

Putting Identifiers in the Context of eHealth

Introduction of a Model

Rieks Joosten¹, Diane Whitehouse², and Penny Duquenoy³

¹ TNO, Netherlands

rieks.joosten@tno.nl

² The Castlegate Consultancy, United Kingdom

diane.whitehouse@thecastlegateconsultancy.com

³ Middlesex University, United Kingdom

P.Duquenoy@mdx.ac.uk

Abstract. eHealth is becoming an increasingly noteworthy domain in terms of public sector exploitation of information and communications technologies. Appropriately identifying the users of electronic health systems is a major contemporary challenge. The appropriate identification of eHealth systems' and services' users is one of its core areas of concern. This paper develops a particular problem statement that relates to the notion of identifiers in eHealth, outlines its conceptual background, and defines a set of solutions to the problem outlined. It lists a variety of use cases or examples against which the issues can be tested (these are further explored in a parallel paper [13]), and proposes some possibilities for future work. In particular, the paper describes the results of a 2007 workshop that explored all of these notions. While the paper bases its orientations in a general European framework, the main examples and illustrations used by the authors come from experiences in the Netherlands.

1 Introduction

Probing the concept of identifiers, what they are, how they are used, and how they can be managed in an information systems environment – particularly one that has relevance to the eHealth domain – is the focus of this paper. The notion of a unique identifier, and what it is intended to do, is not difficult to understand. However, there is a considerable challenge in transferring an identifier that is unique from its own small context to a wider context or scope. A conceptual shift is required that can take into account the nature of the technologies with which we are surrounded whether in eHealth or in some other public sector services field. In this paper, therefore, we present an approach that takes a perspective based on the requirements of the computer system and its capabilities. The model presented here illustrates the various scoping difficulties involved (see also, [8, 10]).

The writing of this paper was stimulated by a workshop held at the IFIP WG 9.2, 9.6/11.7, 11.6/FIDIS summer school held in Karlstad, Sweden (August 6-10, 2007) on 'Identifiers and eHealth' run by Diane Whitehouse and Penny Duquenoy. As a result of the combination of persons present, particularly Rieks Joosten of TNO (Netherlands), the session focused on developing an underpinning model of

Please use the following format when citing this chapter:

Joosten, R., Whitehouse, D. and Duquenoy, P., 2008, in IFIP International Federation for Information Processing, Volume 262; The Future of Identity in the Information Society; Simone Fischer-Hübner, Penny Duquenoy, Albin Zuccato, Leonardo Martucci, (Boston: Springer), pp. 389–403.

identification that can be applied to the eHealth domain. The underlying purpose of the model is clearly explained in the second section of this paper that follows.

2 Purpose

The particular contribution of the workshop, and of the paper resulting from it, is to propose a model for identifiers, and a working procedure for using that model. When applying the contribution in a given setting, the following results can be expected:

1. ability **to distinguish identifier-related issues from non-identifier issues** in that particular context. This approach is necessary so as to verify that a selected issue is **actually identifier-related** and that the model can be used according to its intentions (rather than as a solution to larger, global, problems).
2. opportunity to obtain an absolutely **clear perspective on the particular problem's identifier-related issues**. While this opportunity is a pre-requisite for problem-solving, it is not a solution in itself (although, in some cases, it may suffice to see the proper solution).

With these two clear intentions, identifier issues in a particular situation can be addressed by the people/personnel responsible for the situation. It can enable them to understand the setting fully, and apply the relevant contextual and regulatory conditions, and so on. The actual solution of individual problems is outside the scope of our paper. The paper merely seeks to use **an example illustration** of the proposed model, and a suggested method for putting it to work.

The paper is intended to be the technical report of the outcomes of a particular summer school workshop. By applying the model to the eHealth sector, as we did in this workshop, the actual process of examining the various test cases shows how the model can unravel the complexities of combining data sets. As this possible transformation lies at the core of eHealth, it is crucial to understand where the problems arise. Having such a model or tool available may clarify the relationships between different systems of data organisation, and may help in resolving these difficulties. The paper particularly aims at helping to make software architects and designers more aware of the pervasive nature of identifiers, whether in eHealth or more widely, and helps to indicate how they should be handled. In terms of the personnel working in eHealth, it enables those concerned to understand the considerable complexities facing the domain. Hence, the paper does not solve particular problems. Rather it aims to clear a kind of 'fog', and to make issues clearer for the personnel in charge of a particular problem area.

