The eID-Terminology Work of FutureID

Bud P. Bruegger, Moritz-Christian Müller

Identity Management
Fraunhofer IAO
Nobelstraße 12
70569 Stuttgart
bud.bruegger@iao.fraunhofer.de
moritz-christian.mueller@iao.fraunhofer.de

Abstract: The paper reports on the experience of the FutureID project in the creation and use of an eID terminology so far. A major part of work has reviewed the state of the art in eID Terminologies. Five existing terminologies have been compared and analyzed in detail to yield unexpected and surprising results. On this basis, FutureID has designed its approach for creation and use of an eID terminology that is currently being implemented in the project. It is hoped that the terminology, its approach, and the related infrastructure will constitute a general community resource, well beyond the scope and duration of the project.1 Section heading

1 Section heading

FutureID[1] – Shaping the Future of Electronic Identity – is a Collaborative Project of the EC's Seventh Framework Programme and is coordinated by Fraunhofer-Gesellschaft. It started in November 2012 for a duration of three years. The consortium is composed of 19 partners from 11 European countries, and combines the multi-disciplinary and complementary competence. The project builds a comprehensive, flexible, privacy-aware and ubiquitously usable identity management infrastructure for Europe. It integrates existing eID technology, trust infrastructures, emerging federated identity management services, and modern credential technologies. It creates a user-centric system for the trustworthy and accountable management of identity claims.

FutureID faces particularly difficult challenges in respect to terminology. Its work is often interdisciplinary and combines technical, legal, economical, and societal aspects. On top of this, FutureID's objective is to reach an unprecedented level of integration that comprises many different types of eIDs, federation protocols, trust infrastructures, platforms, and more, each of which often comes with its specific concepts and thus terms. In comparison, Stork, one of the most prominent eID interoperability projects,

[1] www.futureID.eu

decided to use a single federation technology for its whole infrastructure. This is the setting in which the terminology work described in the paper is embedded.

2 Purpose of Terminology Work in FutureID

In such an environment, the efficiency with which project partner can discuss, conceive, and implement various integration concepts depends largely on precise communications and a **common understanding** of the involved concepts. A common understanding is achieved when the meaning associated with words is the same for all participants. Hence, a well-defined terminology that describes the precise meaning of the most important concepts is of vital importance for efficiency. When different partners create components of a whole, different understandings can lead to serious problems when assembling these. A good terminology can go a long way to avoid such problems.

A terminology identifies and defines the concepts that we use to reason and communicate in the field of eIDs. eIDs are a young discipline in the process of moving from a collection of technologies to a more mature science. Our current position in this process is closely related to the concepts we use, the degree of consensus that exists on their meaning, and the degree to which concepts are defined by a single technology or are more general and thus applicable across technologies.

To design a strategy for dealing with terminology in a project such as *FutureID*, it is important to first understand what elements of an eID terminology already exist and at what level of maturity they are.

3 Review of the State of the Art in eID Terminology

To review the state of the art in eID terminology, existing terminologies and glossaries were identified, loaded on a MediaWiki, and compared and analyzed with ad-hoc scripts. The description of this work comprises the main part of this paper.

3.1 Previous Work on eID Terminology

First, links to **fourteen** terminologies were compiled and **seven** were selected. For each, an ad-hoc parser was written to extract terms and their definitions from the original format (mostly PDF) and load them on the MediaWiki. F**ive** were selected since they were comparable in scope and all focused on technical aspects of eIDs (see Table 1). The year col. indicates the range between first and last available version.

Terminology	Label	Year	Reference
Modinis-IDM Glossary	Modinis	2005	[Mo05]
Identity Management for eGovernment (IDEM)	IDEM	2005-2007	[HuAl07]
STORK Glossary	Stork	2008-2009	[Pi08]
U.S. IdM Task Force Glossary	US		[Id05]
ISO/IEC 24769-1	ISO	2011	[Is11]

Table 1: Overview of the Intersection of five Terminologies

The terminologies that were also parsed and loaded but excluded from this analysis were the Eurosmart Glossary [Eu13] and the glossary contained in the European Draft Regulation on Privacy Protection [Eu12].

