

Oracle Clinical for SAS Programmers

Kevin Lee, Adolor Corporation, Exton, PA

Introduction of Oracle Clinical

Oracle Clinical (OC) is the software package that is widely used in the pharmaceutical environment as a data entry and storage tool. It captures the data entered from CRF (Case Report Form) and stores the data in its database system. Unlike SAS, the OC builds the database differently.

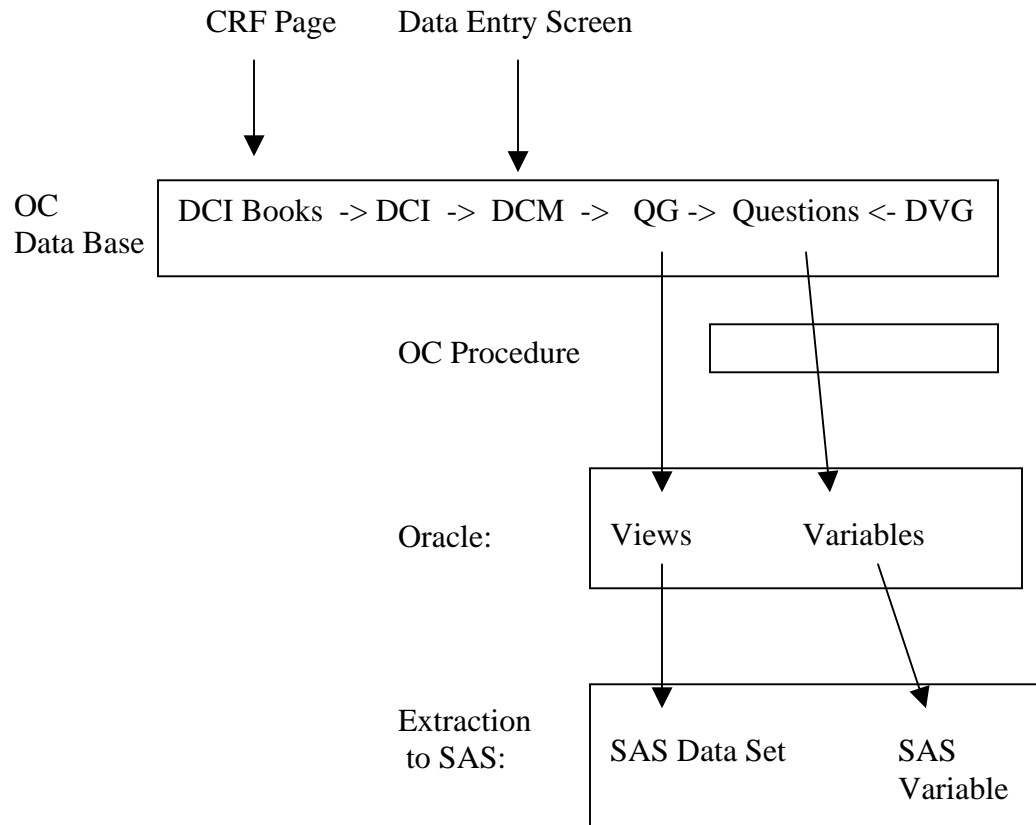
Introduction of Oracle Clinical Database

Building OC database requires the understanding of hierarchy structure in OC. The OC database developer generally builds up the OC database in the order of DVG (Discrete Value Group), Questions, QG (Question Groups), DCM (Data Collection Module), DCI (Data Collection Instrument) and DCI Books. However, its hierarchy structure is the opposite way.

The Hierarchy Structure of Oracle Clinical Database

- **DCI (Data Collection Instrument) Books:** The DCI books are a collection of DCI. It provides the order of data entry in OC.
- **DCI (Data Collection Instrument):** The DCI is a DCM or a group of DCM. One DCI usually represents each page of CRF.
- **DCM (Data Collection Module):** The DCM is a QG or a group of QG, in which the related data can be entered in a single clinical visit. As we will discuss later, one DCM represents each screen of the data entry.
- **QG (Question Group):** The QG is, like its wording explain, the group of questions. It collects the related data by grouping related questions.
- **Questions:** The question captures and stores the data entered. In SAS, it can be viewed as a variable.
- **DVG (Discrete Value Group):** The DVG is the discrete value for a question. It is also known as code list. In SAS, we can see it as a format.

The diagram for OC Database structure and its relationship with SAS



The Oracle Clinical Procedure

There are 2 types of OC Procedures – Validation and Derivation.

The Validation Procedures are usually used for validation of the data entry – SAS programmers can do this validation part by the edit check programs.

The Derivation Procedures create the derived variables in OC database. For example, the variable of age can be derived from both consent date and birth date that are entered from CRF.

The Real Example

We will discuss the part of real CRF and its hierarchical structure of Oracle Clinical database. We will also introduce its source codes in Oracle and SAS codes that can extract into SAS environment. Finally, we will show how the data entry screen looks like in Oracle Clinical environment.