We wish to emphasise that there are yet other domains in which this paper does not stray. The paper is not intended to survey the entirety of the identifier problems throughout the whole of the European Union, nor more widely internationally: even so, it does use as a particular illustration the eHealth context in a particular European country (the Netherlands). It is nevertheless evident that the example could be explored in terms of other European Member States. Surveys of the different Member States' approaches to identification issues have taken place under the aegis of other

projects and initiatives, e.g., the European Commission co-financed i2health project. Nor does the paper provide a critical social, ethical, or legal review of the issues surrounding identifiers¹. Overall, therefore, the paper does not seek to promote the notion of a unique personal identifier². While it is clear that some Member States and some other international countries have chosen that particular approach³, the orientation for direction on such issues is the domain of European, and Member States, policy-makers and decision-makers.

3 Background on eHealth

The domain of eHealth has grown from research and development initiatives undertaken in Europe from the 1960s onwards, and which have thrived especially during the past twenty-year period. eHealth has progressed from an area of theoretical exploration to one which is being put into practice today throughout a growing number of European countries. As the research team of the i2Health project on eHealth, an eTEN project co-financed by the European Commission, emphasised:

“Information and information technologies (ICT) are currently deployed on a broad scale in healthcare. European Member States are struggling to deploy promising eHealth services under high pressure ... The fundamental use case of eHealth is making medical information quickly available for a better and more efficient treatment of patients. ... eHealth can be reduced to a transmission of information on someone (who?) between actors (again who?). An appropriate implementation of identification management guarantees correct attribution of information and access control to guard privacy.” [7]

eHealth comprises a wide range of organisational forms and technological applications. The domain is said to describe ‘the application of information and communications technologies across the whole range of functions that affect the health sector’ [2]. Each application has a different requirement for the quality of identification associated with it: whether the application is, for example, reimbursement, electronic booking, clinical information, electronic patient records (or ‘patient summaries’), emergency data sets, ePrescribing, or some Internet-based publicly-available information.

¹ For readers interested in these issues, see the referenced European Directive 95/46/EC [3] and the FIDIS [5] project deliverable on profiling techniques in the field of ambient intelligence.

² Note that doing so would have been a proposition for a solution to a problem, which is explicitly not one of the results we expect of our model.

³ Examples include the national registry number in Belgium, the BSN – the citizen’s service number or social security number - in the Netherlands, and the social security number in the United States of America and also in Canada.

While individual European Member States are responsible for the provision of healthcare⁴, it is becoming increasingly evident that the challenges facing national health systems are no longer separate dilemmas but rather problems that confront the whole of Europe. A focus on the issues surrounding patient, health professional, and institutional identification has expanded particularly as the different European Member States have progressed to focus on the provision of electronic healthcare (or medical) records for their respective individual citizens. As the potential connectedness and interoperability of Europe's health systems and services expands both organisationally and electronically, it is timely to ensure that the users of eHealth systems are appropriately identified. It is only right and proper that the appropriate person (or persons) is accessing, using, and manipulating the appropriate data at the appropriate time and in the appropriate place (this is an adaptation of a statement by Duquenoey et al. [4] and HEHIP [6]). These issues will increasingly come to the fore as cross-border provision of healthcare is promoted, and a launch of a large-scale pilot on eHealth interoperability takes place. In this paper, the example of a single country (the Netherlands) is used rather than using exemplars from the whole of the European Union.

4 Problem statement with regard to identifiers

Fundamental to the 2007 Karlstad summer school workshop was the notion that identifiers are key to eHealth, and somehow also seem to present a number of challenges. Given the perceived importance of the identifier issue, the original focus of the workshop, namely "the exploration of social and stakeholder-related issues related to identifiers in eHealth" was shifted towards "developing an underpinning model of identification that can be applied to the eHealth domain". Hence, the workshop was inductive in character: it involved a quest or search for the problem statement.

This section therefore explores some of the general characteristics of identifiers in the articulation of the problem that is needed to be addressed. Wherever possible, rather than selecting a generic workplace, organisation, or surveillance setting, we have deliberately chosen – as we did in the workshop – a healthcare context to illustrate the points made.

