The Evolution Towards a Uniform Referencing Mode in Fact-Based Modeling

Peter Bollen(✉)

Department of Organization & Strategy, School of Business and Economics,
Maastricht University, Maastricht, Netherlands
p.bollen@maastrichtuniversity.nl

Abstract. Since its inception in the 1970's, fact-based modeling has evolved. In this article we will sketch this evolution mainly from the referencing mode's point of view. In this article we compare referencing modes from the N-ary 1989 NIAM model to the contemporary incarnation of fact-based modeling known as CogNIAM.

1 Introduction

Fact-based modeling (FBM) is a durable modeling approach that has been pioneered by Falkenberg [1, 2, 3] and Nijssen [4, 5] and in the 1970's was referred to as a conceptual modeling approach. Over the years the FBM approach has seen many incarnations and many (sometimes competing) dialects have evolved over time. Many of these dialects were characterized by being *binary* or *N-ary* (among other distinctive characteristics). The first-generation of fact-based modeling dialects, i.e. ENALIM [6, 7, 8] was N-ary and was soon converted to NIAM [9] that was restricted to *binary* fact types. The limitations of the binary dialects in terms of being able to capture 'N-ary' domain semantics has led to the introduction of multiple referencing modes [11].

In subsequent fact-based dialects additional ways of referencing objects or concepts have been introduced and some of these referencing modes have evolved over the years. In [11] these referencing modes have been evaluated with respect to the *conceptualization principle*, the principle *of semantic stability* and *ease of validation*. In this article we will follow-up on the results of [11]. In sub-section 1.1. we will present our running example that we will use in this article. In sub-section 1.2 we will show the conceptual schema for the running example using the NIAM 1989 N-ary FBM dialect. In section 2 we will summarize the results of the assessment of referencing modes in 'classic' FBM. In section 3 we will introduce a further development in referencing modes within the 'contemporary' CogNIAM incarnation of fact-based modeling. Finally in section 4 conclusions will be given.

1.1 The Running Example: University Staff Languages Spoken Abroad

We will illustrate the evolution of the referencing modeling concepts in this article by a running example (based on [10]) that we will introduce here. In this communication example we are interested in the language that university staff members have spoken abroad. We will give examples of instances of ground facts in this domain:

I. Ciuciu et al. (Eds.): OTM 2015 Workshops, LNCS 9416, pp. 226–234, 2015.
DOI: 10.1007/978-3-319-26138-6_26

```
Staff member Peter B. has spoken English in the United States
     Staff member Peter B. has spoken English in Canada
Staff member Peter B. has spoken Spanish in the United States
Staff member Thijs M. has spoken English in the United States
```

These four ground facts constitute a significant set of instances with respect to the permitted states of the fact base. The communication pattern for this example can be generalized as follows:

```
AT1: Staff member <staff member name> has spoken <language>
                        in <country>
```

In combination with the following naming convention fact types:

```
AT2: Staff member name <staff member name> identifies a
specific member of the University staff among the group of
              all University staff members

AT3: Language name <language name> identifies a specific
       language among the group of all languages

AT4: Country name <country name> identifies a specific
       country among the group of all countries
```

1.2 The N-ary 1989 NIAM Model

In figure 1 we have shown the durable model of our running example using the N-ary fact modeling conventions from the N-ary version of NIAM 1989 [12]. We will call this way of referencing 'flat', since we will implicitly use the abbreviated referencing scheme for every entity type that plays (a) role(s) in the fact type.

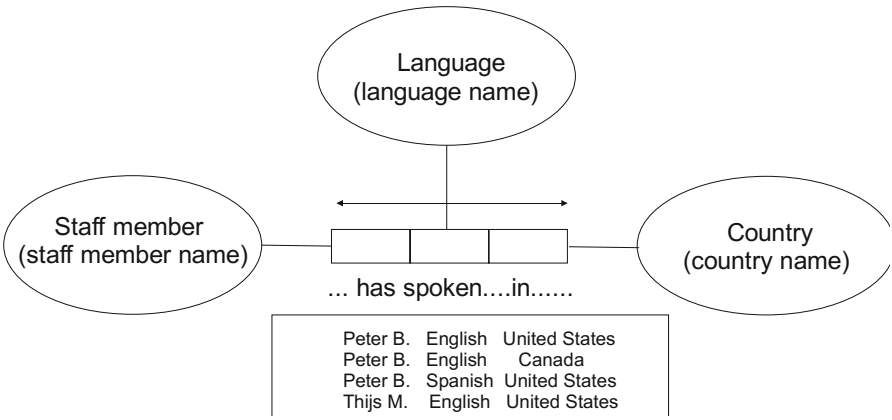

Fig. 1. Durable model of running example expressed in NIAM 1989 using flat referencing mode

Another modeling feature that was introduced in NIAM 1989 was the concept of *nesting* (or *nominalization*). An example of nominalization is the situation in which the initial user verbalization is as follows:

```
Staff member Peter B. has visited the United States.
  During that (those) visits he has spoken English.
```

Applying the conceptual schema design procedure (CSDP) on these verbalized sentence instances will lead to a 'nominalized' (or nested) NIAM 1989 durable model in figure 2.

