

A Study of a Health Enterprise Information System, Part 6 - Coalescing the Analysis of the ER Diagrams, Relational Schemata and Data Tables

Jon Patrick, PhD
Health Information Technology Research Laboratory
The School of Information Technologies
The University of Sydney

Abstract

Consistent weaknesses in sections of the Millenium clinical information System (CIS) are revealed in the combined study of the ERD, logical schema and the data tables. PK values are not always defined unambiguously at the design level and data tables reveal inconsistencies in declarations and data validation. There is evidence that keys are managed by software within the application rather than by the in-built functions available in the database management system leading to less confidence in data integrity.

Introduction

The purpose of this part on the discussion of the design of the Cerner software is to coalesce the observations made in the previous two parts of this study (Pts 4 & 5). The aim is to show consistencies that will warrant attention for their consequences and inconsistencies that will be assessed for their impact. This section assumes knowledge of the previous sections written in the Technology Theme.

The purpose of ER Diagrams are to construct a conceptual model of the problem that has to be represented in the final working information system. There is a well established method for translating the ER model into the relational tables of the logical model that is to be implemented, so much so that much of the translation can be performed automatically. Hence we should see standard representations in the ER Diagrams directly transferred across to the relational tables without any change. Where we see substantial differences between the two mental models there is evidence that the database implementation has deviated from the model for particular reasons that may not be at all transparent. Some of the differences noted below are explicable but others are not. The consistency between the implemented schema and the actual data can be established by studying actual data samples extracted from the database. This enables checking of data verity across tables and the checking of data as it is loaded or entered into the system.

The set of entities that are considered in this phase of the analysis is a small list. They are exemplars of different types of weaknesses and inconsistencies found in the three different information sources. The ten entities considered in this phase are shown in Table 1 along with the form of their source document. Detailed descriptions across the three information sources are presented in Appendix 1.

Entity	ERD	Schema Glossary	Data Sample
ACCESS_CONTROL_POLICY	SECURITY_APPLICATION.pdf	Screenshot	Screenshot
DCP-FORMS-ACTIVITY	DCP_DOCUMENT_MANAGEMENT.pdf	Screenshot	Screenshot
DCP-FORMS-REF	DCP_DOCUMENT_MANAGEMENT.pdf	Screenshot	Screenshot
ITEM_MASTER	not available	Screenshot	Data Table
MEDICATION-DEFINITION	not available	Screenshot	Data Table
MED_IDENTIFIER	not available	Screenshot	Data Table
PERSON	Various	Screenshot	Screenshot
PHA_PRODUCT	not available	Screenshot	Screenshot
PRSNL	Various	Screenshot	Screenshot
REF_TEXT	DCP_DOCUMENT_MANAGEMENT.pdf	Screenshot	Screenshot
REF_TEXT_RELTN	DCP_DOCUMENT_MANAGEMENT.pdf	Screenshot	Screenshot

Table 1. The list of ERDs, schemata and data samples for a subset of entities used in this analysis.

Analysis

The comparison of the three sources of information help complete the consistency check on the architecture, logical design and data storage of a sample of the Millenium suite. The ERDs demonstrate that there is an incompleteness in design identified by description inconsistencies between entity names and PKs, missing PKs, unidentified FKs, missing relationship links.

The schema glossary confirms that some of these weaknesses have permeated into the schema although some have been corrected. FK declaration seems to be whimsical at times and the lack of separation between indices and FKs makes the nature of certain declarations more uncertain. Datotyping of PKs is also inconsistent and creates uncertainty as to the behaviour of searches for PK and FK values. The apparent mis-assignment of appropriate PKs also opens up speculation as to the consistency of reconstruction of complete records for a given processing task.

The data tables confirm the weaknesses that invalid values for PKs can be introduced into the system, and that PK uniqueness is not always preserved. They display PK values as floating point numbers apparently confirm that they are stored in an ill-advised data format. The consecutive sequencing of PK values across tables implies a strategy that does not use the database management software to manage PKs and also implies a dependency between the values of the two PKs which is generally ill-advised.

The Role of Configuration in System Functionality

There is one source of weakness that we have not been able to investigate to any depth, that is, the configuration process. Configuration is a major feature of this software as the strategy of the manufacturer is to provide both the software and significant effort to configure the system to the “user requirements”. These requirements however are not created in the usual sense of a de novo systems analysis and design of user requirements before the software system is constructed, but rather defined within the limitations of the predefined software. Any requirement that falls outside the current software functionality cannot be provided.

Furthermore, the cost of providing “configuration” services has been reported variously to be up to three times the cost of the initial software license purchase. Thus configuration services are controlled by the ongoing budgetary processes of an organisation, and so the determination of the budget priorities constrain the services that are provided to the operational units of the institution.

Another constraint on achieving an “ideal” configuration of the software is the lack of adequately trained staff to do the work. We have already indicated the scale of this software is so large that it is virtually impossible for one person to understand all its variables and how to use them properly. In Australia it has been a common comment that it is difficult to find staff who are experienced in the use of this software and the manufacturer is reputedly very slow to react to any requests to changes in the system indicating they also might be hard-pressed for adequate staff.

Ultimately, while it is said that any user requirement can be fulfilled by the Cerner software, it is the clinical staff in the operational units of the organisation who fail to get the configuration they desire due to any combination of budget constraints, bureaucratic constraints, unavailability of programming staff, or missing functionality from the underlying software.

Conclusions

Cerner Millenium software has been shown to have a number of weaknesses in conceptual design, logical design and data processing that should be considered in any risk assessment of its installation. The risk assessment should evaluate the extent to which each of these weaknesses will effect the ability of an organisation to deliver on:

1. Patients’ health outcomes; and
2. Process productivity objectives of its stakeholders.

The nature of the risks and the potential cost of moderating them needs to be assessed in comparison to the cost of any culpability that should it be established on the organisation’s part where a patient has suffered some clinical misadventure or there has been a fraud perpetrated on (or even inadvertently by) the organisation. *These evaluations should be made promptly.*

The weaknesses in terms of clinical work practices, that have been identified are only likely to show up in occasional circumstances with a combination of processing and data values separated in time. Staff are not likely to associate one instance of missing or mis-processed data with another. This spasmodic nature tends to lull staff into a false sense of security that the mis-processing is either inconsequential or an accident of their own making. We recommend that each and every mis-processing experience be recorded as accurately as possible so that appropriate computational forensic analysis can correctly identify if

weaknesses in the underlying technology have been the source. In this regard, establishing a robust log of such events will be helpful.