**The sample page of annotated CRF
Screening**

Protocol Study000	Site No. 000	Subject No. □□□□	Subject Initials □□□
--------------------------	-------------------------------	----------------------------	--------------------------------

Informed Consent	
Date Informed Consent signed: (must be prior to all study procedures)	CONSDT □□/□□□□/200□ d d m m m y y y y
Time Informed Consent signed:	CONSTM □□:□□ (24 hour clock)

Demographics	
Date of Birth: □□ / □□□□ / □□□□□□ d d m m m y y y y BIRTHDT	Gender: SEX <input type="checkbox"/> Male (1) <input type="checkbox"/> Female (2)
Ethnicity (Check one box only):	RACE <input type="checkbox"/> Caucasian (1) <input type="checkbox"/> Hispanic (4) <input type="checkbox"/> African American (2) <input type="checkbox"/> Native American (5) <input type="checkbox"/> Asian (3) <input type="checkbox"/> Other (99) specify: _____ RACESP

Please review inclusion criteria below and check (√) appropriate boxes.

Inclusion Criteria		Yes (1)	No (0)	N/A (2)
IEINEX				
1.	Subject is a healthy man or woman.	<input type="checkbox"/>	<input type="checkbox"/>	
2.	If male, subject must be surgically sterile or agree to use an appropriate method of contraception.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> (Females Only)
3.	If female and of childbearing potential, subject must be surgically sterile.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> (Males Only)
4.	Subject is at least 18 years old or older.	<input type="checkbox"/>	<input type="checkbox"/>	
5.	Subject is able to understand the study procedures, agree to participate in the study program, and voluntarily provide written informed consent.	<input type="checkbox"/>	<input type="checkbox"/>	

This is the real example for CRF page. The investigator at site gets the above information (informed consent, demographics and inclusion criteria) from the patient, fills the CRF and sends it out to Data Manager. The data manager enters the data in the data entry screen.

The hierarchy structure of OC Database from the sample CRF page

DCI Books for Study000

Name	Description
CRF_STUDY000	CRF pages for Study000
TRACKING_STUDY000	page tracking for Study000

DCI Book Pages for CRF_STUDY000

DCI Name	Display	Start Page	Clinical Event
CONS / DEMO / INCL	1	1	SCREENING

The DCI Name of 'CONS / DEMO / INCL' represents the first page of the sample CRF and its visit is 'Screening'

DCM for the DCI Name of 'CONS / DEMO / INCL'

DCM Name	Disp Seq #	Subset Name	DCM Domain
DEMO	1	DEMO	STANDARD
INCL	2	INCL	STUDY000

DCI Name has 2 DCM, DEMO and INCL. The 'standard' of DCM domain means that the DCM of DEMO is a universal DCM and 'study000' means that the DCM of INCL is the study specific for STUDY000. In other word, INCL is the new DCM or the modified DCM.

Question Group (QG) for the DCM of DEMO

DCM Question Group Name	Library Question Group Name	Question Group Domain	Short Name	Disp Seq #
CONS	CONS	STANDARD	CONS	1
DM	DM	STUDY000	DM	2

The DCM of DEMO has 2 QG, CONS and DM. Like explained above, the QG of DM is the study specific. The 'Disp Seq #' represents the order of the prompt in data entry screen.

Question Group (QG) for the DCM of INCL

DCM Question Group Name	Library Question Group Name	Question Group Domain	Short Name	Disp Seq #
INCL	INCL	STANDARD	INCL	1

The DCM of INCL has only one QG, INCL.

Questions for the Question Group CONS

Question Name	Question Domain	Disp Seq #	Data Type	Data Time Format	Length	DVG Name	SAS Name	SAS Label
CONSDT	STANDARD	1	DATE	DMY	8		CONSDT	Consent Date
CONSTM	STANDARD	2	TIME	HM	6		CONSTM	Consent Time

The QG of CONS has 2 Questions, CONSDT and CONSTM. The sequence number also represents the order of variable prompted in data entry screen. The data type and format are the Oracle Clinical data type and format. The length represents the maximum of data can captured in data entry. The SAS Name and SAS label will be the name and label of the variable name and label when extracted to SAS environment.

Questions for the Question Group DM

Question Name	Question Domain	Disp Seq #	Data Type	Data Time Format	Length	DVG Name	SAS Name	SAS Label
BIRTHDT	STANDARD	1	DATE	DMY	8		BIRTHDT	Birth Date
SEX	STANDARD	2	CHAR		1	SEX	SEX	Sex
RACE	STANDARD	3	CHAR		2	RACE	RACE	Race
RACESP	STANDARD	4	CHAR		40		RACESP	Other Race

The QG of DM has 4 Questions, BIRTHDT, SEX, RACE and RACESP. Unlike QG of CONS, GENDER and RACE have DVG. We will discuss DVG later.