We have established a number of criteria for identifiers. This section explores two of these identifier criteria (IC), that is, IC1 and IC2, and the process used to develop them, which are explained in greater detail below. Sections 4 and 5 refine the criteria still further to arrive at IC3. IC4 is outlined briefly in section 9.

⁴ Title XIII Public Health, Art. 152 of the European Community Treaty ensures a high level of human health protection in the definition and implementation of all Community Policies and activities.

To this end, we started a quest for a criterion⁵ for identifiers based on an initial proposal criteria IC1) that: *an identifier is (1) a text that (2) names someone or something*. Throwing a number of test cases at these criteria helps us to make them increasingly strong and more effective. At the same time, we come to an understanding of some of the underlying issues that need to be addressed. A number of test cases follow. They lead to an enhanced identifier criteria (IC2).

4.1 Testing first criteria for identifiers

Our first test case is this: is the text "Donald Duck" an identifier⁶? According to IC1, identifiers must satisfy two requirements. First, an identifier must be a text and, secondly, it must name someone or something. "Donald Duck" satisfies both requirements: it is a text and it names a Disney character with whom we are all familiar.

Our second test case places us in a hospital ward where there is a patient called Jane Smith. According to IC1, the text "Jane Smith" is an identifier because it names this particular patient. Now suppose that another patient is brought in, and she happens to be called Jane Smith too. What occurs next is that the ward nurses would detect⁷ the ambiguity of this name and as a result would start to concoct some form of consensus as to how to distinguish both patients. For example, they might come to call one of the patients "Jane", and the other one "Mrs. Smith" or "Jane with the grey hair". Whatever the consensus, the result is that, in the ward setting, both patients will be assigned (different) names that actually uniquely identify them. Note that "Jane Smith" has continued to be an identifier, at least according to IC1. However, the nurses started a process that resulted in there being other identifiers (according to IC1). They did this automatically and unconsciously to ensure that these identifiers would have the property of uniquely identifying the patients in the ward.

Apparently, the nurses (unconsciously) used a criterion for identifiers that incorporates this property of uniqueness. So we propose another criteria for identifiers, IC2: *an identifier is a text that names precisely one someone or something*. IC2 (which still works in the "Donald Duck" example). This example means dismissing "Jane Smith" as an identifier as soon as a second patient called "Jane Smith" enters the ward.

The ward nurses start a process to come up with new IC2-compliant identifiers as soon as there are patients in the ward without a compliant identifier⁸. Note that "Jane Smith" has remained a name for either patient, which implies that identifiers and

⁵ We use a criterion (or criteria) rather than a definition. This is because the correctness of a definition may be disputed, but a criterion is something that anyone can simply apply.

⁶ Although we use a fictional character to illustrate our test, we generally assume that the eventual identifier used will (if it is a name) be a factual and authentic name.

⁷ This is an example of how the violation of a criterion leads to an action. This is a fundamental principle and is explained further in section 6.

⁸ A consequence of using criteria for identifiers rather than definitions, is that violations of criteria may serve as triggers for processes that restore compliance. See section 6.

names are different things. The generally accepted idea that the terms "identifier" and "name" have the same meaning is hence flawed! The nurses' behaviour of automatically disambiguating patient identifiers by creatively thinking of different names for the patients is readily recognisable in other organisational settings. Apparently, it is something we humans do unconsciously. This does not mean however that, as human beings, we can actually and accurately describe what it is we are doing. To illustrate this point, everyone can tell that sentence this not proper English grammar to adheres, even though readers are unlikely to be able to list the grammatical rule(s) that are violated in the phrase.

This approach might explain why many information technology (IT) specification documents do not explicitly specify how the system should handle identifiers so as to guarantee the unique identification property; after all, their authors are only human. As a consequence, systems in the eHealth domain, as well as in other domains, may contain flaws related to identifiers, e.g., there will be systems that assume names and identifiers to be the same (which, as we have seen, is a flawed assumption).

From this last test case, we conclude that the identifier problem is one of automation design rather than of knowing or not knowing how to deal with identifiers. In particular, it is necessary, at least for software architects and designers, to have an explicit model of what constitutes an identifier at their disposal.

