The following full text is a publisher's version.

For additional information about this publication click this link.
http://hdl.handle.net/2066/28418

Please be advised that this information was generated on 2017-12-23 and may be subject to change.
1. INTRODUCTION

Many conceptual modelling methods have as point of departure an initial specification of the world to be modelled, also referred to as the Universe of Discourse (UoD) [1]. This initial specification is usually expressed in natural language. The goal of the modelling process is to obtain a complete and consistent formal specification from the initial specification. Only a few conceptual modelling methods provide the system analyst with clues and rules which can be applied to the initial specification such that this specification can actually be used in the modelling process. Examples of these modelling methods are NIAM [2], [3] for conceptual data modelling and KISS [4] for object-oriented modelling.

In order to structure this way of working, the (sentences in the) initial specification should be in accordance with some unifying format. In NIAM these sentences are called elementary sentences, while KISS uses the term structure sentences. Elementary and structure sentences provide a simple and effective handle for obtaining the underlying conceptual model of so-called Snapshot Information Systems [5], i.e. information systems where only the current state of the UoD is relevant.

However, even though these initial specifications are an important aid in modelling information systems, they still are too poorly structured. This lack of structure results in several problems during the modelling process. As an example, a system analyst has to come up with a lot of properties of the UoD, which do not follow directly from the initial specification, but have to be specified in order to satisfy the 100% principle for the formal specification. One of these properties is the order and history of events. More concretely, since the mutual order of the sentences in an initial specification is lost, the analyst has to reconstruct this order. Other UoD properties are for instance related to the instances involved in events and the role in which they are involved. This combination, together with properties about the specific involvement, is referred to as an associate of the event. Usually, associates are not clear from the informal specification.

In this paper the notion of logbook is introduced as a common basis for various models to be produced during system analysis. A logbook has a unifying format which contains a complete description of the history of some Universe of Discourse. It is intended as a structuring mechanism for initial specifications based on natural language. A typing mechanism is provided as an abstraction mechanism for logbook instances, leading to object-oriented analysis models.

The logbook is intended as a structuring mechanism for initial specifications in the context of Snapshot Information Systems as well as Temporal Information Systems [6]. Furthermore, it supports the development of Evolving Information Systems (e.g. [7], [8], [9]). Note that Evolving Information Systems can be used to handle versioning problems.

The logbook is intended as a structuring mechanism for initial specifications in the context of Snapshot Information Systems as well as Temporal Information Systems [6]. Furthermore, it supports the development of Evolving Information Systems (e.g. [7], [8], [9]). Note that Evolving Information Systems can be used to handle versioning problems.

The main reason for introducing the concept of logbook is that it can be used effectively to bridge the gap between a UoD and its specification. The history of a UoD may be described by an application model history, such as that introduced in [7]. In this view, the evolution of a UoD is seen as the evolution of its elements. An element evolution describes the state of the element at each point of time. This dual vision on evolving information systems will not be elaborated further in this paper.

The organization of the paper is as follows. In Section 2 logbooks are formally introduced. We present a meta model
in terms of PSM [10], followed by a further elaboration of
logbook functions and properties. These properties include
the so-called creation (birth) and destruction (death) of
objects. Section 3 describes the modelling process, i.e. the
process of obtaining a conceptual model from a logbook.
This is centered on the definition of a typing mechanism.
Our approach is further illustrated by relating different
views on a logbook to concrete models which constitute the
overall conceptual model. We consider the object action
involvement model, the object life model and the object
property model. Finally, conclusions and directions for
further research are presented in Section 4.

2. LOGBOOKS
The most simple description of events occurring in a UoD
consists of the following components for each event: (1)
when does it occur, (2) what is happening, (3) who are
associated, (4) what is the role of each associate and (5)
which properties of associates are relevant. A logbook is a
set of such event descriptions, ordered by their time-stamps
in some unifying format (seconds, hours, days, etc.).

2.1. Informal specifications
The starting point for constructing a logbook is an informal
specification consisting of natural language sentences
describing time-stamped events. Obtaining an informal
specification is usually an incremental and iterative process
[11], [12] which requires skills for both the domain expert
and system analyst [13]. As an example of an informal
specification consider Table 1, where the unifying time
format is chosen to be on a daily basis.

2.2. Meta-model of a logbook
Formally, the format of a logbook is defined by the meta
model of Figure 1. This meta model is presented as a PSM
schema which is a formalized and extended version of the
object role modelling technique NIAM. For more literature
about the background and formalization of PSM, the reader
is referred to [10] or [14].

