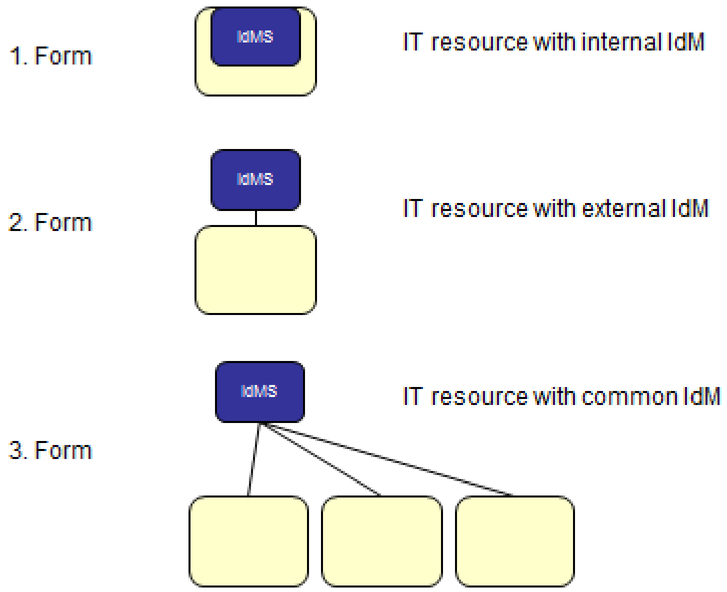



In principle, with transition to higher layers, the information density of personal information in the digital identities decreases, so that at the regional level, the attribute frame of a digital identity has only the (intersecting) set of attributes necessary to describe the person to be represented in the higher context of the association of universities and organizations. At the same time, the number of these digital identities is increasing, since at this level the digital identities of all universities and associated institutions involved are to be stored.

#### *2.6.1.1 Partial Level*

IdM (end) source systems, in which the raw data of possible identities are generated for the IdM, and IdM (end) target systems that are related to the main task of IT resources user and / or account management, authentication, authorization and, if necessary, billing (accounting) are located on a partial level. Such systems are often subsumed under the term triple-A systems[4].

IdM here have primarily the focus on usage with more or less finely granular roles and rights to ensure a secure and authorized access to IT resources using authentication. In this case, an IdM system for user management is assigned to the relevant IT resources representing the so-called (end) target systems, whereby the degree of assignment can be different.




The first form stands for IT resources that have an internal proprietary user management that is not readily visible from the outside, that’s mean that it is either "implicitly" wired in the application software or its data is stored in an integrated data base. If there is no LDAP interface, adequate adapters have to develop to integrate these IdM systems.

The second form presents IT resources, which have explicitly an LDAP-enabled respository or directory.

In the third form, a central account or user directory or repository represents the IdM system[5] whose authentication service is used by multiple IT resources, such as hyper-applications, where multiple applications deliver their part of a given service.