Finally, the weaknesses portrayed herein are only from analysis of less than 1% of the full suite of programs. This weakness rate is higher than we would have expected otherwise. The density of these weaknesses in such a small sample of the system is concerning.

We recommend that organisations who use this software carefully analyse the faults they perceive in their systems and address them with the manufacturer as a matter of their first course of action.

APPENDIX 1 - Detailed Comparison of Weaknesses in ERD, Schema and Data

1. ERD, Schema and Data Comparison for ACCESS_CONTROL_POLICY

The ERD of the Security Application shows the entity ACCESS_CONTROL_POLICY with a repeated set of 4 attributes that are used for auditing purposes (see fig. 1). It is difficult to understand this construct apart from it being accidental reduplication. The table schema does not show this duplication and we should consider that as an appropriate correction of the logical model (see fig. 2). However, of the remaining attributes in the ERD, none of them appear in the relation including the PK. The attributes in the schema glossary show a consistency in naming and an appropriately named PK. It appears likely that the entity in the ERD has been incorrectly named as the PK name suggests it is the description of the entity ENCINTR_PRSNL_RELTN.

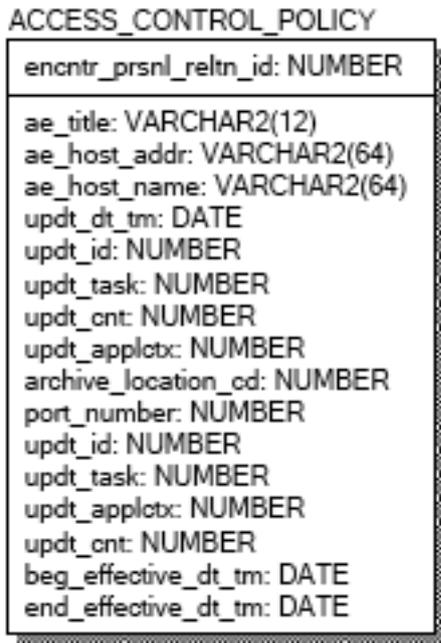


Figure 1. ACCESS_CONTROL_POLICY ERD. The PK or more likely the entity is inappropriately named and the audit stamping attributes are duplicated.

Fields	Type	Code Set	Flag	Definition
ACCESS_CONTROL_DECISION_CD	F8	4001898		The consumer's authorization decision. (I.E Permit or Deny)
ACCESS_CONTROL_POLICY_ID	F8			The primary key of the ACCESS_CONTROL_POLICY table.
ACCESS_CONTROL_POLICY_NAME	VC100			The name of the Access Control Policy
ACCESS_CONTROL_POLICY_NAME	VC100			Identical to column Access_Control_Policy_Name except it is in all capitals with special characters and blanks removed.
ACCESS_CONTROL_TYPE_ENTIT...	F8			The entity ID for the type of access control to which this policy applies.
ACCESS_CONTROL_TYPE_ENTIT...	C30			The entity name for the type of access control to which this policy applies. (I.e. CODE_VALUE)
ACTIVE_IND	I2			The table row is active or inactive. A row is generally active unless it is in an inactive state such as logically deleted, com
BEG_EFFECTIVE_DT_TM	DQ8			The date and time for which this table row becomes effective. Normally, this will be the date and time the row is added, b
DATA_SOURCE_ENTITY_ID	F8			The entity ID for the data source that defines the policy.
DATA_SOURCE_ENTITY_NAME	C30			The entity name for the data source that defines the policy. (I.e Contributor_system, Organization)
END_EFFECTIVE_DT_TM	DQ8			The date/time after which the row is no longer valid as active current data. This may be valued with the date that the row

Figure 2. Part of the Schema Glossary for the table ACCESS_CONTROL_POLICY.

2. ERD, Schema and Data Comparison for DCP_FORMS_ACTIVITY and DCP_FORMS_REF

The DCP_FORMS_ACTIVITY entity has 2 outgoing 1:M relationship links to DCP_FORMS_ACTIVITY_PRSNL and DCP_FORMS_ACTIVITY_COMP which each show the correctly declared FK *dcp_forms_activity_id*. However there is an incoming 1:M relationship from DCP_FORMS_REF, which has a seemingly misnamed PK *dcp_form_instance_id* which is not shown in the ERD of DCP_FORMS_ACTIVITY (See Fig 3). There is however in the ERDs an attribute in DCP_FORMS_REF, *dcp_forms_ref_id*, which would be its correctly formed name for the PK but it is shown only as an attribute. The same attribute is shown as an index/FK in DCP_FORMS_ACTIVITY, as we would expect if it were the

true PK of DCP_FORMS_REF. This ERD model design is repeated exactly in the schema glossaries for the two equivalent logical tables (See Figs 4 and 5).

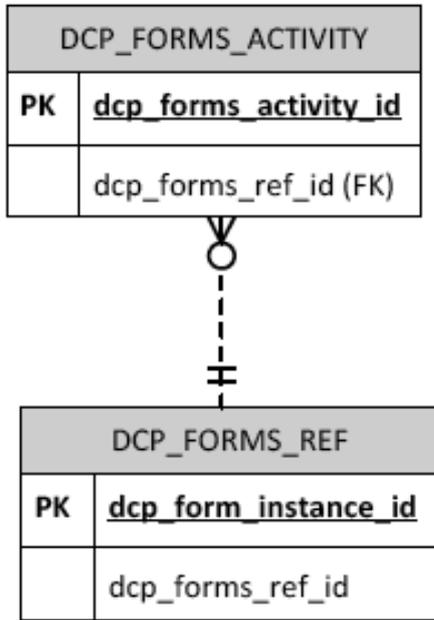


Figure 3. The ERD showing the 1:M link from DCP_FORMS_REF to DCP_FORMS_ACTIVITY and the apparently incorrectly defined PK name of *dcp_form_instance_id*, with the seemingly correct PK name declared as an attribute.

