

# The SDA\* Data Model v1.0

D. Riaño, J. A. Bohada, F. Real, J. A. López-Vallverdú

*Research Group on Artificial Intelligence*

Rovira i Virgili University

banzai-deim.urv.net

david.riano@urv.net

The data related to health-care procedures like diagnosing or treating a patient use to be structured in different ways depending on the health-care institution where the procedure is carried out. On the contrary, the medical and assistive principles hidden behind the data use to be the same. So, for example the treatment of a chronic ailment is used to be structured as a sequence of encounters between the patient and the physician(s) through time, that reflect a cycle with steps admission, follow-up, re-evaluation, and discharge.

This document describes a general purpose data model to store the data in health-care procedures. It is not intended to be a standard or a complete model able to deal with all the health-care situations that may occur in real life, but a departing model to allow medical informatics systems and artificial intelligence technologies (e.g., inducing SDA\* knowledge structures [1]) to exploit health-care procedural data. This is the *SDA\* data model*.

The departing points of this model are: a particular way of deploying health-care assistance, and an ontology of the terms involved in this assistance.

**Health-Care assistance:** the treatment of a patient is ruled through an episode of care. We take the meaning of this concept as the one published in [2]: *"The term episode of care refers to a health problem from its first encounter with a health care provider through the completion of the last encounter related to that problem [3]. An episode of care, therefore, differs from an episode of disease, which is a health problem from its onset through its resolution or until the patient's death, and an episode of illness, which is the period during which a person suffers from symptoms or complaints experienced as an illness. Not every disease and certainly not every illness results in an episode of care. Most episodes of care, however, are part of an episode of disease and, less often, of illness. Health maintenance episodes can be considered a special form of episodes of care"*.

Episodes of care (EOC) are then sequences of patient-professional encounters, in which the professional may observe the patient condition and decide some action measures either to provide the sort of care the patient requires or to ascertain some facts that are relevant to the treatment of the patient. Optionally, these action measures have some reasons that justify their application (i.e., evidence-based health-care). In practice, real encounters may be related to more than one episode if the professional uses the same encounter to treat the patient of several co-morbid health problems. In these cases, a reference to the sort of health problem each action measure in the encounter is addressed to is required to make it compatible with the definition of EOC of the SDA\* data model.

Time is also an important component of the EOC. Each encounter has a date and a day time, and some action measures have a period of application (start time, duration, and frequency). For example, take antibiotics starting this afternoon, during 1 week, and every 6 hours. Medical observations may also have a time dimension, as for example "has the patient been taking antibiotics during the last week?". Patient conditions and decisions may also be conditioned to time.

**The assistance ontology:** when stored, all the data of the EOC must be expressed in some specialized vocabulary in which each term has a concrete and, hopefully, free of vagueness meaning. Codifying terms (or expressions) is also a common practice in health-care. These terms are called the *SDA\* vocabulary terms*. In the EOC described above, the following vocabularies are required: state vocabulary (terms to describe patient conditions, e.g. signs and symptoms, etc.), action vocabulary (terms to describe the action measures of the health-care providers, e.g. medical procedures, medication, counsels, etc.), and decision vocabulary (terms used to justify the application of the action measures).

SDA\* vocabulary terms comprise standard codifications [4, 5, 6] as ATC, IDC9, ICD10-CM, CIAP, etc. as well as other specific terminology that the users of the SDA\* model may get the permit to define. All this vocabulary is structured in the assistance ontology which is the hierarchy of classes of Figure 1. Each term of the vocabulary is an instance of one or several of such ontology classes which are defined not to be disjointed.

The classes in the SDA\* structure layer of the ontology cannot be extended. The SDA\* health-care layer can be extended with new subclasses of STATE, DECISION, and ACTION, or with subclasses of the other classes in this layer. The SDA\* code layer include either standard or *ad hoc* medical codifications of terms.