4.2 Testing enhanced criteria for identifiers

Let us return to the example of the two women who are both called "Jane Smith". We can easily think of names that they will be called in another setting, such as in their homes. There, they might be called "Jane" but they will also have other names such as "Mum" or "Darling". However, these last two names, while they are unambiguous in Jane's own home, are also used in other households; there they are unambiguous as well. In the setting of most (English language) households, the name "Mum" would unambiguously identify one person. So, here we see that one name can be associated with multiple identifiers, depending on the setting. Here too, people have an automatic way of dealing with identifiers. By being sensitive to the particular setting in which they find themselves and switching contexts unconsciously when necessary, people have no problem with the fact that a single name may have the property of uniquely identifying different people or things in different settings. People apparently make these kinds of adaptations all the time.

Looking at this case in another way, we see that while "Mum" has a unique identification property in a given household, this is no longer the case when we stretch the scope to include the entire street, city, or country. A text can be used as an identifier when we consider that a certain scope (setting) may no longer be an identifier and when the scope under consideration is enlarged.

This description might explain what happens in eHealth and other software applications. For example, a software application in a hospital is likely to have been designed for use in a particular setting (or scope). However, as the scope that the application has to function in grows (e.g. as a consequence of a hospital merger or if

external parties have been given access to the particular hospital system), the application may not be able to follow suit. It may be the case that its identifiers are no longer identifiers in the new, enlarged scope and that it is also not equipped with the program code that allows the systems to resolve the issues in a similar way that human beings do unconsciously.

From this exploration of a number of tests of IC1 and IC2, we conclude that the identifier problem is also one of increasingly large, or rather, enlarging scopes rather than the previous small ones. In eHealth terms, this could not only be a problem when local systems are increasingly connected. It could also be a challenge when nationwide systems try to bring smaller systems together, and even – quite possibly – if European Member States were at some hypothetical point in time in the future ever to start to look at ways of integrating their health-related systems. With regard to the transfer of information, from hospital to hospital, for example, this paper deliberately does not – as already stipulated – address those legal or regulatory issues which might be relevant.^{9,10}

To conclude, in the 2007 Karlstad summer school workshop, our quest for identifier criteria led us to the four following observations:

1. Human beings deal with identifiers mostly unconsciously. However, the architects and designers of software systems need a conscious, explicit idea of what identifiers are if they are to design software programs that exhibit similar behaviour to that of human beings.
2. Whether or not a text (label, name, number) is considered as an identifier depends on the setting (scope) in which it is used.
3. A text may identify (i.e., be the name of) different things in different settings.
4. An identifier may lose its unique identification property when the scope (setting) that it is defined or used in is enlarged.

In the following section, we outline how we proposed in the workshop to deal with the identifier challenges uncovered.

⁹ Readers interested in these aspects may refer to the FIDIS [5] project deliverable which draws attention to the ‘finality principle’ and the ‘legitimate purposes principle’. International transfer of data is covered by the ‘safe harbour principle’. The notion of ‘scope’ used in this paper provides a formal of demarcation but one that is not necessarily related to legalities or the legitimacy principle, jurisdictions, or domains. The challenge of transferring information between two hospital jurisdictions are not identifier issues that arise when transferring information between two scopes – in the sense used in this paper.

¹⁰ It might be more realistic to add the notion of introducing the person’s date of birth. However, this model is not intended to determine whether it is right or wrong to use names and dates of birth as identifiers. Rather, the model is used to verify or validate the quality of a criterion that can distinguish identifiers from non-identifiers.

5 A proposal for a resolution of the problem statement: a new model

In this section, we set out to deal with the particular challenge of enlarged scope. Dealing with this problem leads us to outline two more identifier criteria, IC3 and IC4. IC3 is dealt with in detail here. IC4 is treated only briefly in section 9.