Fields	Type	Code Set	Flag	Definition
ACTIVE_IND	I2			The table row is active or inactive. A row is generally active unless it is in an inactive state such as logically deleted, co
BEG_ACTIVITY_DT_TM	DQ8			Date & Time form was first charted
DCP_FORMS_ACTIVITY_ID	F8			Unique identifier for assessment
DCP_FORMS_REF_ID	F8			Identifies form definition for activity
DESCRIPTION	VC255			This field is the display that will show up in the FormBrowser.
ENCNTR_ID	F8			This is the value of the unique primary identifier of the encounter table. It is an internal system assigned number.
FLAGS	I4			//Completion Status: 0 - Unknown 1 - Incomplete 2 - Complete
FORM_DT_TM	DQ8			Date/Time that assessment was done
FORM_STATUS_CD	F8	8		Status of the results on the form
FORM_TZ	I4			Time zone associated with the corresponding DT_TM column.
LAST_ACTIVITY_DT_TM	DQ8			Last Date/Time form was modified
LOCK_CREATE_DT_TM	DQ8			The time this row was locked for use.
LOCK_PRSNL_ID	F8			The personnel who is currently in control of this row.
PERSON_ID	F8			This is the value of the unique primary identifier of the person table. It is an internal system assigned number.
ROWID	C18			
TASK_ID	F8			Activity Task Id
UPDT_APPLCTX	F8			The application context number from the record info block.
UPDT_CNT	I4			Set to 0 on insert. Incremented by 1 on update. Used to recognize update conflict where data in a row updated by one
UPDT_DT_TM	DQ8			The date and time the row was last inserted or updated.

Figure 4. Part of the DCP_FORMS_ACTIVITY schema glossary. The PK is correctly named and an index/FK reference to *dcp_forms_ref_id* is declared.

Fields	Type	Code Set	Flag	Definition
ACTIVE_IND	I2			The table row is active or inactive. A row is generally active unless it is in an inactive state such as logically deleted, co
BEG_EFFECTIVE_DT_TM	DQ8			The date and time for which this table row becomes effective. Normally, this will be the date and time the row is added.
DCP_FORMS_REF_ID	F8			Each form has a unique reference id that is used to identify the form. Each version of the form is identified by a unique ir
DCP_FORM_INSTANCE_ID	F8			Each form has a unique reference id that is used to identify the form. Each version of the form is identified by a unique ir
DEFINITION	VC200			Textual definition of the form.
DESCRIPTION	VC200			Textual description of the form
DONE_CHARTING_IND	I2			if this indicator is set then done charting will be allowed otherwise it is not allowed on this form
END_EFFECTIVE_DT_TM	DQ8			The date/time after which the row is no longer valid as active current data. This may be valued with the date that the rc
ENFORCE_REQUIRED_IND	I2			Indicates whether to make the user enter all required fields before they are allowed to save. Without this option users ca
EVENT_CD	F8	72		Event_cd associated with this form to be used for various purposes. Currently is not used
EVENT_SET_NAME	VC100			specify an event set that is associated with this form to enable charting to print results.
FLAGS	I4			1 - Do not allow any charting unless all required fields are filled out. 2 - Show form in a maximized state
HEIGHT	I4			Height the form should be sized
ROWID	C18			
TASK_ASSAY_CD	F8	14003		Associates a discrete task assay with the form. Currently not used.
TEXT_RENDERING_EVENT_CD	F8	72		Event code associated with the text rendered from the form.

Figure 5. Part of the DCP_FORMS_REF schema glossary. The PK is seemingly incorrectly named as *dcp_form_instance_id*. The seemingly correct PK name, *dcp_forms_ref_id* is shown as an index/FK.

3. ERD, Schema and Data Comparisons for ITEM_MASTER

There is no ERD available for this entity. The ITEM_MASTER schema of itself does not show any self-evident weaknesses. It is a simple table with only a handful of attributes (see Fig 6). There are no index/FK attributes and a single PK of *item_id*. The dates for updating records show a data incompatibility warning for all dates from the 1st to the 12th of each month but not for later dates (see Fig 7).

Fields	Type	Code Set	Flag	Definition
COST_CENTER_CD	F8	14058		Cost Center
COUNTABLE_IND	I2			Indicates whether or not this item is considered countable.
CRITICAL_IND	I2			Indicates if this item is critical in order for a procedure to be performed.
FDA_REPORTABLE_IND	I2			Indicates this item is reportable to the FDA.
ITEM_ID	F8			Primary key.
ROWID	C18			
SCHEDULABLE_IND	I2			Indicates if this item is scheduable.
STERILIZATION_REQUIRED_IND	I2			Indicates if this item needs sterilization prior to use.
STORAGE_REQUIREMENT_CD	F8	11020		Describes how to store an item.
SUB_ACCOUNT_CD	F8	11030		Sub Account

Figure 6. ITEM_MASTER Schema Glossary

9579939	10	10/22/10	723943
9579939	4	10/22/10	723943
9579939	1	10/22/10	723943
9691275	11	10/25/10	723943
1443413	5	10/26/09	721957
9760967	9	10/26/10	723943
9760967	2	10/26/10	723943
1453380	2	10/28/09	1
1453380	2	10/28/09	1
1459944	6	10/30/09	721957
1467909	1	11/03/09	723943
1492478	0	11/10/09	723934
1495006	5	11/11/09	723943
9831401	16	11/15/10	723943
9831438	8	11/15/10	763958
9831401	9	11/15/10	723943
9831763	12	11/16/10	763958
9831961	10	11/16/10	723943
9831961	15	11/16/10	723943
9831709	7	11/16/10	723943

Figure 7. ITEM_MASTER data table in a spreadsheet with icons indicating invalid dates for the days between the 1st and 12th of each month.

4. ERD, Schema and Data Comparisons for MEDICATION-DEFINITION

There is no ERD available for this entity. This schema table has the same PK as the ITEM_MASTER and so in principle should be merged with it to form one table (see fig 8). However the *cki* attribute is an index/FK used for connecting to other tables, but the data values are neither unique and in about 17% of records have no allocated values.