3.2 Intersection Analysis of Five Existing eID Terminologies

Each of these terminologies was considered to be a set of terms and the intersection of all five terminologies was computed. Where synonyms were defined, they were treated like normal terms (see Table 2). It lists the total number of terms (incl. synonyms) defined by the terminologies, the number of "isolated terms", i.e., terms that occur only in a single terminology, and the percentage of isolated terms in each terminology.

Characteristics	Stork	IDEM	Modinis	ISO	US
Total terms	123	195	45	43	125
Isolated terms	67	132	4	18	72
Percentage	54%	68%	9%	42%	58%

Table 2: Overview of the Intersection of five Terminologies

It is evident, that the percentage of isolated terms is above 50% for all terminologies but Modinis, which was taken into account by Stork and IDEM, and partly by ISO. These latter two terminologies are also relatively small. So it is more likely to focus on the most important concepts that are shared by other terminologies.

Table 3 shows in how many terminologies each unique term is contained. Out of the 377 unique terms, only 10, i.e. a bare 2.7% are contained in all five terminologies. Even the terms contained in at least three out of the five terminologies are only 36, i.e., less than 10%. Consequently, 90.5% of all unique terms are contained by at most two terminologies, and of these 77.7% in only one.

Characteristics	No. Terms	Perc.	accum. No.	accum. Perc.
terms contained in all 5 terminologies	10	2.7%	10	2.7%
terms contained in only 4 terminologies	14	3.7%	24	6.4%
terms contained in only 3 terminologies	12	3.2%	36	9.5%
terms contained in only 2 terminologies	48	12.7%	84	22.3%
terms contained in only 1 terminology	293	77.7%	377	100.0%

Table 3: Detailed Intersection of five Terminologies

The terms contained in all terminologies are:
anonymity, attribute, authentication, context, credential, entity, federated identity, identifier, identity, identity management

The terms that are included in four terminologies are:
identification, pseudonym, enrolment, identity provider, relying party, access control, assertion, delegation, digital identity, principal, privacy, role, trust, trusted third party

The terms that are included in three terminologies are:
identity federation, verifier, partial identity, verification, personally identifiable information, characteristic, confidentiality, corroboration, identified entity, profile, registration, token

This intersection analysis illustrates that there is an unexpectedly low level of agreement on the key terms/concepts of identity management. This is not what would be expected from a mature field. It much rather indicates that the field is young and immature and that its maturation will be measurable in an increasing level of agreement. The hypothesis that the lack of agreement is caused by cultural differences between a Europan and a North American view cannot be sustained as shown in Table 4. Here, only the three European terminologies, that are even linked by their lineage, are intersected. While the situation improves some, the lack of agreement still subsists.

Characteristics	No. Terms	Perc.	accum. No.	accum. Perc.
terms contained in all 3 terminologies	26	9.2%	26	9.2%
terms contained in only 2 terminologies	28	9.9%	54	19.1%
terms contained in only 1 terminology	229	80.9%	283	100.0%

Table 4: Intersection of the three European terminologies

Further evidence for this is given in Table 5 where the agreement between pairs of terminologies is measured. Through the band, the overlap remains very low.

Terminologies	Common Terms	Percentage
Modinis-Stork,	30	21.7%
Modinis-IDEM	36	17.6%
Modinis-ISO	45	20.5%
Modinis-US	23	15.6%
Stork-IDEM	40	14.4%
Stork-ISO	15	9.9%
Stork-US	34	15.9%
IDEM-US	36	12.7%
IDEM-ISO	17	7.7%
ISO-US	22	15.1%

Table 5: Overlap between Pairs of Terminologies

With the help of Euler diagrams, graphical representations of intersections are particularly suited to understand the level of agreement of terminologies. The resulting diagrams are presented in the following figures 1 through 5. Each set of terms is represented by an ellipsis. Each subset of the intersection is labeled with the number of terms it contains. For example, the number under the terminology name represents its isolated terms.