Steps towards a solution to this set of observations would be to make the relation between identifiers, scopes and texts (names, labels, and bit strings) explicit, to apply this model in various use cases to see whether it works, and to detect any possible or associated limitations. The model we propose is quite simple, and can be worded as identifier criterion IC3: *an identifier is a symbol¹¹ that, when interpreted in a given scope¹² (setting), uniquely identifies an entity*. We use the term ‘identifier-symbol’ for symbols that are part of an identifier. We define ‘name’ as a pair (s, e) where s is an identifier-symbol of some identifier that uniquely identifies entity e . A visual representation of this model is shown in Fig. 1 below:

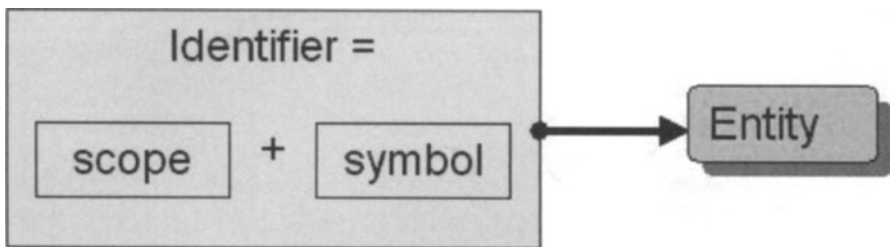


Fig. 1. An identifier is a symbol that, when interpreted in a given scope, uniquely identifies an entity.

Registration within a given scope (i.e., the creation of an identifier), means that the combination of this scope and a symbol (i.e., an identifier) is stored, and is linked to an entity¹³. Hence, the symbol is a name for that entity within the given scope. Within the scope of the hospital ward outlined in our second test case example, the new names of the two Mrs. Smiths were registered as the consensus developed on these two names was established. Likewise, symbols may be registered as identifiers in the databases of the various healthcare institutions that point to examples such as electronic health (or medication) records, and employees’ files.

¹¹ As this model is intended for personnel such as information technology (IT) professionals, we prefer to use the term ‘symbol’ rather than ‘text’ so as to allow for other data types, such as pictures or photographs to be used as a symbol.

¹² Similarly, we prefer to use the term ‘scope’ over ‘setting’ as it aligns well with terminology used in design and programming languages. In particular, the term ‘identifier scope’ is well defined in programming languages, and it is our sense that the model might be an extension of what is used there already. In other circumstances (in the 2007 Karlstad summer school itself, for example), we used the term ‘context’.

¹³ An e .

Pursuing this particular model or criterion has several consequences:

1. Every identifier uniquely identifies an entity¹⁴. Note that in computers, entities can be files, database records, objects, etc. In eHealth, entities can be patients, medical staff, or hospital administrators. Also, note that this does not imply that every entity must have a unique id-symbol.
2. Interpretation of an identifier symbol, i.e., associating a meaning to that symbol, is specific to the setting (scope) in which this is done, and its meaning is the entity (associated with the identifier).
3. Within every scope, there must be guarantees for the unique identification property of its identifiers, i.e., guarantees that every identifier satisfies IC3 and continues to do so over time. Which guarantees are selected and how they are operated is a design decision. However, particular attention should be paid to the registration of symbols to ensure that not only the symbol (text label) itself and the entity it references are registered, but also the scope within which the symbol names that entity. Also, attention must be paid to the situation where symbols cross the borders of scopes, e.g., when a system receives a message containing symbols from another system in another scope, or if the scope is enlarged, e.g., the system is supposed to work with data from other resources.
4. Within a given scope, symbols may only be used reliably if they are an identifier-symbol within that very scope, or if the scope in which they are an identifier-symbol is explicitly known. If not, such symbols bear the risk of being interpreted ambiguously (i.e., not being identifier-symbols any more). Note that a violation of this rule can happen when the scope is enlarged – it is then no longer the same scope. Also, this may happen if the identifier's symbol is interpreted in a completely different scope.
5. If a system within a certain scope is required to handle a symbol that is part of an identifier from another ('foreign') scope, the software designers of that system must have defined a way to interpret this symbol in a way that is useful within the system's scope/setting. This can be done e.g., by means of a 'dictionary' (i.e., a translation table or a correspondence table) that, to use this analogy, translates the identifier-symbol of the first identifier into an identifier in the system's scope. Note that both identifiers may in fact point to the same or different entities. For example, both identifiers may point to a single patient. However, one of them may point to a patient while the other may point to the patient's electronic patient record. Whatever translation is useful depends on the purpose pursued. This, we feel, is the actual basis of the situation in which we deal properly with identifiers¹⁵.