Fields	Type	Code Set	Flag	Definition
ALTERNATE_DISPENSE_CATEGO...	F8	4008		The category used for grouping this product, when ordered in a partial unit, on fill lists and batches for continuous dispensing.
ALWAYS_DISPENSE_FROM_FLAG	I2		Y	This field helps to indicate where this product is dispensed from.
CKI	VC255			Cerner Knowledge Index field for Multum MMDC numbers. Syntax is "MUL.FRMLTN!<mmdc>"
COMMENT1_ID	F8			Link to long_text table for the first order note.
COMMENT1_TYPE	I4			Type of Order comment #1
COMMENT2_ID	F8			Key to long_text table for second order entry note.
COMMENT2_TYPE	I4			Type of second order entry comment
COMPOUND_TEXT_ID	F8			Key to long_text table for compounding instructions.
CONTINUOUS_FILTER_IND	I2			Indicates whether this product should be displayed when entering a continuous IV order
DEFAULT_PAR_DOSES	I4			The default par which overrides any default par value from the frequency schedule. Eg: PRN bulk item is ordered Q4H. T
DISPENSE_CATEGORY_CD	F8	4008		The category used for grouping this product, when ordered as a whole unit, on fill lists and batches for continuous dispensi.
DISPENSE_QTY	F8			The number of units of a pharmacy product that you should dispense for the ordered dose
DISPENSE_QTY_UNIT_CD	F8	54		Values from PHA_GET_PHARMUNIT or code set 54
DIVISIBLE_IND	I2			Defines whether this product can be split, broken, ect. To create a dose. The divisible is defaulted from the form.divisible.
FORMULARY_STATUS_CD	F8	4512		Defines the acceptance of this product by the institution.
FORM_CD	F8	4002		The dosage form of the product.
GIVEN_STRENGTH	C25			Strength of the product as retrieved from the drug database.
INTERMITTENT_FILTER_IND	I2			Indicates whether this item is selectable when building in Intermittent IV.
INV_MASTER_ID	F8			Identifies the Inventory item which QOH is tracked, in place of the formulary item. This is an FK column from table ITEM_D.
ITEM_ID	F8			Item id inherited from item_master
LEGAL_STATUS_CD	F8	4200		The legal status of the drug as assigned by the governing body.
MAX_PAR_SUPPLY	I4			For a single dispense event of this item as PRN, the maximum number of units to supply. Mostly, this will be used by multipl.
MDX_GFC_NOMEN_ID	F8			Nomenclature id for the Micromedex identifier for this product's generic formulation, including active ingredients, strength ar
MED_FILTER_IND	I2			Indicates whether this item is selectable when building a medication order.
MED_TYPE_FLAG	I2		Y	Indicates the type of this formulation.
MEQ_FACTOR	F8			strength/volume ratio expressed in millequivalents
MMOL_FACTOR	F8			strength/volume ratio expressed in millimoles.
OE_FORMAT_FLAG	I2		Y	Preferred order format for this item
ORDER_ALERT1_CD	F8	4033		First order alert code for this item.
ORDER_ALERT2_CD	F8	4033		Second order alert code for this item.
ORDER_SENTENCE_ID	F8			Order sentence id value
PARENT_ITEM_ID	F8			Identifies the parent formulary item that the child item is grouped to. This is an FK column from table ITEM_DEFINITION.
PREMIX_IND	I2			Indicates whether the medication product is a premix containing multiple ingredients.
PRICE_SCHED_ID	F8			Link to price schedule for this item.
PRIMARY_MANF_ITEM_ID	F8			Link to the manufacturer item that is currently being dispensed.

Figure 8. MEDICATION_DEFINITION Schema Glossary.

5. ERD, Schema and Data Comparison for MED_IDENTIFIER

There is no ERD available for this entity. There are a number of issues in the med_identifier table that are relevant to data quality. The data table has added to it two attributes which are not shown in the schema glossary labelled *value* and *value_key* (see fig 9). The latter is the same as the former except all non-letter and non-digit characters are removed. The motivation for adding these two fields is uncertain but we know from other schema the variation between the two attributes is created to support searching for records, for this reason the form of the values are important if the search actions are to be successful. A number of anomalies found in these columns would lead to such searches failing for particular records.

1. Incorrectly formatted National Drug Code (see fig 10.1). This is a very unexpected phenomena as the *ndc* is a PK in PHA_PRODUCT table and so it is expected the format should be checked when the values are loaded into that table. This could indicate that invalid PK values can occur in the system.
2. A new *item_id* value is allocated as the next number in the sequence of *med_identifier_id* number and the *med_identifier_ids* run as a continuous sequence of numbers. The greater majority of PK values are assigned in this way (see fig 10.2). This indicates a process external to the database management software computing the PK values across two different tables and retaining a dependency between them which is generally inadvisable.
3. Unusual value for medical device suggesting a lack of data validation (see fig 10.3).
4. Description strings with seemingly superfluous leading characters, which would lead to failed matches in search processes (see fig 10.4).

Fields	Type	Code Set	Flag	Definition
ACTIVE_IND	I2			The table row is active or inactive. A row is generally active unless it is in an inactive state such as logically deleted, combi...
FLEX_SORT_FLAG	I2		Y	flex sort flag
FLEX_TYPE_CD	F8	4062		FLEX TYPE CD
ITEM_ID	F8			ITEM ID
MED_DEF_FLEX_ID	F8			MED DEF FLEX ID
MED_IDENTIFIER_ID	F8			med identifier id
MED_IDENTIFIER_TYPE_CD	F8	11000		med identifier type cd
MED_INGRED_SET_ID	F8			med ingred set id
MED_PACKAGE_TYPE_ID	F8			med Package type id
MED_PRODUCT_ID	F8			med product id
MED_TYPE_FLAG	I2		Y	med type flag
PARENT_ENTITY_ID	F8			parent entity id
PARENT_ENTITY_NAME	C32			parent entity name
PHARMACY_TYPE_CD	F8	4500		pharmacy type cd
PRIMARY_IND	I2			primary ind
ROWID	C18			
SEQUENCE	I4			Sequence
UPDT_APPLCTX	F8			The application context number from the record info block.
UPDT_CNT	I4			Set to 0 on insert. Incremented by 1 on update. Used to recognize update conflict where data in a row updated by one ap...
UPDT_DT_TM	DQ8			The date and time the row was last inserted or updated.
UPDT_ID	F8			The person_id of the person from the personnel table (prsnl) that caused the last insert or update of the row in the table.
UPDT_TASK	I4			The registered (assigned) task number for the process that inserted or updated the row.
VALUE	VC200			value
VALUE_KEY	VC200			value key

Figure 9. The whole entry in the schema glossary for the table MED_IDENTIFIER.

T	U	V	W	X	Y
05/27/10	723934	340200	Lactated Ringers wit	LACTATEDRINGERSWITHKCL10MEQ	
05/28/10	805880	340200	99999-0000-00	99999000000	
05/28/10	805880	340200	Lactated Ringers wit	LACTATEDRINGERSWITHKCL10MEQ	
05/28/10	805880	340200	Lactated Ringers wit	LACTATEDRINGERSWITHKCL10MEQ	
05/28/10	805880	340200	Lactated Ringers wit	LACTATEDRINGERSWITHKCL10MEQ	
02/04/10	723943	340200	TPN	TPN	
02/04/10	723943	340200	TPN	TPN	
02/04/10	723943	340200	TPN	TPN	
02/04/10	723943	340200	- TPN - Inpatient -	TPNINPATIENTALLINACTIVE	
02/04/10	723943	340200	TPN	TPN	
02/04/10	723943	340200	9999000001	9999000001	
02/04/10	723943	340200	TPN Placeholder	TPNPLACEHOLDER	

Figure 10.1. MED_IDENTIFIER data table. An example of National Drug Code (NDC). In column W, Row 2 shows an correctly formatted NDC whilst row 11 shows an incorrect format. This suggests a lack of validation of PK values on loading data into the PHA_PRODUCT table.