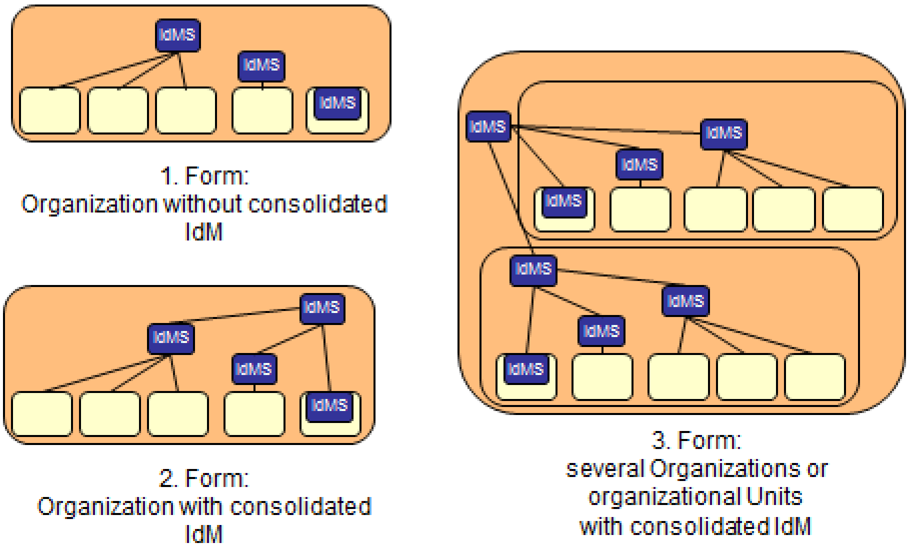
---

[4] Triple-A systems (or AAA systems, AAA) are widely used by wired and mobile network carriers as well as Internet service providers. The three “A” stands for authentication, authorization, and accounting for network access of customers (end users). Triple-A systems are often coupled with customer management systems (IdM systems may deliver parts of these customer data or manage the commercial aspects and data of end users) and supply other systems (e.g., billing, accounting). (See http://de.wikipedia.org/wiki/Triple-ASystem).

[5] In this context, these systems are also referred to as Corporate or Enterprise Directory but they would have to be located on the next level. Also, their task seems ambivalent because, on the one hand, they serve the distribution of identity information via provisioning and / or simple distribution and replication mechanisms (Referrals, chaining, shadowing) and on the other hand, their use is reduced only to the authentication process. However, these directories have their origin in the fulfillments of collaborative tasks caused by the introduction of company or intranet portals and are primarily directories in the form of address books, such as white or yellow pages, which mainly serve the

### 2.6.1.2 Local Level

At the local level, the universities themselves are located with their faculties, departments, affiliated institutes (e. g. Data Center) and administrative institutions as well as facilities associated with higher education institutions (e. g. university library), but also virtual organization units in the form of, for example, common services offered by several universities. The main task of this level with respect to IdM is to consolidate the distributed IdM information on an attribute frame that adequately describes the person to be represented, as well as the creation of well-defined transfer points.




Here, mainly systems (2nd and 3rd form) are used for the IdM task in which (possibly preliminary) digital identities are stored in order to obtain organization-related information about the person to be represented. This is a kind of meta directory which consolidates the data for IdM aspirants from different external data sources and (if necessary) distributes them back and / or to other data repositories as targets (provisioning).

If there are some organizational, procedural and data-related conditions fulfilled with regard to the IdM systems, such as a relatively small organizational unit with low data volume as well as direct provisioning from a consolidated IdM system of higher or the same, that means the local level, the task of the consolidation by means of an explicit IdM system is not necessary (1st form). This task moves then towards authentication, in order to grant persons with their access rights according to their organizational affiliation. Thereby, a strict separation between the partial and local level is no longer given, so that these level can merge here.

---

consolidation and central administration of personal data of larger organizational units. However, other objects, such as organizational units and / or IT components, which must be kept consolidated as part of system management and / or distributed to other directories, can also be inserted into such directories.

If the IdM systems of this level are directly related to the regional level, the consolidation and provisioning takes place on the basis of a well-defined, as far as possible, identical attribute frame for all IdM systems. There are two variants for data consolidation:

- All identities are built entirely from the different external data repositories and form a full set of all identities of an organization o organizational unit respectively
- A subset of the identities is extracted from the various external data repositories

For the attribute frame, mentioned above, the attributes are selected from the external data sets