A PSM schema is a structure consisting of an information
structure and a set of constraints. The semantics of a
PSM schema is the set of its possible populations, which
is restricted by both the information structure and the
constraints.

The information structure consists of the following basic
components:

1. A set of predications. A predication is intended to specify
the role played by an object type in a fact type (see
below).
2. A set of object types which can be classified as follows:
   (a) Entity types and label types. The difference is that
labels can, in contrast to entities, be represented
(reproduced) on a communication medium.
   (b) Fact types. The set of fact types is a partition of
the set of predications.

FIGURE 1. Meta model for the format of a logbook.

(c) Complex object types such as power types (also
called set types), sequence types (for modelling
list structures) and schema types (which can be
used as an abstraction mechanism for information
structures).

3. Relations for specifying (a) specialization and (b)
   generalization.
4. Functions for specifying (a) the object type associated
   to a predication, (b) the fact type of an associated
   predication and (c) the underlying element type (or
   object types) for power types and sequence types (and
   schema types).

In Figure 1 we have a number of entity types, e.g. Event,
represented by circles. Label types, also represented by
circles, appear between parentheses. Examples of label
types are Name and Stamp. The convention is used that
if a label type is an identifier for an entity type, the label
type is represented within the same circle. For example,
in the logbook meta-model, an Action is identified by its
Name. Fact types consist of predications represented by
boxes connected to circles for their associated object types.
For example fact type t is a binary fact type consisting of
two predications, with associated object types Event and
Time. Object type Properties is modelled as a group type
with underlying element type Property. A Logbook is seen
as a sequence type of tuples consisting of an Event and Time
part. Finally, the information structure of a logbook contains
a generalization hierarchy. Object type Involved is either an
Instance or an Event/Time tuple.
For example each handle the case of multiple agents on a conceptual level. As a consequence, the modelling technique must be able to be elementary and thus cannot be split into smaller events.

In order to be able to represent logbooks in a way such as that presented in Table 1, some additional (non-graphical) constraints are required. The PSM modelling technique is equipped with the query language Lisa-D to express such constraints. Lisa-D as such falls outside the scope of this paper, for more information see [15]. The additional constraints are informally expressed as:

1. the order of the tuples of a logbook is according to the order of its time stamps and
2. if two events occur at the same time, these events cannot be the same.

Table 2 represents a population of the meta model of the logbook in Figure 1 and thus is a unifying format for the sample logbook of Table 1. Note that the action to write may have multiple agents. However, each of these events is elementary and thus cannot be split into smaller events. As a consequence, the modelling technique must be able to handle the case of multiple agents on a conceptual level.

A population of the logbook meta-model can be obtained by grammatical analysis of the sentences of the informal specification. Methods such as KISS and SACIS (as described in [16]) provide the analyst with clues for grammatical analysis of informal specifications. Grammatical analysis as such falls outside the scope of this paper.

### 2.3. Logbook functions

In this section a number of functions are defined in order to elaborate the notion of a logbook as defined by its meta model. The function \( \tau(e) \) is such a function which produces the point of time of an event \( e \). The function \( \alpha(e) \) results in the performed action and \( \Delta(e) \) is the mapping which assigns the set of associates occurring in event \( e \).

Let \( a \) be an associate of event \( e \), i.e. \( a \in \Delta(e) \). The function \( \iota(a) \) yields the involved object, \( \pi(a) \) assigns the (possibly empty) set of properties of this object, whereas \( \rho(a) \) provides its role in this event. Note that a role may be played more than once in an event.

An event \( e \) is thus identified by the action \( \alpha(e) \) performed, and the set \( \Delta(e) \) of associates. An associate \( a \) is identified by its involved object \( \iota(a) \), its corresponding set of properties \( \pi(a) \) and its role \( \rho(a) \).

We introduce for given functions \( f : B \rightarrow C \) and \( g : A \rightarrow P(B) \) the composition operator * as:

\[
(f \ast g)(a) = \{ f(b) \mid b \in g(a) \}
\]

**EXAMPLE 1.** Consider the event \( e_1 \) with time stamp (23-06-1967) from Table 2:

\[
\{ \text{is produced}, \{ (e_2, \text{object}, \{ \text{tape nr 666, studio nr 11} \}), \}
\]

\[
\{ (\text{Mick Jagger, agent, 0}), (\text{Keith Richards, agent, 0}) \} \}
\]

and the event \( e_2 \) with time stamp (21-06-1967):

\[
\{ \text{is recorded}, \{ (\text{Paint It Black, object, 0}), \}
\]

\[
\{ (\text{The Rolling Stones, agent, 0}) \} \}
\]

which correspond with the following sentences of the informal specification:

*The recording of Paint It Black, on tape nr 666 in studio nr 11, is produced by Mick Jagger and Keith Richards.*

*Paint It Black is recorded, on tape nr 666 in studio nr 11, by The Rolling Stones.*

Applying the functions \( \tau, \alpha, \Delta, \iota, \rho \) and \( \pi \) to events \( e_1 \) and \( e_2 \) results in:

\[
\tau(e_1) = 23-06-1967
\]

\[
\alpha(e_1) = \text{is produced}
\]

\[
\Delta(e_1) = \{ a_1, a_2, a_3 \}
\]
TABLE 2. A sample population of Figure 1

<table>
<thead>
<tr>
<th>Time</th>
<th>Action</th>
<th>Involved</th>
<th>Associate</th>
<th>Role</th>
<th>Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>01-05-1963</td>
<td>set up</td>
<td>Mick Jagger</td>
<td>agent</td>
<td></td>
<td>µ</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Keith Richards</td>
<td>agent</td>
<td></td>
<td>µ</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The Rolling Stones</td>
<td>object</td>
<td></td>
<td>µ</td>
</tr>
<tr>
<td>03-05-1967</td>
<td>write</td>
<td>Mick Jagger</td>
<td>agent</td>
<td></td>
<td>µ</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Keith Richards</td>
<td>agent</td>
<td></td>
<td>µ</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Paint It Black</td>
<td>object</td>
<td></td>
<td>µ</td>
</tr>
<tr>
<td>21-06-1967</td>
<td>record</td>
<td>Paint It Black</td>
<td>object</td>
<td></td>
<td>µ</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The Rolling Stones</td>
<td>agent</td>
<td></td>
<td>µ</td>
</tr>
<tr>
<td>23-06-1967</td>
<td>produce</td>
<td>{21-06-1967, record, (Paint It Black, object, µ), (The Rolling Stones, agent, µ)}</td>
<td>object</td>
<td>{tape nr 666, studio nr 11}</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mick Jagger</td>
<td>agent</td>
<td></td>
<td>µ</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Keith Richards</td>
<td>agent</td>
<td></td>
<td>µ</td>
</tr>
<tr>
<td>12-12-1989</td>
<td>write</td>
<td>I Want You</td>
<td>object</td>
<td></td>
<td>µ</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A. Knijff</td>
<td>agent</td>
<td></td>
<td>µ</td>
</tr>
<tr>
<td>10-02-1989</td>
<td>set up</td>
<td>P. Frederiks</td>
<td>agent</td>
<td></td>
<td>µ</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A. Knijff</td>
<td>agent</td>
<td></td>
<td>µ</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The Playful Plebs</td>
<td>object</td>
<td></td>
<td>µ</td>
</tr>
<tr>
<td>29-04-1991</td>
<td>record</td>
<td>The Playful Plebs</td>
<td>agent</td>
<td></td>
<td>µ</td>
</tr>
<tr>
<td></td>
<td></td>
<td>I Want You</td>
<td>object</td>
<td></td>
<td>µ</td>
</tr>
<tr>
<td>05-05-1991</td>
<td>produce</td>
<td>H. Honer</td>
<td>agent</td>
<td></td>
<td>µ</td>
</tr>
<tr>
<td></td>
<td></td>
<td>{29-04-1991, record, (I Want You, object, µ), (The Playful Plebs, agent, µ)}</td>
<td>object</td>
<td>{tape nr 3, studio nr 2}</td>
<td></td>
</tr>
</tbody>
</table>

2.4. Properties of a logbook

In this approach, a concrete logbook is an instantiation of the meta model from Figure 1, i.e. a mapping which instantiates each object type. A population thus assigns a set of values to each object type from the meta model. Let $\text{Log}$ be such a mapping (population). Applying function application leads simply to some results. For example the set of all involved objects occurring in this logbook can be obtained with $\text{Log}(\text{Involved})$. This set corresponds to what is called the extra temporal population in [7].

The logbook can be restricted to the history (evolution) $\text{Log}_x$ of a single involved object $x$ by removing all events from the logbook in which $x$ is not involved. This population of the meta model is obtained by first determining the relevant events:

$$\text{Log}_x(\tau) = \{(t, e) \in \tau \mid x \in (\iota \ast \Delta)(e)\}$$

where, for convenience, $(t, e) \in \tau$ is used as a shorthand for $(t, e) \in \text{Log}(\tau)$ (thus overloading the $\in$-operator). The instantiations of the other object types of the meta model are obtained as the minimal subset of their original instantiation required for $\text{Log}_x(\tau)$.