	A	B	C	D
238	1252230	1252171	01/12/10	#NAME?
239	1252254	1252231	04/29/09	Precose
240	1252255	1252231	04/29/09	
241	1252256	1252231	04/29/09	Acarbose 25 mg Tab
242	1252257	1252231	12/14/09	Acarbose
243	1252258	1252231	04/29/09	PREC25
244	1252259	1252231	04/29/09	
245	1252260	1252231	04/29/09	
246	1252261	1252231	04/29/09	
247	1252262	1252231	04/29/09	acar25Tab - Inpatient - All - Active
248	1252287	1252231	02/16/10	00026-2863-51
249	1252288	1252231	02/16/10	Precose
250	1252289	1252231	02/16/10	Acarbose 25 mg Tab
251	1252290	1252231	02/16/10	
252	1252314	1252291	04/29/09	Precose
253	1252315	1252291	04/29/09	
254	1252316	1252291	04/29/09	Acarbose 50 mg Tab
255	1252317	1252291	12/14/09	Acarbose
256	1252318	1252291	04/29/09	PREC50
257	1252319	1252291	04/29/09	
258	1252320	1252291	04/29/09	
259	1252321	1252291	04/29/09	
260	1252322	1252291	04/29/09	acar50Tab - Inpatient - All - Active
261	1252347	1252291	08/31/09	00026-2861-48
262	1252348	1252291	08/31/09	Precose
263	1252349	1252291	08/31/09	Acarbose 50 mg Tab
264	1252350	1252291	08/31/09	
265	1252374	1252351	04/29/09	Sectral
266	1252375	1252351	04/29/09	
267	1252376	1252351	04/29/09	Acebutolol 200 mg Cap
268	1252377	1252351	12/14/09	Acebutolol
269	1252378	1252351	04/29/09	SECT200
270	1252379	1252351	04/29/09	
271	1252380	1252351	01/12/10	
272	1252381	1252351	04/29/09	
273	1252382	1252351	04/29/09	aceb2Cap - Inpatient - All - Active
274	1252407	1252351	02/25/10	00378-1200-01
275	1252408	1252351	02/25/10	Sectral
276	1252409	1252351	02/25/10	Acebutolol 200 mg Cap
277	1252410	1252351	02/25/10	
278	1252434	1252411	06/01/10	Tylenol
279	1252435	1252411	04/29/09	

Figure 10.2. MED_IDENTIFIER data table. An example of a consecutive increments of PK values of med_identifier_id (col A, row 1) has value=1252230 and its synchronicity with item_id values (col B, row 2) has value 1252231. See other examples 1252291, 1252351, 1252411.

	T	U	V	W	X
46627	07/01/10	721957	340200	NEU400	NEU400
46628	07/01/10	721957	340200	NIA500	NIA500
46629	07/01/10	721957	340200	NOL10	NOL10
46630	07/01/10	721957	340200	NOR25	NOR25
46631	07/01/10	721957	340200	NOR5	NOR5
46632	07/02/10	721957	340200	CYA1000	CYA1000
46633	07/02/10	721957	340200	CYM30	CYM30
46634	07/02/10	721957	340200	DAP100	DAP100
46635	07/02/10	721957	340200	DAR25	DAR25
46636	07/02/10	721957	340200	DDA1	DDA1
46637	07/02/10	721957	340200	DEC150	DEC150
46638	07/02/10	721957	340200	40516	40516
46639	07/02/10	721957	340200	DEM20	DEM20
46640	07/02/10	721957	340200	DEP125C	DEP125C
46641	07/02/10	721957	340200	DEP500	DEP500
46642	07/02/10	721957	340200	DEPA250	DEPA250
46643	07/02/10	721957	340200	DEPER250	DEPER250
46644	07/02/10	721957	340200	DEPER500	DEPER500
46645	07/02/10	721957	340200	DES50	DES50
46646	07/02/10	721957	340200	DET1	DET1

Figure 10.3. MED_IDENTIFIER data table. An example of a set of medical device codes with one coding value apparently anomalous suggesting a lack of data validation processes on loading data from external sources.

	B	C	D	E
1	ITEM_ID	UPDT_DT_TM	VALUE	VALUE_KEY
2	1383297	06/29/10		0
3	1413068	09/17/10		0
4	1377295	09/02/10		0
5	1413106	07/21/10		1
6	1413144	07/21/10		2
7	1599278	07/16/10	00000-0000-06	6
8	1456478	10/07/10	00000-0000-07	7
9	1456300	02/25/10	00000-0000-09	9
10	1456434	10/07/10	00000-0000-14	14
11	1460966	07/13/10	00000-0000-15	15
12	1456344	10/04/10	00000-0000-16	16
13	1599102	06/22/10	00000-0000-18	18
14	1492275	07/27/10	00000-0000-19	19
15	1492321	07/28/10	00000-0000-20	20
16	1589433	10/07/10	00000-0000-49	49
17	1589357	02/25/10	00000-0000-50	50
18	1589395	02/25/10	00000-0000-51	51
19	1589471	10/07/10	00000-0000-53	53
20	1589509	10/07/10	0000-0000-55	55
21	1589547	10/07/10	00000-0000-56	56
48049	1336059	04/29/09	Zyvox	ZYVOX
48050	1336185	07/02/10	Zyvox	ZYVOX
48051	1600788	03/23/10	zzD5250ADV	ZZD5250ADV
48052	1600836	03/23/10	zzD5250ADV2	ZZD5250ADV2
48053	1600836	03/23/10	- zzD5250ADV2 - Inpatient - All - Inactive	ZZD5250ADV2INPATIE
48054	1600788	03/23/10	- zzD5250ADV - Inpatient - All - Inactive	ZZD5250ADVINPATIE
48055	1600788	03/23/10	zzDextrose 5% in Water	ZZDEXTROSESINWATE
48056	1600788	03/23/10	zzDextrose 5% in Water	ZZDEXTROSESINWATE
48057	1600836	03/23/10	zzDextrose 5% in Water	ZZDEXTROSESINWATE
48058	1600836	03/23/10	zzDextrose 5% in Water	ZZDEXTROSESINWATE
48059	1600788	03/23/10	zzDextrose 5% in Water 250 mL ADV	ZZDEXTROSESINWATE
48060	1600836	03/23/10	zzDextrose 5% in Water 250 mL ADV	ZZDEXTROSESINWATE
48061	0	09/30/04		

Figure 10.4. MED_IDENTIFIER data table. Description strings with seemingly superfluous leading characters (last 11 rows) suggest a lack of data validation processes on loading data from external sources. These entries would fail any search for them that didn't contain the spurious characters, thus making the records inaccessible.