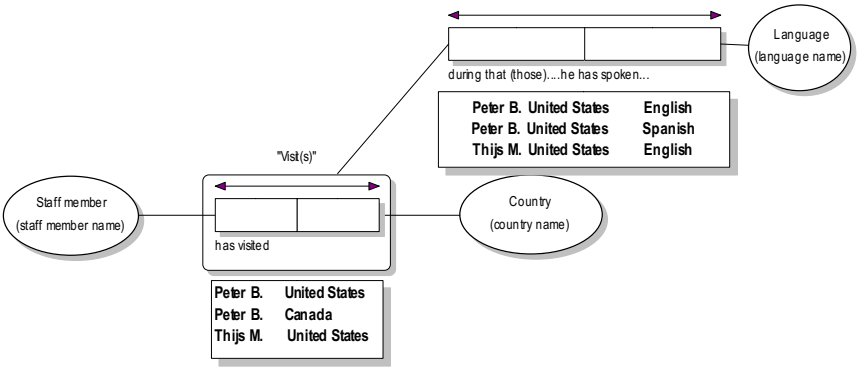

Fig. 2. Durable model expressed in NIAM 1989 using nesting (or nominalization).

We note that the application of the nesting modeling construct is user-driven, i.e. the domain user explicitly acknowledges the existence of the nested or nominalized concept (i.e. *visit*) in the domain. So the application of the nesting concept in this way is in line with the conceptualization principle. It should also be noted that the nested 'object type' is information-bearing, e.g. it can be phrased as a sentence, e.g. *Staff member Peter B. has visited the United States*.

2 Assessment of Referencing Modes in Fact-Based Modeling

In [11] we have assessed the different referencing modes that have been introduced in fact-based modeling until the FBM ORM-2 dialect (as presented in [13]). The referencing modes that were introduced and compared in [11] were: *flattening*,

Table 1. Evaluation of referencing modes (from [11])

Criterion\mode	Flat	Nesting	Co-referencing
Conceptualization principle	*Complies*	*Complies, as long as objectifications are not forced upon users that communicate in a 'flat' way*	*violates*
Semantic stability	*Complete additivity*	*nesting in retrospect might be required, when some domain experts adapt the objectified concept.*	*complete additivity*
Ease of validation	*Very good: efficient*	*good*	*not possible*

nesting and *co-referencing*[1]. These referencing modes haven been evaluated on the extent to which they comply to the *conceptualization principle*, *semantic stability* and *ease of validation* [11]. In table 1 we have summarized the main findings of [11].

From table 1 we can conclude that the 'flat' referencing mode and under certain conditions, the 'nesting' referencing mode comply to these evaluation criteria. The co-referencing referencing mode violates the conceptualization principle and does not enable the application of validation steps.

3 Further Evolution of Fact-Based Modeling

In parallel to the evolution from referencing modes from the binary 1982 NIAM to ORM-2 (2008) that was described in [11], another evolutionary path in the development of fact-based modeling can be distinguished. Developed and applied in hundreds of projects in the Netherlands the approach is now known under the name CogNIAM (an acronym for 'cognition enhanced natural language information analysis method'). In CogNIAM [14] and its pre-decessors [15], the focus is on a generic knowledge architecture that also covers the process and behavioural perspectives in durable modeling.

In this article we will illustrate the application of the fact-based approach by using CogNIAM's knowledge architecture and notational convention for semantic grammars. A theoretical foundation for CogNIAM can be found in [16, 17, 18].

In fact-based modeling we will use tangible documents or 'data-use cases' as a starting point for the modeling process. In most, if not all cases, a verbalizable knowledge source is a document that often is incomplete, informal, ambiguous, possibly redundant and possibly inconsistent. As a result of applying the fact-based knowledge extracting procedure (KEP), we will yield a document that only contains structured knowledge or a semantic grammar which structures verbalizable knowledge into the following elements (*semantic grammar* or *knowledge reference model(KRM)*)[19]:

1. Knowledge domain sentences
2. Definitions and naming conventions for concepts used in domain sentences
3. Knowledge domain fact types including sentence group templates
4. Population state (transition) constraints or validation rules for the knowledge domain
5. Derivation rules that specify *how* specific domain sentences can be derived from other domain sentences.
6. Rules that specify *what* fact instances can be inserted, updated or deleted.
7. Event rules that specify *when* a fact is derived from other facts or when a fact must be inserted , updated or deleted.