Restricting a logbook $\text{Log}$ to a particular point of time (time stamp) $t$ results analogously in a snapshot of the logbook:

$$\text{Log}(\tau) = \{(t, e) \in \tau \mid \tau(e) = t\}$$
Note that Log^ (Involved) results in the set of objects describing the life of objects. The creation (birth) of an object has an especially prominent position in these life models. The creation Inv(x) of an object x is defined as the first event in the logbook which mentions this object:

\[ \text{Init}(x) \in \text{Log}_x(\text{Event}) \land \forall_{e:\text{Log}_x(\text{Event})} [\tau(\text{Init}(x)) \leq \tau(e)] \]

As a result Inv(Init(x)) denotes the action which causes the birth of object x. Note that the death of an object cannot be derived from the logbook since objects are not removed from the logbook. Usually there will be a criterion Alive(x,t) which decides whether object x is alive at some point of time t. The following property then is obvious:

\[ \text{Alive}(x,t) \land \forall_{x',t'} [\sim \text{Alive}(x,t') \Rightarrow \tau(\text{Init}(x)) = t] \]

The death of an object is not derivable, however, from the Alive predicate but the death of an object may be concluded. Suppose object x is alive at time t, but not alive at a later point of time t', then x has died somewhere in between. The set Log(Role) of all roles has a central position during the modelling process. In the case of evolving information systems, this set will also be subject to evolution. In this paper, we restrict ourselves to a stable set of roles and thus to the specification of temporal information systems.

3. MODELLING A LOGBOOK

A number of conceptual data modelling techniques are based on using natural language. The goal of such modelling techniques is to derive the grammar that governs the communication within the UoD. Some examples of such techniques are NIAM and PSM. This grammar, also referred to as information grammar (see e.g. [17], [18], [19], [20]), can be depicted as an information structure diagram. Information grammars achieved with these techniques only describe static aspects of the communication in the UoD. In order to find an information grammar which also includes dynamic parts of the UoD, the modelling process may start from an informal specification such as presented in Table 1.

Once a stable sample logbook is obtained from this informal specification the main task of the system analyst is to find abstractions for the concrete instances within the sample logbook, leading to a structural description (conceptual model). The information grammar can also be used to paraphrase the conceptual model, resulting in a textual description of the UoD. A major advantage of this approach is that the domain expert can validate this textual description [21].

In the remainder of this section we will first informally introduce a number of analysis models for object-orientation which compose the conceptual model. Each analysis model describes a specific view on the logbook. For an in-depth treatment of the analysis models see [22]. Then we will introduce a typing mechanism which describes the modelling process, i.e. a way to obtain the analysis models from the sample logbook.

3.1. Views on a logbook

During the several stages of the modelling process the sentences of a logbook are analysed according to three perspectives, leading to a number of abstractions of the logbook, also referred to as analysis models. Each abstraction provides a specific view on the nature of the application domain and results in a corresponding model. The models together compose a conceptual model of the UoD. This conceptual model is also referred to as the information architecture. The information architecture composes the following analysis models:

- object action involvement model,
- object life model and
- object property model.

The purpose of the object action involvement model is to develop a first approximation of the information grammar, covering the main structure of the logbook language. This model provides an abstraction (type) for three columns of the logbook, i.e. Action, Involved and Role (see Table 2). What is special is that different types of events may have the same type of action. Eventually this can lead to the introduction of generalized object types. Subtyping hierarchies as such falls outside the scope of this paper.

The object action involvement model does not restrict the order of the action types. The object life model elaborates on the object action involvement model. It (the object life model) considers the application domain from a historical perspective and describes for each object type the course of life of its instances. Each object type starts with its birth action type (creation). From the object action involvement model one can derive in which action types an object type is involved. Using the column Time of the logbook, clues for the order of action types can be derived. Note that the object life model describes a possible order in which action types can be performed for each object type. By collecting all life courses the (the restriction on the) dynamics of the UoD can be derived.