¹⁴ An entity is something that has a separate and distinct existence and objective or conceptual reality. See <http://www.webster.com/dictionary/entity/>

¹⁵ The model presented in this paper has already been used in an internal prototype tool by TNO for the translation of identifiers from one scope to another as a means of coming to grips with use cases. This tool has been developed at the same time as the model and has influenced its construction.

For the automated settings that can be found within computer systems (software applications), this means there must be symbol-registries, each entry of which is associated with exactly one scope and precisely one entity. Current database technology is capable of supporting this option; it is more a question of using it appropriately.

6 Conceptual background

This section describes in some detail the conceptual background to the problem statement we have developed. The way that we have treated the identifier problem here is rooted in the recently developed, yet still unpublished, technique called ‘Ampersand’ [11, 12]. The Ampersand method can be applied by anyone who can think conceptually, is familiar with relation algebra¹⁶, and is capable of mapping the conceptual models onto reality. The method starts by making one or more statements (rules, criteria) about a given topic, e.g. IC1. The idea is that such statements are universal truths’ within the context for which they are created. Usually, such statements are about concepts and the relations between such concepts. The statements together form a conceptual model or a conceptual pattern.

The universal truth of the conceptual model is challenged by throwing at it multiple use cases (cases that actually happen in practice). For every relevant use case, an attempt is made to map it onto the conceptual model. If the attempt succeeds, the model has passed yet another test. If it fails, then this forces us either to modify the model or to think about the topic in a different way. Both possibilities brings us further forward. Either the model becomes better and more robust or our understanding of the topic is enhanced.

This technique requires that the statements behind the conceptual model are not only represented in natural language (as IC1 through IC4 are), but are also represented in relation algebra (we have, however, omitted this algebraic representation in this paper). This latter representation ensures that the statements have a formal underpinning, which means that combining multiple conceptual models can be done with the aid of automated reasoning tools. Topics that have been modeled this way can consistently and fluently be related to other topics that have been modeled using the same method. Also, the relation algebra representations form a solid, mathematical basis from which both process and service designs can be created.

An important characteristic of this technique is that the universal truth statements can be operationalised. By this we mean that violations of such statements can be detected, in particular, in automated settings. Any such event is to be treated as a trigger for an action that changes the situation in which violations exist into a situation where this is no longer the case. This is illustrated already in the second test in section 4.2, where IC3 was violated when a second "Mrs. Smith" entered the hospital ward.

¹⁶ Relation algebra is based in mathematics. For a short explanation, see e.g. http://en.wikipedia.org/wiki/Relation_algebra. Relational algebra [1] differs subtly from relation algebra. Relational algebra has to do with databases.

This violation (that was unconsciously signalled by the nurses), caused them to come up with a solution that removed the violation and restored the truth of IC3. In fact, detecting violation of universal truth statements is a powerful mechanism for designing processes, which is covered in a paper that describes the use of the statements (rules) as the definition of a process [9]). This technique could be useful in eHealth for designing healthcare processes as a set of criteria that should become fulfilled during that process.

7 Results of the workshop

In a proposed associated paper [13], we outline some of the results that emerged during the workshop and immediately afterwards. A number of examples, some of which – but not all – are directly pertinent to the field of eHealth, may have become less problematic when viewed in the light of the proposed model. We chose six examples to illustrate the problem issues:

1. an application of a first set of questions about identifiers, scope, symbol and entity;
2. identification of patients in a hospital setting;
3. how to deal with an enlargement of scope (for example in hospital systems);
4. selecting among identifiers;
5. a particular challenge uncovered in Health Level 7 (HL7);
6. a number of possible software solutions to the problem statement outlined.

8 Discussion and conclusions

The focus of this workshop was on probing the concept of identifiers – what they are, how they are used, and how they can be managed in an information systems environment. As this paper has made clear, the notion of a unique identifier and what it is intended to do, is in itself not difficult to grasp. However, what is problematic is the transfer of an identifier that is unique in one small context to a wider context (or scope, to use the term of the model). Each human being has a set of characteristics that can be used to differentiate one person from another (and are used in our everyday lives). However, transferring these characteristics to the specific, and explicit, rules that are required by an information system is, as we have seen, not feasible – at least in the same way as they are used, adapted and extended by ourselves as human beings. The example of ‘Jane Smith’ in the hospital ward demonstrated this. A conceptual shift is needed, one that takes into account the characteristics and limits of the technologies with which we are working.