6. ERD, Schema and Data Comparison for PERSON

There is no ERD available for this entity. The schema glossary shows attributes that would commonly be associated with clinical patients e.g. *deceased_dt_tm* confirming this is the table for storing the patients key personal information (see fig 11). There is no data table for PERSON but one screen shot shows the values of the PK *person_id* are listed with 2 decimal places indicating that its datatype is a floating point number (see fig 12).

Figure 11. Part of the schema glossary for the table PERSON. Attribute names indicate this table is designed to represent basic information about patients.

Figure 12. PERSON screenshot of data. This data indicates that the PK, *person_id* is stored as a floating point number.

7. ERD, Schema and Data Comparison for PHA_PRODUCT

There is no ERD available for this entity. The schema glossary for PHA_PRODUCT is very large and shown here in two portions (fig 13.a, 13.b). The PK is shown to be the National Drug Code, *ndc*, with values shown from a data screenshot (see fig 14). The important issue here is that there are duplicate values of the PK which makes it impossible to know which record will be retrieved when accessing this table by the PK. The effect of this design weakness is additive to the weakness found in MED_IDENTIFIER table with ill-formatted *ndc* values.

GCR_CKI	VC100	Cerner knowledge index value for generic cross reference. CKI is MULDRUG!<multum drug_id>
GCR_CODE	C6	Generic Cross Reference code as defined by Medical economics (Redbook/Micromedex) - NO LONGER IN ACTIVE USE
GCR_DESC	VC255	Generic Cross reference description
GENERIC_NAME	VC100	Generic name for the product
GENERIC_NAME_KEY	VC100	All uppercase generic name for the product. Used for sorting the reports
GEN_LEDGER_ACCT_LUDE	VC50	General Ledger Account Code
GFC_CKI	VC100	Cerner Knowledge Index for generic formulation as defined by Multum. CKI is MULFRMLTN!<main_multum_drug_code>
GFC_CODE	C6	Generic Formulation Code as defined by Medical Economics (Redbook/Micromedex) - NO LONGER IN ACTIVE USE
GFC_DESC	VC255	Description for Generic Formulation
INFUSE_OVER	F8	Determines how long to infuse an IV

Figure 13.a. PHA_PRODUCT_1. References to *gcr_cki* and *gfc_cki* FK attributes.

Fields	Type	Code Set	Flag	Definition
MDX_GFC_NOMEN_ID	F8			Nomenclature id assigned to the Micromedex generic formulation code for this product.
MED_FILTER_IND	I2			Used to indicate product should appear on product selection lists including medication order products
MED_TYPE_FLAG	I2		Y	Flag to indicate the type of order
MEQ_FACTOR	F8			Factor for determining ratio for ml equivalents
MMDL_FACTOR	F8			Factor for ratio to millimoles
MNEMONIC	VC50			Order entry mnemonic for product
NDC	C13			National Drug Code (US FDA) identifier
NDC_ACTIVE_IND	I2			Is NDC active for product
NON_RB_IND	I2			IS item in reference database
OE_FORMAT_FLAG	I2		Y	Default order Entry Format to display for this product
OE_FORMAT_ID	F8			Order Entry Format for this type of order
ORDER_ALERT1_CD	F8	4029		First order alert for the product
ORDER_ALERT2_CD	F8	4029		Second order alert for the product
ORD_SENT_ID	F8			Order Sentence associated with the product
OS_DISPLAY_LINE	VC255			Denormalized order_sentence display line.
OUTER_QTY	F8			Outer package quantity from package_type.
OUTER_UOM	F8			Unit of measure for outer package quantity from code set 54
PRICE_SCHEDULE_ID	F8			Price schedule ID used for billing
PRICE_SCHED_SHORT_DESC	VC50			Text description for price_schedule
PRIMARY_IND	I2			Indicates whether this item_id/ndc combination is the primary choice for dispensing
PRIMARY_MNEMONIC	VC100			Primary Mnemonic used for PowerChart order entry
PRN_IND	I2			Scheduled/PRN indicator - derived from def_prn_val
PROD_ACTIVE_IND	I2			IS product active in formulary
PYXIS	C30			PYXIS interface ID for product.
QTY	F8			Dispense Quantity
REF_DOSE	C25			Reference dose, a string value, combination of strength/strength unit and volume/volume unit. Shows up in Given Strengt
REUSABLE_IND	I2			Indicates whether this product can be re-used when returned to the pharmacy for credit.

Figure 13.b. PHA_PRODUCT_2 with National Drug Code (NDC) as PK.