Figure 1 shows the intersection of the European terminologies Stork, IDEM, and Modinis. Since Modinis was taken into consideration in both, the creation of the Stork and IDEM terminologies, it is not surprising that most of its terms were adopted. Only five terms remained unadopted by the other terminologies; they are:

identity management application, nym, privacy enhancing technology, proxy, unique identity

Since the Modinis Glossary is relatively small and was mostly absorbed in the other two terminologies (30 and 36 terms out of 45 for Stork and IDEM), it is less relevant than the other two to understand the relationship with the US and ISO terminologies.

The intersection in Figure 1 shows further, that apart from the common Modinis terms, only 14 additionally added terms are in common. This compares to 93 additional terms in Stork and 159 in IDEM. Surely, this strong divergence after adaptation of a common core of Modinis term is amazing.

Fig. 2 shows the relation of the two major European terminologies with that of the U.S. IdM Task Force. Again, a strong divergence of the three terminologies is evident. The hypothesis of possible cultural differences must be rejected also here, since the communality between the European terminologies is in all respect comparable to the communalities with the US terminology.

Fig. 3 shows that the major European terminologies have about the same level of agreement with the ISO terminology. The level of agreement seems even less that with the US IdM Task force. This may be partially explained by the smaller size of the ISO terminology.

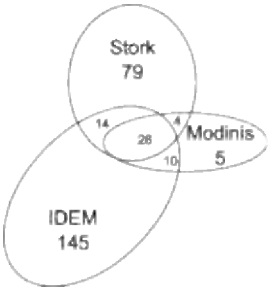

Figure 1: Stork-IDEM-Modinis

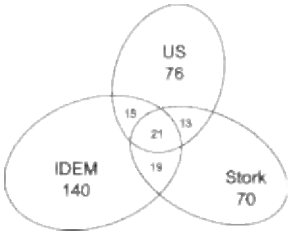

Figure 2: Stork-IDEM-US

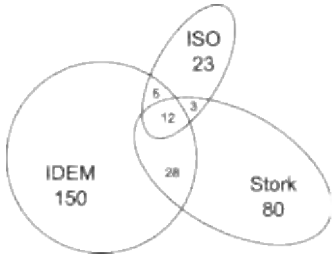

Figure 3: Stork-IDEM-ISO

Also partially explainable by the size of the ISO terminology is the conclusion from Figures 4 and 5 that the European terminologies have more communality with the US IdM Task Force one than with ISO. Also visible is that ISO has only a slightly higher level of agreement with the US terminology compared to the European ones. Figure 6 compared to Figure 1 illustrate that while Modinis has strongly influenced Stork and IDEM, it has much less communalities with US and ISO. This is probably explainable by the fact that the European terminologies have taken Modinis into account.

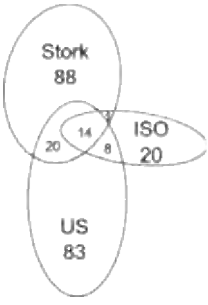

Figure 4: Stork-ISO-US

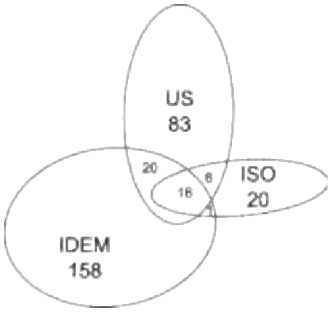

Figure 5: IDEM-ISO-US

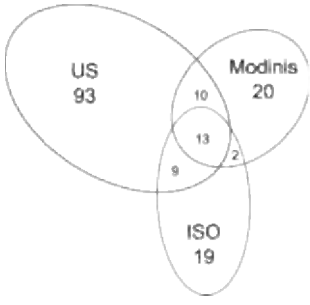

Figure 6: Modinis-ISO-US

3.3 Conclusions on the State of the Art of eID Terminologies

The detailed analysis of communalities of the existing terminologies illustrates, that the level of agreement is unexpectedly low. Apart from Modinis, whose choice of terms show a strong overlap with Stork and IDEM, no clusters or families of terminologies with a higher level of agreement could be found.