What we have presented here is an approach that takes a perspective based on the requirements of the system and its capabilities. This model illustrates the scoping issues. It allows us to see more clearly what is needed in an identifier when it is expected to cross the boundaries of a constrained setting to much broader settings –

i.e. when its scope is extended. By demonstrating, through the examples and cases we have used, the changing nature of uniqueness (from ‘Donald Duck’, to ‘Mum’, to ‘Jane Smith’) we have a more effective understanding of the criteria necessary for identifiers in these settings. Through this perception, we were able to see the challenge more clearly. The problem is one of automation design and the need to have an explicit model with which software architects and designers can work (cf., Section 4.1). For example, by noting that a change in scope or number of entities constitutes a violation of the criteria, it is possible to set out in technical terms the rules necessary to restore the situation (Section 5). Addressing the problem in this way allows the use of existing technical operations (e.g., symbol-registries) rather than devising new techniques.

In applying the model to the eHealth sector the process of examining the different cases showed how the model could begin to unravel the complexities of combining data sets. As controlled, reliable database accessibility is at the heart of eHealth, it is vital to understand where the problems arise, and to have some tool available that can highlight any conflicting or inconsistent relationships that arise between different systems of data organisation in order to attempt to resolve these difficult issues.

As we have said, the techniques and standards are already in place to support our identifier model. What remains to be done is twofold: firstly, to make software architects and designers aware of how pervasive identifiers really are and how they should be handled, in particular with respect to scoping and extending this IT-notion to organisational entities, organisations and even countries; and, secondly, to make eHealth and other business professionals aware of the pervasiveness of this issue, and the impact it must have if we are to work with identifiers effectively.

9 Future work

By its very nature this workshop adopted an exploratory approach, and thus could only scratch the surface of identifier challenges as they appear in eHealth. The model, as it currently stands, has limitations: it does not yet explicitly address the problem that identifiers may no longer have a unique identification property when the scope grows. Given the possibilities for such enlargement occurring, whether in the domain of eHealth or within other public sector domains, this is an issue that requires attention. In this respect we therefore need to pursue our quest further than we have been able to do in this paper.

During the workshop we also touched on the question of whether or not an identifier must actually reference an (existing) entity. An argument can be made that, if there is an identifier that references an entity and that entity ceases to exist, this should perhaps mean that the identifier is not declassified. If this situation were to be preferable over the currently suggested solution, the associated identifier criterion IC4: would read: *an identifier is a symbol that, when interpreted in a given scope (setting), identifies at most an entity*. The challenge of the further development of IC4 has been left open for further exploration. Therefore a useful exercise would be to explore the creation of models for similar cases within or outside eHealth. A possible

topic for further work is the verification of whether or not an electronic patient file actually matches the patient that the electronic patient file is allegedly about or if an electronic patient file for a given patient actually exists. Further examples in the eHealth domain could include situations when a patient moves out of his/her country of residence or even what happens when a patient dies.

While identifiers have been addressed in this paper, identities have not. Apart from often being confused with identifiers, identities in the context of information and communication technologies are usually considered to be a set of attributes associated with a specific entity (person, thing). Additional challenges are those of the integrity (correctness, reliability) of the individual attributes, as well as whether all attributes should actually be associated to one and the same entity. Ongoing work on this issue is being undertaken in the FAIM project within TNO.

At the end of the workshop, we were pleased that such a simple model could be so powerful in explaining the use cases we threw at it. We feel that, to a great extent, the mutual learning gained by applying the model to the challenges we posed to it, in terms of examples, was a major factor in experiencing the model's power.

Others who lack such an experience may react quite differently. One participant of the summer school at which the workshop took place initially commented along the lines of "You just make the problem bigger by introducing such [a] model"¹⁷. Talking with him a bit more allowed him to clarify his statement. It put us in a position to see that what happened is that he realised the enormous amount of scopes that exist. The identifiers with the symbol "Mum", for example, show that just about every household must be considered as an individual setting. And this is only a single symbol!