	MNEMONIC	FILT	WWP_FACTOR	GFC_CK	NDC	SYNONYM_ID	BRAND_NAME	MED_TYPE
1	TNF-BUL	OR.	0.00		99999-9999-06	1601997.00	TNF-Bulk Liquid	
2	TNF-BUDP		0.00		99999-9999-06	1287529.00	TNF-Bulk Topical	
3	TNF-CHF		0.00		99999-9999-05	1287567.00	TNF-Chemo Intermittent	
4	TNF-CHDR		0.00		99999-9999-07	1287668.00	TNF-Chemo Med	
5	TNF-IRRRG		0.00		99999-9999-08	1287637.00	TNF-Irrigation	
6	TNF-IVF		0.00		99999-9999-01	1287569.00	TNF-IV Piggyback	
7	TNF-LVF		0.00		99999-9999-03	1287581.00	TNF-Large Volume Parenteral	
8	TNF-TPN		0.00		99999-9999-09	1287583.00	TNF-Total Parenteral Nutrition	
9	TNF-UD	ORA.	0.00		99999-9999-02	1287698.00	TNF-Unit Dose	
10	TNF-UD	IJ	0.00		99999-9999-10	1287563.00	TNF-Unit Dose Injection	
11	TNF-UD	RAL	0.00		99999-9999-04	1287591.00	TNF-Unit Dose Liquid	
12	aceco	RAL	5.00	MUL.FRMLTN!300	65234-0046-16	1269991.00	CAPITAL/CODEINE SUSP	
13	amino	OR.	0.00	MUL.FRMLTN!504	66479-0021-82	1278261.00	AMICAR	
14	anes	IJ	0.00		99999-9999-05	1838703.00		
15	antir	RE	0.00		99999-9999-05	5945347.00		
16	ac	RAL	0.00		99999-9999-05	1735985.00		
17	balsa	OR.	0.00	MUL.FRMLTN!7752	65649-0101-02	1331330.00	COLAZAL	
18	bengate	,	60.00	MUL.FRMLTN!6158	74300-0005-30	1267585.00	BEN GAY GREASELESS	
19	benpe	3c,	42.00	MUL.FRMLTN!7136	99207-0209-01	1267708.00	TRIAZ 3+	
20	beta	1ORA.	0.00	MUL.FRMLTN!802	65473-0703-01	1275405.00	URECHOLINE	
21	beta	1,	0.00	MUL.FRMLTN!807	65473-0704-01	1275407.00	URECHOLINE	
22	beta	2ORA.	0.00	MUL.FRMLTN!807	65473-0704-01	1275409.00	URECHOLINE	
23	beta	5ORA.	0.00	MUL.FRMLTN!811	65473-0700-01	1275413.00	URECHOLINE	
24	bival	2	IV	MUL.FRMLTN!8289	65293-0001-01	2178204.00	ANGIOMAX	
25	bivit	OR.	0.00	MUL.FRMLTN!6091	70030-1328-30	1287702.00	VITAMIN B-1	
26	bupro	1	OR.		99999-9999-05	6180269.00	WELLBUTRIN XL	
27	bupro	3	OR.		99999-9999-05	6180672.00	WELLBUTRIN XL	
28	camp	10	ORA.		99999-9999-05	1287615.00	CAMP ORAL	

Figure 14. PHA_PRODUCT Table. Entries in the PHA_PRODUCT table with multiple occurrences of the PK, *ndc* : 9999-9999-06, 9999-9999-05, 65473-0704-01 and 9999-9999-05.

8. ERD, Schema and Data Comparison for PRSNL

There is no detailed ERD available for this entity. The PRSNL entity is shown in all cases with a PK of *person_id* which is the same as the entity PERSON. However in both the ERDs and schema glossary most relationships that link to this entity refer to *prsnl_id* in some form. A self reference to this table by one of its own attributes, *prsnl_type_cd* can be seen in Figure 15. A screenshot of the data table shows that the PK is represented as a floating point number, and its values show contiguous sequences (see fig 16). The attribute *p_position_disp* contains information seemingly of staff duties, e.g DBA Level III, which supports the interpretation this table is designed to represent relevant information about the employed staff of the institution.

NAME_FIRST	VC200	This is the person's first given name.
NAME_FIRST_KEY	VC100	This is the person's first given name all capitals with punctuation removed. This field is used for indexing and searching for
NAME_FIRST_KEY_A_NLS	VC400	NAME_FIRST_KEY_A_NLS column
NAME_FIRST_KEY_NLS	VC202	First Name Key field converted to NLS format for internationalization requirements
NAME_FULL_FORMATTED	VC100	This is the complete person name including punctuation and formatting.
NAME_LAST	VC200	This is the person's family name.
NAME_LAST_KEY	VC100	This is the person's family name all capitals with punctuation removed. This field is used for indexing and searching for a
NAME_LAST_KEY_A_NLS	VC400	NAME_LAST_KEY_A_NLS column
NAME_LAST_KEY_NLS	VC202	Last Name Key field converted to NLS format for internationalization requirements
PASSWORD	VC100	Encrypted form of the data store password used to authorize Millenium users access to Millenium Mobile devices.
PERSON_ID	F8	This is the value of the unique primary identifier of the person table. It is an internal system assigned number.
PHYSICIAN_IND	I2	Set to TRUE, if the personnel is a physician. Otherwise, set to FALSE.
PHYSICIAN_STATUS_CD	F8 14647	Physician status code identifies the status of the physician. (For Example: In, In Surgery, Out, etc...)
POSITION_CD	F8 88	The position is used to determine the applications and tasks the personnel is authorized to use.
PRIM_ASSIGN_LOC_CD	F8 220	Primary Assigned Location Code identifies the primary location to which a personnel will be assigned.
PRSNL_TYPE_CD	F8 309	The personnel type is used to group personnel with common characteristics (i.e., user, non-user, template user)
ROWID	C18	
SECTION_CD	F8 216	not used, to be deleted!
UPDT_APPLCTX	F8	The application context number from the record info block.
UPDT_CNT	I4	Set to 0 on insert. Incremented by 1 on update. Used to recognize update conflict where data in a row updated by one.
UPDT_DT_TM	DQ8	The date and time the row was last inserted or updated.

Figure 15. Part of the schema glossary for the table PRSNL. The PK is named as *person_id* but an attribute in this schema referring to member records uses the appropriate referencing *prsnl_type_code*.

PERSON_ID	PHYSICIAN_IND	P_PHYSICIAN_STATUS_DISP	P_POSITION_DISP	P_PRIM_ASSIGN_LOC_DISP	UPDT_APPLC
1	590083.00	0		PathNet + RadNet Transcription	8236376E
2	590103.00	0			199289E
3	590104.00	0			199289E
4	590105.00	1			362862377
5	590106.00	0		zPhysician	274564E
6	590107.00	0			199289E
7	590163.00	0		DBA	289125587
8	590183.00	0		DBA	1908513E
9	590223.00	0			232044E
10	590224.00	0			232044E
11	590263.00	0			199289E
12	590563.00	0		DBA Level II	1520861
13	590583.00	0			190851E
14	590603.00	0		DBA	16705141E
15	590657.00	0		DBA	251343E
16	590658.00	0		Analyst - 3rd Level	43011277E
17	590659.00	0		DBA	1820077
18	590705.00	0		DBA Level II	76251667
19	590706.00	0		zDBA Level III	4611180E
20	590707.00	0		DBA Level II	76251667
21	590708.00	0		DBA Level II	8824876E
22	590710.00	0		zDBA Level III	4611180E
23	590711.00	0			48048500E
24	590714.00	0			271544E
25	590717.00	1			273869E
26	590718.00	1		zRadiologist	59997808E
27	590719.00	1		Radiologist EV	34801600E
28	590721.00	0		DBA Level II	6597640E
29	590737.00	0		DBA Level II	76251667

Figure 16. A screenshot of the PRSNL data table. The PK *person_id* is displayed as a floating point number and it has contiguous sequences of values. The *p_position_disp* attribute carries values that suggest the data in this table is about personnel working in the institution.