The object property model deals with such static aspects of the application domain by modelling properties of object types. The properties of an object type form a so-called state record which reports the current state of an object. Usually the property is set during the birth of the associated object type. Properties are only useful if they can be retrieved via retrieval action types and updated by so-called update action types. Strictly speaking, properties thus can also be modelled via their initialization action type, retrieval action types and update action types. The object property model is introduced to provide a more simple description mechanism dealing with properties, obtained by consulting the Property column of the logbook.
3.2. Typing a logbook

In order to obtain structural information from a logbook, all instances that occur in the sentences (tuples) of the logbook have to be typed according to some typing mechanism. This typing mechanism is guided by the three analysis models that constitute the information grammar. The modelling process may be informally described as follows: Find a consistent typing mechanism, i.e., a typing mechanism satisfying:

1. Events with a similar action have (a) similar objects involved (b) in similar roles.
2. Similar objects have a similar course of life.
3. Similar objects have a similar state record.

Two instances of the logbook meta-model are said to be similar if they have the same type assigned. Two sets of instances are called similar if each instance in the first set has a similar instance in the second set and vice versa. Recognizing similar lives of objects is hard to formalize. In fact this remains the main task of the system analyst: ordering and comparing the course of life of objects. In any case, two objects are candidates for similar lives if they have a similar birth action.

Based on this intuition the modelling process can be formally stated as follows: find a set $T$ of types and a typing function

$$\text{Type} : \begin{cases} \text{Log(Involved)} \cup \\ \text{Log(Action)} \cup \\ \text{Log(Role)} \cup \\ \text{Log(Property)} \end{cases} \rightarrow T$$

such that:

1. Events $e_1$ and $e_2$ with a similar action have (a) similar objects involved (b) in similar roles:

   $$\text{IsSim}(\alpha(e_1), \alpha(e_2)) \rightarrow \text{lsSim}(t \ast \Delta(e_1), t \ast \Delta(e_2))$$

2. Similar objects $x$ and $y$ have a similar course of life:

   $$\text{lsSim}(x, y) \Rightarrow \text{SimLife}(x, y)$$

3. Similar objects $x$ and $y$ have a similar state record:

   $$\text{lsSim}(x, y) \Rightarrow \text{lsSim}(\text{Log}_x(\text{Properties}), \text{Log}_y(\text{Properties}))$$

The similarity predicate is defined on instances $x$ and $y$ as:

$$\text{IsSim}(x, y) \leftrightarrow \text{Type}(x) = \text{Type}(y)$$

whereas on sets $X$ and $Y$ of instances:

$$\forall x \in X \exists y \in Y \left[ \text{IsSim}(x, y) \right] \wedge \forall y \in Y \exists x \in X \left[ \text{IsSim}(y, x) \right]$$

Note that this definition does not require similar sets of instances to have equal cardinality. Similarity of life of objects is defined as:

$$\text{SimLife}(x, y) \leftrightarrow \text{IsSim}(\alpha(\text{Init}(x)), \alpha(\text{Init}(y)))$$

The similarity predicate forms an equivalence relation over the instances of the logbook.

This typing mechanism is very strict. The typing mechanism can be enhanced, using the type relatedness predicate (see e.g. [22]), which extends the typing mechanism with a subtyping mechanism (specialization and generalization). For example, in some UoD regarding music companies, different object types, such as Person and Musician, can share common instances. A Musician might be modelled in this UoD as a specialization of a Person. In this UoD Person and Musician are type-related but not of similar type. Introduction of a subtyping mechanism requires a refinement of the typing mechanism. For example, similar objects within the same subtype hierarchy need not have a similar state record, e.g., object type Musician can have properties which are not relevant for object type Person.

**EXAMPLE 2.** Suppose the typing mechanism contains the following:

- Type(Mick Jagger) = Person
- Type(The Rolling Stones) = Person

Then we have no consistent typing, as Mick Jagger and The Rolling Stones do not have a similar life.

**EXAMPLE 3.** Suppose the typing mechanism contains the following:

- Type(Mick Jagger) = Person 1
- Type(Keith Richards) = Person 2

This typing is not plausible, as Mick Jagger and Keith Richards occur in events with similar actions in similar roles.

4. CONCLUSIONS

In this paper the notion of logbook was introduced as a common basis for various information system models. A logbook is intended as a structuring mechanism for initial specifications, which is obtained via a unifying format containing a complete description of the history of some UoD. A typing mechanism is provided as an abstraction mechanism for logbook instances, leading to object-oriented analysis models.

Further research may address the completeness of this approach (for more information, see [23]). For example, does the logbook contain all information relevant for the modelling process? Another direction could be to make a prototype to support the approach of this paper.

ACKNOWLEDGEMENTS

We wish to thank Caspar Derksen and Kees Koster for their interest and advice. We also wish to thank the anonymous referees for their valuable comments which led to an improvement of the paper.
REFERENCES