Making a mental image of all these settings and their symbols can indeed be an overwhelming experience. However, seeing this large number of scopes does not mean that our model makes the problem bigger. These scopes have existed all along – we have not invented them. The only thing we have done is to enable people to become conscious of the wide variety of scopes, so that we can actually start to think about them in a rational way. This approach will equip us better to design appropriate software rather than to create designs that contain 'seat-of-the-pants' decisions. Metaphorically speaking, we clear the murky waters of the so-called 'identity swamp' so that we can actually see what challenges are out there to address. Apparently, we need to prepare people for this reality before (gently) revealing what is in the swamp... and the swamp is indeed large and requires considerable future work.

At the end of the eHealth and identifiers workshop, we experienced that the actual, practical use of this model changes the way in which we come to think about identifier problems (i.e., there was a conceptual change). This observation can be explained if our model is a good explicit model of what we human beings do unconsciously; it could generate some kind of 'Aha-erlebnis'¹⁸.

¹⁷ We are indebted to a number of attendees at the Karlstad August 6-10, 2007 summer school for their feedback in this regard.

¹⁸ We use here the term developed by the early twentieth century German psychologist and theoretical linguist, Karl Bühler, which means a form of 'insight', 'enlightenment' or 'intuition'.

Future work thus might include making people aware of this model in a way that they can actually experience its workings, and to do so using even more specific examples associated with eHealth, e.g., in a workshop setting or in next year's summer school. In this respect, it would be useful to enhance the prototype tool for identifier translation so that it can enable people to acquire the learning experience that we shared together in the workshop.

Acknowledgments

We would like to acknowledge the internal TNO project FAIM support for work in development, and also the EC co-funded project i2Health for providing the starting-point for the workshop. Paul Eveson, Connecting for Health (UK), Marc Griffiths, The Castlegate Consultancy (UK), Wiltfried Pathuis, TNO (the Netherlands), and an anonymous reviewer have also provided useful evidence and criticism. Ultimately, responsibility for the content of the paper is, however, taken by the three authors.

References

1. Codd EF (1970) A Relational Model of Data for Large Shared Data Banks CACM, 13 (6), June 1970
2. COM(2004)356 final (2004) e-Health – making healthcare better for European citizens: An action plan for a European e-Health area. Luxembourg, European Commission
3. Directive 95/46/EC (1995) of the European Parliament and the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. See in particular: Whereas 33; article 6.1 (b); article 8.
4. Duquenoy P, George C, Solomonides A (2007) What ELSE? Regulation and compliance in medical imaging and medical informatics. In: Proceedings of MIMI 2007: Medical Imaging and Informatics Conference, Beijing, China, 15-16th August, 2007
5. FIDIS (2007) project deliverable D7.3 Report on Actual and Possible Profiling Techniques in the Field of Ambient Intelligence. <http://www.fidis.net/resources/deliverables/profiling/> Accessed 5 January 2008
6. HEHIP (2003) A handbook of Ethics for Health Informatics Professionals. London, The British Computer Society
7. i2Health (2007) (Interoperability Initiative for a European eHealth Area) project deliverable D3.1b Identification management in eHealth. <http://www.i2-health.org/> Accessed 5 January 2008
8. Joosten R, Joosten S (forthcoming) Rules for Identity and Access Control. In: Fischer-Huebner S, Duquenoy P, Zuccato A, Martucci L (eds) Proceedings of the IFIP WG9.2, 9.6/11.7, 11.6/FIDIS Summer School on 'The Future of Identity in the Information Society'
9. Joosten S, Joosten R (2005) Specifying business processes by means of rules. In: Proceedings European Business Rules Conference. Amsterdam. June 2005

10. Joosten S, Joosten R (2007) Will rule based BPM obliterate Process Models?, to be published.
11. Joosten S, Joosten R, Joosten S (2007a) Ampersand: foutvrije specificaties voor B&I vraagstukken, Informatie jul/aug (42-50). This article is being translated into English.
12. Joosten S, Joosten R, Joosten S (2007b) Ampersand: Errorfree Specifications for Business-Information Problems, to be published. This article is the translation of the Dutch article cited above.
13. Joosten R, Whitehouse D, Duquenoy P (forthcoming) An identifier model applied to a health context: Results of a workshop (in preparation)