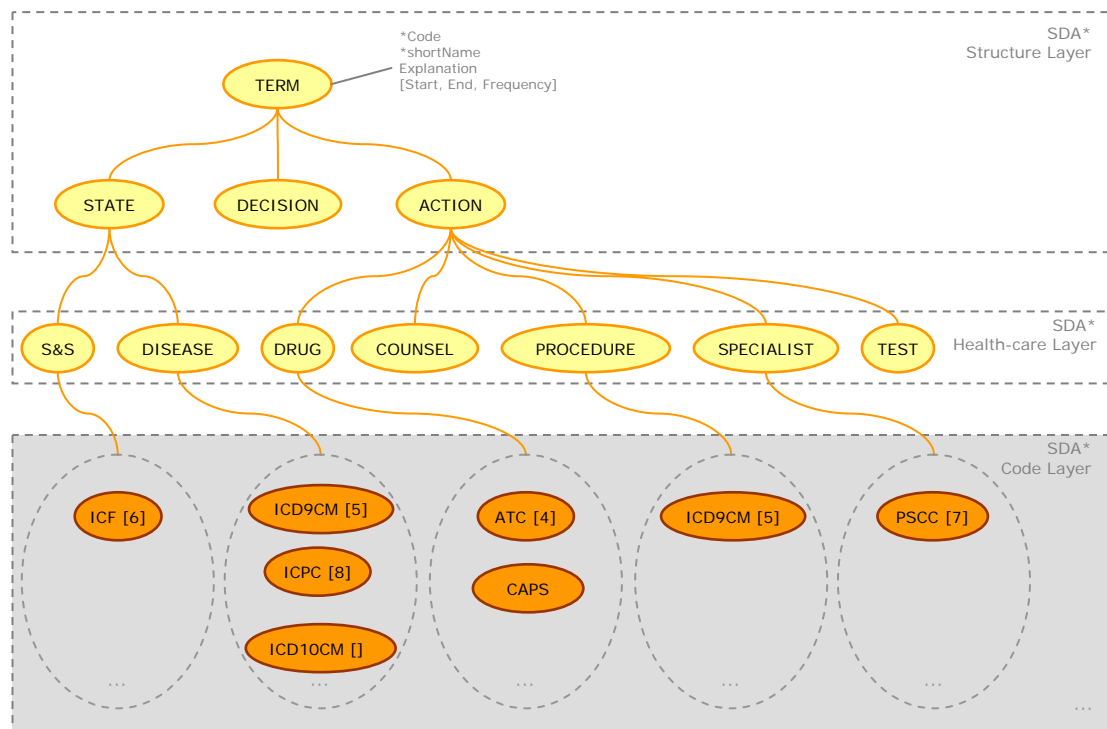


Figure 1. SDA\* term classes.

Time has a particular treatment in the SDA\* Data Model. This model is able to manage *absolute times* as dates, day times or frequencies, and *relative times* as durations. Relative times can be converted to absolute times by putting them in the context of an absolute time; for example, when the relative time TOMORROW is employed in January 15<sup>th</sup> 2008, it refers to the absolute day time January 16<sup>th</sup> 2008 or when a relative duration of 5 days is indicated on January 15<sup>th</sup> 2008, this defines an absolute time going from that absolute date to January 20<sup>th</sup> 2008.

Time can be expressed as a term (e.g. TODAY, YESTERDAY or TOMORROW for dates; NOW for day time, and SID or QD, TID, QID for 1, 3, or 4 times a day, QOD for every other day, Q6H for every six hours, etc. as frequencies) or as an expression. On the one hand, SDA\* time terms are instances of the TIME class and subclasses in Figure 2.

On the other hand, SDA\* time expressions define times by means of a particular syntax: dates are expressed as DD:MM:YYYY, times as HH:mm:ss, and frequencies as NA where DD represent the day of the month, MM is the number of month, YYYY the year (e.g., 15:12:1968 is December 15<sup>th</sup>, 1968), HH is the day hour in the range 00-23, mm is the minutes in the range 00-59, SS the seconds in the range 00-59 (e.g., 16:12:30 is twelve and

a half minutes past 4pm), N is a natural number, and A one of the letters in Table 1 (e.g., 15m represents the frequency once every fifteen minutes).

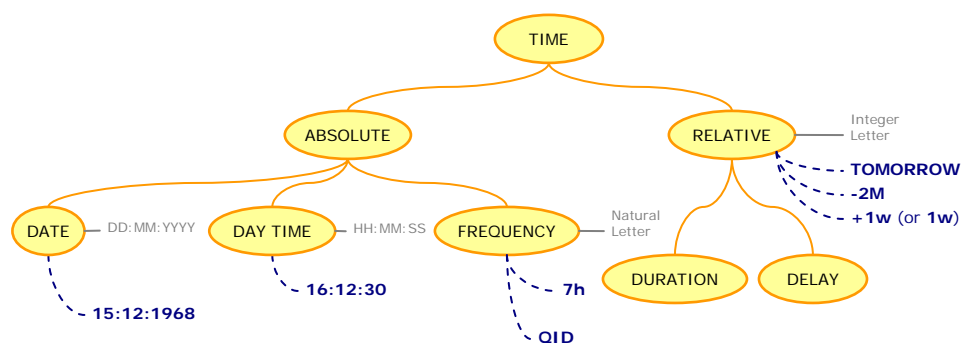


Figure 2. SDA\* time classes.

A	Meaning	Conversion
y	years	
M	months	(12M = 1y)
w	weeks	(4w = 1M)
d	days	(7d = 1w)
h	hours	(24h = 1d)
m	minutes	(60m = 1h)
s	seconds	(60s = 1m)

Table 1. Codification of relative times in the SDA\* Data Model.