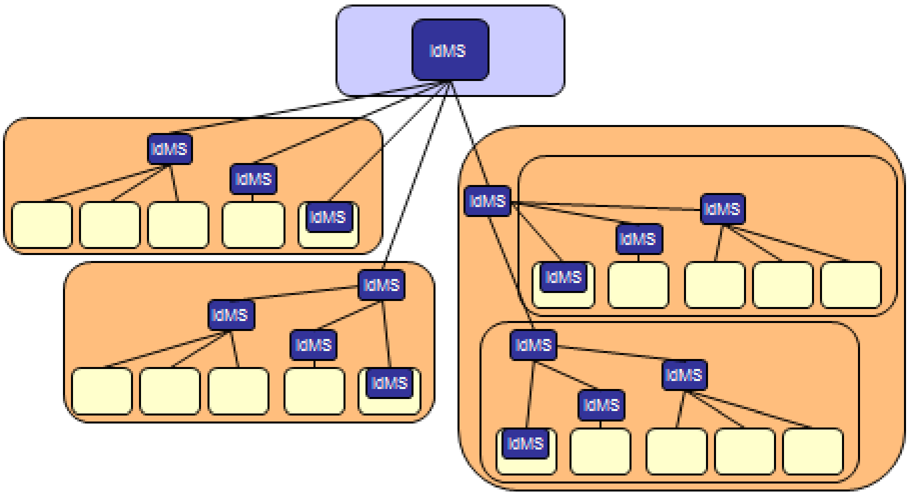
- completely or
- accepted in partial quantities and / or
- will be added by further attributes.

With the help of this structure, well-defined transfer points between the organizations or organizational units are created, and particular responsibility and data integrity, for example, remain in university institutions providing identity data, whereby a certain degree of autonomy is ensured.

#### *2.6.1.3 Regional Level*

At regional level, IdM is provided overall universities or organizations. For this purpose, all relevant identities of the involved universities and organizations are stored in the IdM system. The main task of this level is

- identification and
- registration.



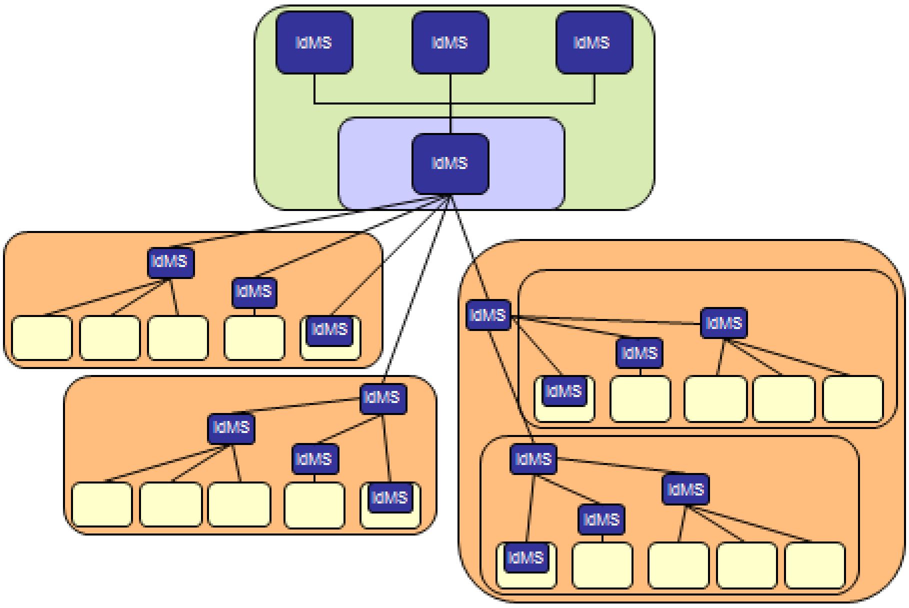
The process of identification comprehends the verification of identities by means of particular attribute combinations and / or unique identifiers in order to be able to determine with certain probability whether a particular person appears at first time or has already been captured. In the latter case, the new attribute frames are assigned to the existing identity with the new or modified information and are made available again.

If the existence of an identity cannot be verified, the person is registered by assigning to the attribute frame an identifier which is generated according to certain rules and which is kept in a one and only one relation with this person during its whole lifecycle. The unique identifier is always identical in the respective namespaces of all subordinate IdM systems if the existence is given.

In order to generate a worldwide unique identifier in this decentralized scenario, the use of so-called Object Identifiers (OIDs) is recommended, which represent unique labels worldwide and are standardized in ISO / IEC 9834-1[6].