9. ERD, Schema and Data Comparison for REF_TEXT

This entity has no PK assigned in the ERD, but *refr_text_id* appears to be a misspelt version of the appropriate attribute *ref_text_id* (see fig 17). The entity shows no relationship links with any other entity. The schema glossary indicates a concatenated PK of *refr_text_id* and *ref_text_name*, but it is mystifying that the latter attribute is being used in such a manner (see fig 18). The data table indicates that there are records without values for the PK attribute *ref_text_name* demonstrating that the uniqueness requirement is violated (see fig 19). There are no anomalies detected for the other PK attribute *refr_text_id*.

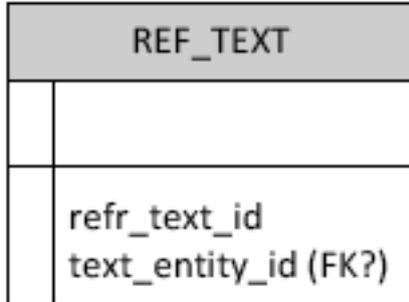


Figure 17. ERD of REF_TEXT

Table: REF_TEXT				
Fields	Type	Code Set	Flag	Definition
ACTIVE_IND	I2			The table row is active or inactive. A row is generally active unless it is in an inactive state such as logically deleted, combi...
REFR_TEXT_ID	F8			The key to the table identifying the reference text.
REF_TEXT_NAME	VC100			Ref text name
ROWID	C18			
TEXT_ENTITY_ID	F8			The id of where the text is being stored, for example may be a long_text_id.
TEXT_ENTITY_NAME	C32			The name of text type that is stored. For example could be 'LONG TEXT'.

Figure 18. REF_TEXT Schema Glossary with inexplicable PK of *parent_entity_name* and conjoint PK *refr_text_id*.

View Program		Results						
	ACTIVE_IND	REFR_TEXT_ID	REF_TEXT_NAME	ROWID	TEXT_ENTITY_ID	TEXT_ENTITY_NAME	TEXT_LOCATOR	R_TEXT
1	0	808021.00		AAA2GkAAAGAAAANKAAA	606923.00	LONG_TEXT		Nurse
2	0	808023.00		AAA2GkAAAGAAAANKAAB	606924.00	LONG_TEXT		Nurse
3	0	808025.00		AAA2GkAAAGAAAANKAAC	606925.00	LONG_TEXT		Nurse
4	0	808027.00		AAA2GkAAAGAAAANKAAD	606926.00	LONG_TEXT		Nurse
5	0	808029.00		AAA2GkAAAGAAAANKAAE	606927.00	LONG_TEXT		Nurse
6	0	808031.00		AAA2GkAAAGAAAANKAAF	606928.00	LONG_TEXT		Nurse

Figure 19. REF_TEXT screenshot of data table. The PK *ref_text_name* is shown to have missing values.

10. ERD, Schema and Data Comparison for REF_TEXT_RELTN

This entity has no PK assigned in the ERD, but an appropriate candidate would seem to be *ref_text_reltn_id*. No index/FKs are declared (see fig 20). REF_TEXT_RELTN contains the attribute *refr_text_id* which may indicate an intended relationship link to REF_TEXT. The schema glossary shows an unusual amalgam of content (see fig 21). The PK is declared to be an attribute *parent_entity_name* of datatype char which begs interpretation. The supposed PK *parent_entity_name* shows the same value for each record violating the uniqueness constraint (see fig 22). There is a declared index/FK *refr_text_reltn_id*, which we proposed as the missing PK from the ERD.

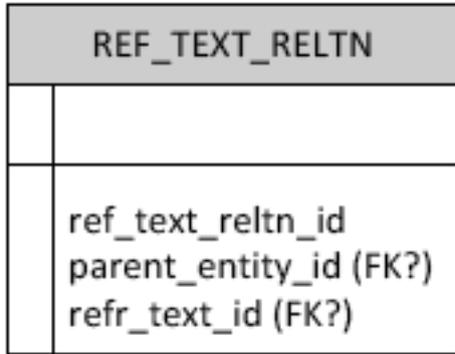


Figure 20. ERD of REF_TEXT_RELTN with no PK and putative FKs *parent_entity_id* and *refr_text_id* with no relationship links.

Fields	Type	Code Set	Flag	Definition
ACTIVE_DT_TM	DQ8			Indicates the last time that the ACTIVE_IND was toggled.
ACTIVE_IND	I2			The table row is active or inactive. A row is generally active unless it is in an inactive state such as logically deleted, cc
ACTIVE_STATUS_CD	F8	48		Indicates the status of the row itself (not the data in the row) such as active, inactive, combined away, pending purge, c
ACTIVE_STATUS_PRSNL_ID	F8			The person who caused the active_status_cd to be set or change.
BEG_EFFECTIVE_DT_TM	DQ8			The date and time for which this table row becomes effective. Normally, this will be the date and time the row is added,
END_EFFECTIVE_DT_TM	DQ8			The date/time after which the row is no longer valid as active current data. This may be valued with the date that the r
PARENT_ENTITY_ID	F8			The id of the Entity the text is being associated with, may be an Order Catalog CD or some other id/cd.
PARENT_ENTITY_NAME	C32			The name of the type of entity you are associating the text too, for example ORDERCATALOG or DISCRETEASSAY.
REFR_TEXT_ID	F8			The id for the text to be associated with this entity.
REF_TEXT_RELTN_ID	F8			The id to identify the relationship between an attribute and a piece of reference text.

Figure 21. Part of the schema glossary of the table REF_TEXT_RELTN with a PK *parent_entity_id* given a datatype char, and no declaration of the putative PK *ref_text_reltn_id*.

	END_EFFECTIVE_DT_TM	PARENT_ENTITY_ID	PARENT_ENTITY_NAME	REFR_TEXT_ID	REF_TEXT_RELTN_ID
1	12/31/00	713143.00	ORDERCATALOG	807702.00	807703.00
2	12/31/00	713147.00	ORDERCATALOG	807706.00	807707.00
3	12/31/00	713149.00	ORDERCATALOG	807708.00	807709.00
4	12/31/00	713153.00	ORDERCATALOG	807712.00	807713.00
5	12/31/00	713155.00	ORDERCATALOG	807714.00	807715.00

Figure 22. REF_TEXT_RELTN screenshot of data table. The supposed PK *parent_entity_name* shows the same value for each record violating the uniqueness constraint. The more likely true PK *ref_text_reltn_id* shows uniqueness at least in these values but is shown as a floating point number.