Relative times are of the form NA where N is an integer number and A is one of the letters in Table 1. For example, +7h (or 7h) means in seven hours, -10m means ten minutes ago, -3M means three months ago, etc. Notice that in the SDA\* data model, this sort of time is always relative to the absolute time of an encounter. Other sorts of temporal constraints as for example “one month after the last intake of drugs” are not possible in the SDA\* model<sup>1</sup> and they have to be converted to a proper relative time.

Relative times are employed to define the temporality of the SDA\* vocabulary terms by means of micro-temporality constraints. A *micro-temporality constraint* is a time restriction of the sort [S, E, F] on a term. If it is a state or decision term it means that the term is valid since S time until E time (S time previous to E time) with a frequency F, with S, E, and F relative times (F with a positive sign). If the term is an action term, it means that the action must be applied between S and E (S time previous to E time), with a frequency F. For example, micro-temporality [-2M, -1w, QID] attached to the intake of some drug D (i.e., DRUG term D) means that the patient has been taking D since two months ago until last week, four times a day. On the contrary, micro-temporality [1w, +2M, 12h] on a DRUG term D means that the patient has to take D starting in one week, for two months from now, and twice a day. In case you want to make explicit that the patient is taking D at some particular time t once a day, there should be a SDA\* term D-t with micro-temporality [1w, +2M, 1d]. If the patient has to take it also during some time a in the afternoon a new term D-a should be introduced.

**The SDA\* Data Model v1.0:** Figure 3 summarizes the proposed SDA\* data model. This model is based on both, the EOC as it is previously defined, and also the SDA\* vocabulary. Formally speaking, the SDA\* data model is a pair (G, O) such that G is a formal context-free grammar representing the EOC in the form:

EOC := EOC-ID PAT-ID ENCOUNTER\*  
 ENCOUNTER := E-DATE E-TIME { S } HC-MEASURE\*  
 HC-MEASURE := { D } { A }

<sup>1</sup> Observe that this is not a limitation of the SDA\* data model, but a medical imprecision that has to be solved either indicating the exact time in which drug intake must stop (e.g., take drugs for two weeks) or leaving the decision for a future event (e.g., take drugs indefinitely and schedule a new visit next week when drug intake will be reconsidered).

... and O the SDA\* vocabulary structured as the SDA\* ontology in Figure 1 and Figure 2.

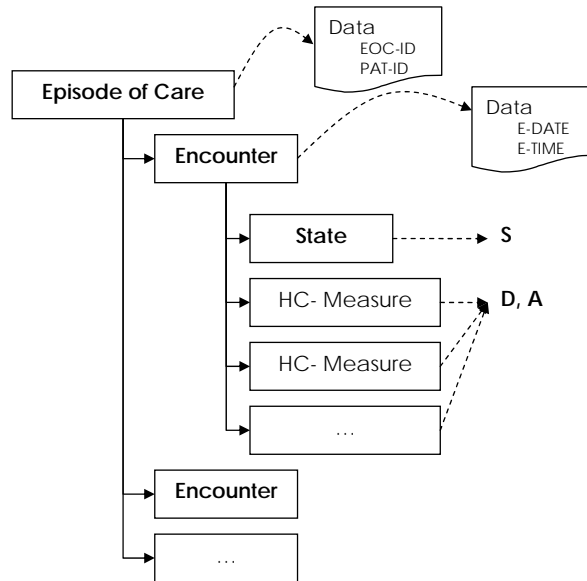


Figure 3. Scheme of Data Structure proposed for SDA\* model.

In the above grammar, EOC-ID and PAT-ID are identifiers of the EOC and the patient, respectively. So, one patient may have several EOC; for example, one for hypertension treatment and another for cardiac insufficiency. E-DATE and E-TIME are the date and the day time of the encounter (both, absolute times); S is the definition of the patient condition in the encounter expressed in terms of the vocabulary in O; D is the justification in terms of the vocabulary in O, and A is an action measure expressed in terms of the vocabulary in O. Recall that, in O, all the instances of a subclass are also instances of the parent class.

In this grammar, S, D, and A must be defined by other grammars that provide the syntax of how states, decisions, and actions are represented in terms of the ontology classes STATE, DECISION, and ACTION, respectively.