#### 2.6.1.4 Interregional Level

For this level here, it shall exemplarily sketched a service on which the DFN-Verein[7] in cooperation with the Albert-Ludwigs-Universität Freiburg that is currently providing. It is about building an infrastructure for authentication and authorization (shortly: DFN-AAI), which is designed to simplify and unify the existing procedures for controlled access to information.




In order to be able to provide this infrastructure, whose technical implementation is based on the middleware "Shibboleth", with a defined quality of service, with the partners as federation

[6] ITU-T Recommendation X.660 (1992), ISO/IEC 9834-1: 1993, Information Technology – Open Systems Interconnection – Systems Management Overview – Procedures for the Operation of OSI Registration Authorities: General Procedures

[7] Security Assertion Markup Language (SAML) is an XML-based scripting language used to exchange confidential and secure information between instances for SSO and authorization services, as well as distributed transactions. SAML consists of SAML assertions, SAML protocol, SAML bindings and profiles. (See http://en.wikipedia.org/wiki/SAML)

members, that means with the users who wish to make available decentralized web services based on this infrastructure to their users, or with providers who want to provide these services on their website are negotiated contractual arrangements.

Shibboleth is a process for distributed authentication and authorization, in order to provide web services in a secured way using SAML. In this case, a user must authenticate only once to his home device (single-sign-on) in order to be able to access to the provided web services of the providers independently of the location.

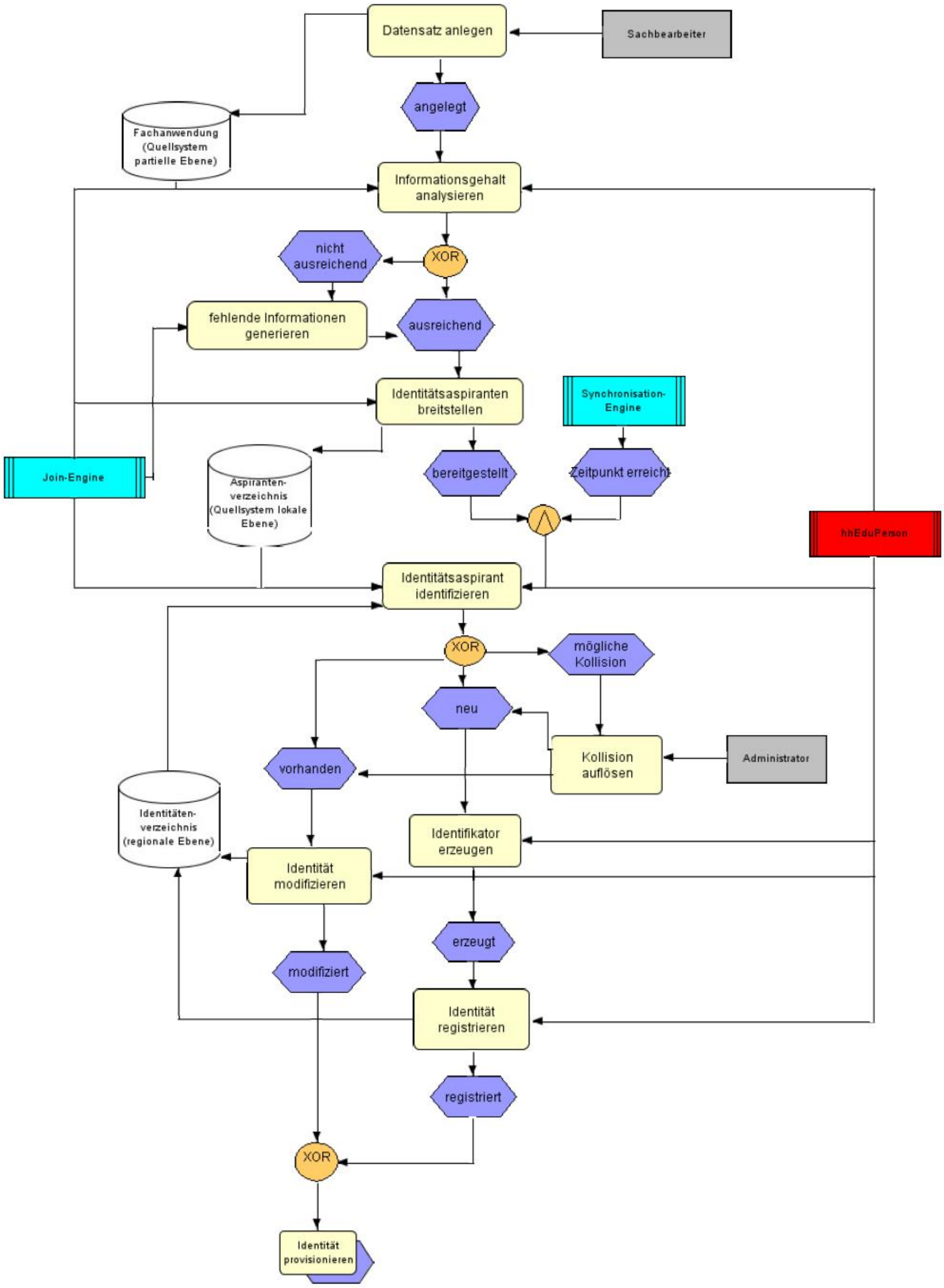
From the set of three services of "Shibboleth", the DFN Association takes over the operation of the localization service (where are you from?), while the implementation and operation of the others is the responsibility of the partners. The user, with his Identity Provider (e. g. IdM system), provides the services of identification (who are you?) and authorization (what are you allowed to do?) as well as the identity management, while the service provider manage and provide services for the secure provisioning web services.