The CogNIAM knowledge extracting procedure (KEP) specifies *how* we can transform a possibly informal, mostly incomplete, mostly undetermined, possibly redundant and possibly inconsistent description of business domain knowledge into the following classes: *informal comment*, *non-verbalizable knowledge* and *verbalizable knowledge* to be classified into types 1 through 7 of the semantic grammar (or KRM).

[1] In co-referencing (compound) reference schemas are modeled as (a collection of) binary fact types that are not populatable.

We note that the knowledge extraction procedure that is needed to instantiate the elements 1 through 5 (of the KRM) is an extension of N-ary NIAM's conceptual schema design procedure (CSDP) [12]. In business domains, furthermore, we can capture the dynamic aspects by defining the exchange rules (element 6 of the KRM) and the event rules (element 7 of the KRM).

3.1 Running Example Expressed in CogNIAM

In this section we will analyze the types of reference modes that exist in CogNIAM and we will relate these referencing modes to the criteria in section 2. First we will give the CogNIAM semantic grammar for our running example. In CogNIAM's semantic grammar we first list the definitions of concepts (see table 2).

Table 2. CogNIAM's list of concept definitions of running example (element 2 of the KRM)

Country
Definition: Political division of a geographical entity, a sovereign territory, most commonly associated with the notions of state or nation and government
Language
Assumed to be known
Staff member
Definition: A Person that has a labour contract within a department of a University and who is assigned to teach and/or perform research.
Has spoken..in..
Assumed to be known

Elements 1,3,4,5,6 and 7 of the semantic grammar (or KRM) of the running example are given in figure 3

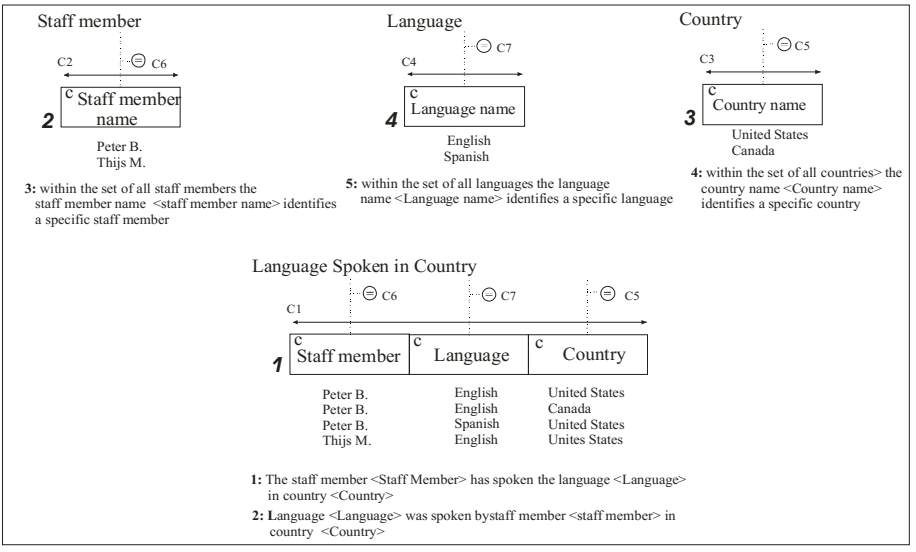

Fig. 3. CogNIAM semantic grammar of running example.

In CogNIAM there is no longer a graphical distinction into roles (or fact types) and objects. In principle CogNIAM uses elementary fact types that can have 1, 2 or N (≥3) place- holders. The reference to the object types (entity types or value types) as we find in the earlier fact-based dialects no longer takes place in a graphical way by connecting roles to entity types (and an abbreviated reference scheme) or value types, but takes place via a hierarchy of subset (or equality) constraints (e.g. see [18]) that ends in a naming convention fact type for the object type that is referenced in the place-holder. If we compare the CogNIAM information grammar from figure 3 with the equivalent 'flat' NIAM 1989 conceptual schema from figure 1 we get the mapping as is given in table 3.