**Example 1:** the simplest way is to use a set theory approach description of EOC.

S := STATE\*  
D := DECISION\*  
A := ACTION\*

In this case, if we have  $\{s_1, s_2, \dots, s_m\}$ ,  $\{d_1, d_2, \dots, d_n\}$ , and  $\{a_1, a_2, \dots, a_p\}$  the sets of instances of the classes STATE, DECISION, and ACTION in O; then “ $\{d_1\} \{a_1\}$ ” is a correct HC-MEASURE representing the fact “since  $d_1$  is observed for the patient, the physician proposes action  $a_1$ ” and “15:01:2008 18:30:00  $\{s_1, s_2\} \{d_1\} \{a_1\} \{d_1, d_2\} \{a_2, a_3\}$ ” is a correct ENCOUNTER hold on January 15<sup>th</sup> 2008 at 6:30pm with a patient having  $s_1$  and  $s_2$ , and with two HC-MEASURES taken. Finally, “00038 P445 15:01:2008 18:30:00  $\{s_1, s_2\} \{d_1\} \{a_1\} \{d_1, d_2\} \{a_2, a_3\} 16:01:2008 8:45:15 \{s_1, s_2\} \{d_1\} \{a_1, a_4, a_5\}$ ” is a correct EOC with two encounters, one on January 15<sup>th</sup>, and the other one on January 16<sup>th</sup>, 2008.

More complex grammars for S, D, and A are possible by combining classes of the SDA\* ontology.

**Example 2:** states and decisions are defined by the ontology terms on signs and symptoms, and actions only include procedures and drugs with the dosage, the frequency and the duration of the treatment.

S := S&S\*  
D := S&S\*  
A := DISEASE MA\*  
MA := PROCEDURE + DRUG DOSAGE FREQUENCY DURATION

Observe that before using this grammar the class DOSAGE should be included in the ontology and their instances defined. It is also worth noticing that actions are related to a disease, which is the disease the action is addressed to in the treatment of co-morbidities.

**Example 3:** the use of micro-temporalities can be included in the actions.

S := STATE\*  
D := DECISION\*  
A := TA\*  
TA := ACTION [ RELATIVE , RELATIVE , FREQUENCY ]

**Example 4:** micro-temporality may affect all the terms used.

S := TS\*  
D := TD\*  
A := TA\*  
TS := STATE [ RELATIVE , RELATIVE , FREQUENCY ]  
TD := DECISION [ RELATIVE , RELATIVE , FREQUENCY ]  
TA := ACTION [ RELATIVE , RELATIVE , FREQUENCY ]

**Hierarchy inversion:** The SDA\* data model defines the relationships EOC-ENCOUNTER, ENCOUNTER-HC-MEASURE of the sort one-to-many. That is to say, for each EOC of a particular disease we may have several encounters, and for each encounter we may have several home-care measures all of them focused in the treatment of only one of the diseases of the patient. This defines an order EOC → ENCOUNTER → HC-MEASURE which may be not satisfied by some databases as for example when the patient is treated of all his diseases simultaneously in a visit. In this case, EOCs of different diseases are merged in the same encounter, and this defines an order ENCOUNTER → EOC → HC-MEASURE that must be inverted. From a grammar point of view these databases may be of the form:

ENCOUNTER := E-DATE E-TIME PAT-ID { S } EOC\*  
EOC := DISEASE HC-MEASURE\*  
HC-MEASURE := { D } { A }

Inversion is possible by sorting the loops that read such databases:

loop on the ENCOUNTER	loop on the EOC
read E-DATE	read EOC-ID
read E-TIME	read PAT-ID
read PAT-ID	loop on the ENCOUNTER
read { S }	read E-DATE
loop on the EOC	read E-TIME
read DISEASE	read { S }
loop on the HC-MEASURE	loop on the HC-MEASURE
read { D }	read { D }
read { A }	read { A }

... and storing the read information in an intermediate database that could be read in the order defined by the SDA\* data model.

**Example 5:** SAGESSA's grammar

EOC := CH-ID CIAP CIAP-DESC DATE DATE DATE TIME PHYSICIAN ...  
HC-MEASURE\*  
ENCOUNTER := E-DATE E-TIME PAT-ID { S } EOC\*  
HC-MEASURE := { D } { A }

**Example 6:** Clinical Hospital of Barcelona's grammar.

- 
1. Riaño D. The SDA Model v1.0: A Set Theory Approach. Technical Report DEIM-RT-07-001, URV, Tarragona, Spain, Feb 2007.
  2. Lamberts, H. Episode of care: a core concept in family practice - Institute of Medicine: Special Issue on Primary Care. Journal of Family Practice, 1996.
  3. Donaldson M, Yordy K, Vanselow N, eds. Defining primary care: an interim report. Washington, DC: National Academy Press, 1994.
  4. Anatomical Therapeutic Chemical (ATC), <http://www.whocc.no/atcddd/>
  5. <http://icd9cm.chrisendres.com/>
  6. <http://www.who.int/classifications/en/>

7. Physician Specialty Classification Codes (PSCC),  
<http://www.wcb.state.ny.us/content/main/hcpp/MedReg/SpecialtyClassifications.jsp>
  8. International Classification of Primary Care (ICPC), version 2,  
<http://www.who.int/classifications/icd/adaptations/icpc2/en/index.html> (and Catalan version CIAP).
- 

Tarragona, September 22, 2008