### 2.6.2 Identification and Registration of digital Identities

The identification of a digital identity encompasses the verification within its production process and after entering the relevant attributes of a data set in a source system in order to determine with a certain probability whether a particular person is present in the system for the first time or whether it has already stored. In the latter case, the new or modified information is assigned to the existing identity and forwarded to the other IDM systems, depending on the rules. If the existence of an identity could not be verified, the person is registered, it means that a digital identity will be created and the resulted information will pass to other systems. (The following graphic describes a possible process sequence[8])

If a so-called collision occurs, that means, if it is not possible to determine reliably whether a digital identity is to be generated, then this collision must be solved by using cognitive capacity of human resources, that is, consulting an expert.

[8] The graphic uses elements and the architecture of event driven process chain (EPC), that can be employ to lay out business process workflows


Datensatz anlegen
Sachbearbeiter
angelegt
Fachanwendung (Quellsystem partielle Ebene)
Informationsgehalt analysieren
XOR
nicht ausreichend
fehlende Informationen generieren
ausreichend
Identitätsaspiranten breitstellen
Synchronisation-Engine
Join-Engine
Aspiranten-verzeichnis (Quellsystem lokale Ebene)
bereitgestellt
Zeitpunkt erreicht
hhEduPerson
Identitätsaspirant identifizieren
XOR
mögliche Kollision
neu
Kollision auflösen
Administrator
vorhanden
Identitäten-verzeichnis (regionale Ebene)
Identifikator erzeugen
Identität modifizieren
erzeugt
modifiziert
Identität registrieren
registriert
XOR
Identität provisionieren

### 2.6.3 Data Cleansing

The data cleansing (or data scrubbing) is one of the most important activities to keep consolidated the identity repositories of the whole IdM architecture. These activities include various methods for removing and correcting data errors in information systems. The errors can consist, for example, of incorrect (originally incorrect or obsolete), redundant, inconsistent or incorrectly formatted data.

Data cleansing provides a contribution to improve the quality of information. However, information quality also affects many other characteristics of data sources (credibility, relevance, availability, costs, etc.), which cannot be improved by means of data cleansing. Significant steps towards data cleanup are the duplicate detection (detection and aggregation of the same data sets) and data fusion (merging and completion of fragmentary data sets).