This unexpected result has a strong impact on terminology work in the field of identity management. It is not possible to start from a consolidated set of core terms. A consensus decision on which the major concepts of identity management are has yet to be made. This is even more astonishing since the divergence is already present in the choice of terms, while semantic discrepancies between different definitions of the same term would have been expected to be more likely. Terminology work has to attempt to avoid creating yet another diverging glossary defined by a relatively small and closed group of experts. Therefore, for furthering the field, it is necessary to launch a collaborative process of consensus building that should attempt to reach out to as many experts as possible, integrating multiple disciplines and backgrounds.

4 The FutureID Approach to Terminology Creation and Use

A glossary *document* managed by an editor, as used by most pre-existing projects, is evidently an ill-suited choice for supporting a large-scale consensus process. FutureID therefore decided early on to use a collaborative semantic wiki (MediaWiki with a semantic extension) to support the terminology work.

A wiki-based approach makes it possible to load the existing terminologies, support analysis as the above, and to more easily see different terminologies side by side in order to compare them. For example, the current wiki makes it possible to see all related existing definitions of a given term in order to make more founded decisions about its semantics. A semantic wiki also allows to model properties and relations that would not be possible in a paper-based approach. This allows for both a richer expression of semantics, as also support of processes. For example, it is possible to capture the technologies for which a concept is relevant, state the community who agrees on a given definition, or model synonyms. A wiki can also support editorial processes such as assigning certain terms to certain editors or capture the status of a review process. The wiki's capability of easily annotating anything is an important support for discussions and consensus building.

Another aspect that the FutureID terminology work attempts to address is motivation. Terminology, surely when managed as a glossary document, is probably "boring" for most experts. Experience of past projects have demonstrated that is far easier to select and define terms, than getting them used in a consistent way by all project participants.

A machine readable management is therefore important in order to create tools and views that confront project partners with terminology and provide useful services. For example, *FutureID* has created a tool that automatically creates custom glossaries for project deliverables. A "motivational design" of the approach seems to be a major success factor for the creation of a valid terminology.

5 Conclusions and Future Work

While *FutureID* has managed to set up an initial infrastructure and set of tools for the creation and use of a terminology within the project, much of the work of deploying to participants them is only just starting. Social challenges and collecting experiences that identify new needs for tools and infrastructure are still ahead of us. We hope that *FutureID's* work will make a major step in the maturation of the eID terminology and thus the "science of eIDs". We also plan to gradually extend the community beyond the project itself and create an infrastructure that can remain as a resource beyond *FutureID's* duration.

References

[Eu12] Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), Article 4 'Definitions', page 41 (printed, 42 electronic),http://ec.europa.eu/justice/data-protection/document/review2012/com_2012_11_en.pdf, 2012.

[Eu13] Eurosmart Glossary, http://www.eurosmart.com/index.php/glossary.html.

[HuAl07] Xavier Huysmans and Brendan Van Alsenoy, IBBT, Identity Management for eGovernment (IDEM), Deliverable 1.3 Conceptual Framework, Annex I. Glossary of terms (v1.07), https://projects.ibbt.be/idem/index.php?id=161, 2007.

[Id05] U.S. National Science and Technology Council (NSTC), Subcommittee on Biometrics and Identity Management, Identity Management Task Force Report, Annex K, Identity Management Glossary, http://www.biometrics.gov/documents/idmreport_22sep08_final.pdf, 2008.

[Is11] ISO/IEC24760-1, Information technology — Security techniques — A framework for identity management —Part 1: Terminology and concepts, First edition, 2011-12-15, chapter 3: Terms and definitions, http://standards.iso.org/ittf/PubliclyAvailableStandards/c057914_ISO_IEC_24760-1_2011.zip, 2011.

[Mo05] Common Terminological Framework for Interoperable Electronic Identity Management, Consultation paper, v2.01., https://www.cosic.esat.kuleuven.be/modinis-idm/twiki/bin/view.cgi/Main/GlossaryDoc, November 23, 2005

[Pi08] Ana Piñuela (Ed.), STORK Glossary (v2.0), https://www.eid-stork.eu/index.php?option=com_processes&Itemid=&act=streamDocument&did=615, 2008.