Table 3. Mapping model elements from NIAM 1989 and CogNIAM

NIAM 1989 durable model (figure 1)	**CogNIAM semantic grammar(figure 3)**
Fact type predicate: *..has...spoken in..*	Fact type form 1 of fact type *1*
Entity type *Staff member*	Naming convention fact type *2*
Declaration of name class *staff member name* as identifier for entity type *staff member*	Fact type form 3 of fact type *2*
Entity type *Language*	Naming convention fact type *4*
Declaration of name class *language name* as identifier for entity type *Language*	Fact type form 5 of fact type *4*
Entity type *Country*	Naming convention fact type *3*
Declaration of name class *country name* as identifier for entity type *Country*	Fact type form 4 of fact type *3*
Uniqueness constraint spanning the fact type	Uniqueness constraint *C1*
Declaration that the entity type *Staff member* plays the left-hand role of the fact type	Equality constraint *C6*
Declaration that the entity type *Language* plays the middle role of the fact type	Equality constraint *C7*
Declaration that the entity type *Country* plays the right-hand role of the fact type	Equality constraint *C5*
(implied) Mandatory role of Entity type *Staff member* in fact type	Equality constraint *C6* (instead of subset)
(implied) Mandatory role of Entity type *Country* in fact type	Equality constraint *C7* (instead of subset)
(implied) Mandatory role of Entity type *Language* in fact type	Equality constraint *C5* (instead of subset)

We will now illustrate how an adaption of our running example will be implemented in CogNIAM. We will first of all illustrate how CogNIAM will handle an initial verbalization in a nominalized format:

```
Staff member Peter B. has visited the United States.
  During that (those) visits he has spoken English.
```

In table 4 and figure 4 we have given the CogNIAM semantic grammar for our adapted running example in which the concept of *visit* is used in the nominalized verbalization.

Table 4. CogNIAM's list of concept definitions of adapted running example

Country
Definition: Political division of a geographical entity, a sovereign territory, most commonly associated with the notions of state or nation and government
Language
Assumed to be known
Staff member
Definition: A Person that has a labour contract within a department of a University and who is assigned to teach and/or perform research.
Has spoken..in..
Assumed to be known
Visit
Definition: To reside temporarily at a place

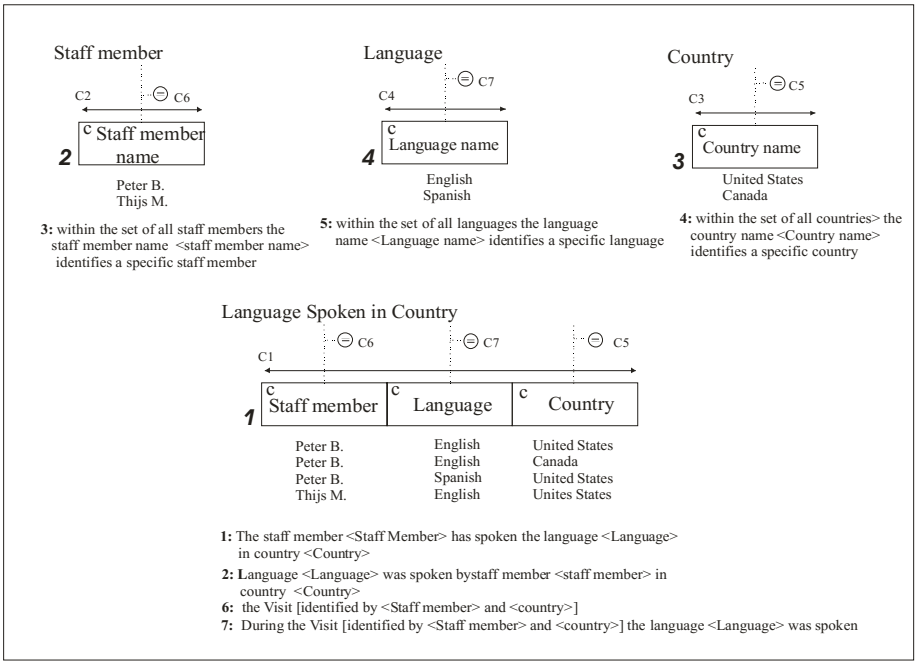

Fig. 4. CogNIAM semantic grammar of adapted running example

In fact type form 7 in figure 4 we have captured the nested or nominalized verbalization for fact type 1. In fact type form 6 in the CogNIAM semantic grammar in figure 4 we see that a nominalized verbalization can be transformed into an object-verbalization [18]:

```
The visit [identified by <Staff member> and <Country>]
```

We can easily add a fact type form that contains such an objectified verbalization for a target group that uses this verbalization in their communication. We note that by using the objectification symbols [ ], CogNIAM captures the name of the objectification while allowing to use non-objectified communication patterns for different target groups.