Caused by different data quality, data conflicts grow up at different levels of semiotics while identifying duplicates and fusing data of identities or their data records from different sources, which result of

- syntactic (data format),
- semantic (single word meaning) and / or
- pragmatic (context-related meaning, domain knowledge)

deficiencies. For example, duplicates are often unintentionally produced by spelling errors, hearing errors, letter and word scrambling, abbreviations, different spellings, and so on. Many data errors can be corrected right from the start by domain-specific normalization and / or transformation (e. g. name surname → surname, name) as well as by standardization (e. g. MM/DD/YY →TT/MM/YYYY) of the data.

## 2.7 Data Privacy Aspects

Utilizing an IdM system, personal data (e. g. name, address, roles and authorizations, relations to organizational units, etc.) are processed. In this case, the data can be collected in one place, stored on another and / or used on a third. The legal basis for this data processing is based on §8 Hochschulgesetz Nordrhein-Westfalen (HG) in conjunction with the fulfillment of tasks allocated to the universities according to §2 Hochschulrahmengesetz (HRG).

Within the framework of these processes, certain data protection regulations must be observed, which more or less influence the way in which data processing is handled in the IdM. The corresponding requirements are regulated in the Federal Data Protection Act (Bundesdatenschutzgesetz BDSG) and Datenschutzgesetz Nordrhein-Westfalen (DSG NRW), and others.

In addition, corresponding operating (Betriebsvereinbarung) and / or service agreements (Dienstvereinbarung) are taken into account, in which the processing of personal data as well as the exercise of the rights of informational self-determination are more clearly specified and manifested.

### 2.7.1 Types of Data

The different types of data that occur during their processing operations can be distinguished as follows:

- Stock data, such as master data, are permanently assigned to the concerned person. They are necessary in order to use provided applications and services. Normally, these data are such as name, address, e-mail address, telephone or fax number, date of birth etc.

- Usage and transactional data are such information those are necessary to enable the utilization and the billing of applications and services. Mainly, this includes, in particular, features for identifying the user, information on the beginning and the end, the scope of the utilization as well as information on the features used by the user. If these user data are again necessary for the calculation of costs, they are called billing data for which special compliances must apply.
- Connection data, such as IP addresses, is generated when access to provided services. They are stored, for example, to detect security violations.

### 2.7.2 Admissibility

Since the collection, processing and use of personal data affects the basic right to informational self-determination, before the implementation of an IdM system it have to be examine how far  the storage of the assigned data is permissible with regard to appropriate legislation and tariff arrangement regulations and / or the conscious agreement of the concerned person.

### 2.7.3 Necessity, Data Reduction and Data Economy

The collection, processing and use of personal data is only necessary if the respective task cannot be fulfilled without the specific data or alternatively with disproportional difficulties or prohibitive effort as well as to late or incompletely.

A priori, a data collection "in stock" (data retension) is not permitted and the stored data should be deleted at the earliest possible date, or at least the personal reference can be loosened by anonymization or by pseudonymization.

In the design of the IdM system, it is important to ensure that during processing and storing the quantities of personal and / -related data can be kept to a minimum or even avoided.

### 2.7.4 Purpose Limitation

The rule of the purpose limitation is to ensure that data of digital identities are processed only for the purpose for which they were collected. The purpose of the data processing follows the specific task for the fulfillment to which the data were collected. Data processing for a purpose other than the originally defined purpose is only permitted on a legal basis or then if the concerned person has given its consent. This also applies if the data within, for example, a university is to be passed to another place with another task beyond simple auxiliary functions.

### 2.7.5 Transparency and Rights of Rectification

The informational self-determination right for those whose digital identities are stored requires knowledge of the structure of data processing, the data processing processes, the technology used and the data streams. Each specialized application of, for example, a universities must inform the concerned persons about the processing of their personal data and the data processing bodies.

Affected persons whose digital identities are stored are entitled to correct, delete and block their stored data stored. Data must also be blocked if their correctness is not clear.