With respect to the explicit declaration of naming convention fact types in CogNIAM (i.e. fact types 2,3 and 4 in the semantic grammar of figure 4) we notice that the fact type forms explicitly capture a description of the 'name-space' within which a name can identify an instance of a concept (or object). We note that these explicit referencing semantics could not be captured in earlier fact-based dialects (e.g. NIAM 1989).

4 Conclusion

In this paper we have summarized the results of a comparison of the three most important referencing formats in 'classic' Fact-Based Modeling: *flat*, *nested* and *co-referencing*. In the second part of the paper the evolution of the referencing modes into CogNIAM have been demonstrated. Being able to define multiple fact type forms for any fact type in CogNIAM allows us to incorporate nesting (or nominalization) and the flat referencing mode into the same uniform fact type structure in FBM and at the same time allow different stakeholder communities to use their preferred 'local' way of referencing in a 'global' model. With respect to the overarching objective of this article we can conclude that in CogNIAM on a diagrammatic level there is only one way to reference objects. However, CogNIAM allows for adding multiple fact type forms to one diagram, thereby making it possible to record nominalized and flat fact type verbalizations at the same time for a given fact type. A major advantage of the CogNIAM naming convention fact types is the ability to capture 'name space' semantics in an explicit way.

References

1. Falkenberg, E.: Strukturierung und Darstellung von Information an der Schnitstelle zwischen Datenbankbenutzer und Datenbank-Management-System, Ph.D thesis, University of Stuttgart (1975)
2. Falkenberg, E.: Significations: the key to unify data base management. Information Systems **2**, 19–28 (1976)
3. Falkenberg, E.: Concepts for modelling information. In: Modelling in Database Management Systems (1976)
4. Nijssen, G.: A brief introduction to common data base languages DDL-DML. Codasyl Base Task Group Report (1972)
5. Nijssen, G.: Present and future possibilities of database technology. In: Medinfo, pp. 1−4 (1974)
6. Nijssen, G.: On the gross architecture for the next generation database management systems. In: Information processing, IFIP 1977 (1977)

7. Nijssen, G.: A gross architecture for the next generation database management systems. In: Nijssen, G. (Ed.) Modelling in Database Management Systems, North-Holland, pp. 1−24 (1976)
8. Nijssen, G.: A framework for discussion in ISO/TC97/SC5/WG3 and comments on 78.04/01 and 78.05/03 (1978)
9. Verheijen, G., van Bekkum, J.: NIAM: an information analysis method. In: IFIP TC-8 CRIS-I conference, North-Holland, Amsterdam (1982)
10. Nijssen, G.: The first increment of the SIMF proposed by the SIMF EU response group (2012)
11. Bollen, P.: referencing modes and ease of validation within fact-based modeling. In: Meersman, R., et al. (eds.) OTM 2014. LNCS, vol. 8842, pp. 289–298. Springer, Heidelberg (2014)
12. Nijssen, G., Halpin, T.: Conceptual schema and relational database design: a fact oriented approach, New-York: Prentice-hall. 337 (1989)
13. Halpin, T., Morgan, T.: Information Modeling and Relational Databases; from conceptual analysis to logical design 2nd edn., San-Francisco, California. Morgan-Kaufman (2008)
14. Nijssen, G., Le Cat, A.: Kennis gebaseerd werken: de manier om kennis productief te maken. PNA Publishing, Heerlen (2009). (in dutch)
15. Lemmens, I.M., Nijssen, M., Nijssen, S.: A NIAM 2007 conceptual analysis of the iso and omg mof four layer metadata architectures. In: Meersman, R., Tari, Z. (eds.) OTM-WS 2007, Part I. LNCS, vol. 4805, pp. 613–623. Springer, Heidelberg (2007)
16. Bollen, P.: On the applicability of requirements determination methods, in Management and Organization, University of Groningen, Groningen, p. 219 (2004)
17. Bollen, P.: The natural language modeling procedure. In: Halevy, A.Y., Gal, A. (eds.) NGITS 2002. LNCS, vol. 2382, p. 123. Springer, Heidelberg (2002)
18. Bollen, P.: Natural Language Modeling for Business Application Semantics. Journal of Information Science and Technology **2**(3), 18–48 (2005)
19. Nijssen, M., Lemmens, I.M.: Verbalization for business rules and two flavors of verbalization for fact examples. In: Meersman, R., Tari, Z., Herrero, P. (eds.) OTM-WS 2008. LNCS, vol. 5333, pp. 760–769. Springer, Heidelberg (2008)