Questions for Question Group INCL

Question Name	Question Domain	Disp Seq #	Data Type	Data Time Format	Length	DVG Name	SAS Name	SAS Label
IEINEX	STANDARD	1	CHAR		6	INCLQ	INQ	Inclusion Criteria
IEINA	STANDARD	2	CHAR		2	YESNO	INA	Inclusion Response

The QG of INCL has 2 Questions, IEINEX and IEINA. Notice that the questions of IEINEX and IEINA will be renamed to INQ and INA in output. This kind of practice is not recommended, but it can be done.

Discrete Value Group(DVG) for Questions, SEX, RACE, INQUEST and INANS

DVG Name	DVG Domain	Description	Seq #	DVG Value	DVG Long Value
SEX	STANDARD	Gender	1	M	Male
			2	F	Female
RACE	STANDARD	Race	1	1	Caucasian
			2	2	African American
			3	3	Asian
			4	4	Hispanic
			5	5	Native American
			99	99	Other
YESNO	STANDARD	Yes/No	0	NA	Not Applicable
			1	Y	Yes
			2	N	No
INCLQ	STUDY000	Inclusion Criteria	1	INCL1	Subject...
			2	INCL2	If male...
			3	INCL3	If female...
			4	INCL4	Subject is ...
			5	INCL5	Subject is able...

For the question of SEX, the data entry screen will prompt the choice 1 or 2. If you choose 1, the screen will display the value of ‘M’. The question of SEX with DVG SEX creates 3 variables – SEX(Sex), SEXN(Sex – DVN) and SEXL(Sex – DVL). Here, SEX contains the value from ‘DVG Value’; likewise, SEXN the value from ‘Seq #’ and SEXL the value from ‘DVG Long Value’.

The above Oracle Clinical structure creates the views in Oracle Database. We have to understand that Oracle Clinical is the interface software for Oracle, so whatever we do in OC will be saved in Oracle.

The views created by the above Oracle Clinical

The views are CONS, DM and INCL.

The variables created in the view of CONS are the following. You have to see that the standard variables such as SUBJID, SITEID, SUBJINIT, VISIT, VISITNUM and more (we won’t be using all the standard variables) will be created for all the views.

View Name	Variable Name	Variable Length	Variable Label	Variable Type
CONS	SUBJID	8	Subject ID	NUM
	SITEID	10	Site ID	CHAR
	VISIT	30	Visit	CHAR
	VISITNUM	8	Visit Number	NUM

			
	CONSDT	8	Consent Date	CHAR
	CONSTM	4	Consent Time	CHAR
DM			
	BIRTHDT	8	Birth Date	CHAR
	SEX	1	Sex	CHAR
	SEXN	1	Sex – DVN	NUM
	SEXL	6	Sex – DVL	CHAR
	RACE	2	Race	CHAR
	RACEN	2	Race – DVN	NUM
	RACEL	15	Race - DVL	CHAR
	RACESP	40	Other Race	CHAR
INCL			
	INQ	4	Inclusion Criteria	CHAR
	INQN	1	Inclusion Criteria – DVN	NUM
	INQL	200	Inclusion Criteria - DVL	CHAR
	INA	2	Inclusion Response	CHAR
	INAN	1	Inclusion Response – DVN	NUM
	INAL	14	Inclusion Response - DVL	CHAR

The source code for view CM

The OC is the interface of Oracle. So, whatever is created in OC is also written in Oracle. We can view the source code for view CM in PL/SQL Developer. Please notice that real source code is a lot more complicated and the following is the simplified version.

```

create or replace view study000.demo as
  select select 'Study000' as study, a.subjid, a.siteid, a.visit,
         a.visitnum,,,
         b.consdtd, b.consttm
  from rdcn a, response_view b,,,
  ..
quit;

```

The sample code for extracting program

This view can be extracted by SAS program. The following is the example code for extraction program.

```

libname raw "C:/Study000/Raw";
libname oadb oracle path="" user="" password="" schema=" Study000";
proc sql;
  create table raw.cons as
    subjid, siteid, visitnum, visit,,,
    consdt, constm
  from oadb.cm
quit;

```

The above code creates SAS data set, cons, in c:/Study000/Raw. We can create other SAS raw data sets in the same way.

The data entry screens in Oracle Clinical

The database entry screen looks like the following. This is the DCI of 'CONS / DEMO / INCL'. There two screens for the sample CRF page.

The first page is the DCM of DEMO. It has two Question Groups, CONS & DM.

INFORMED CONSENT

Consent Date

Consent Time

DEMOGRAPHY

Date of Birth Sex 1=Male/2=Female

Race 1=Caucasian/2=African American/3=Asian/4=Hispanic/5=Native American/99=Other

Race, Specify

The second page is DCM of INCL. It has only one Question Group, INCL.

INCLUSION CRITERIA

Criteria 1=Yes/0=No/2=NA

<input type="text" value="IEINEX"/>	<input type="text" value="IE"/>
<input type="text" value="IEINEX"/>	<input type="text" value="IE"/>
<input type="text" value="IEINEX"/>	<input type="text" value="IE"/>
<input type="text" value="IEINEX"/>	<input type="text" value="IE"/>
<input type="text" value="IEINEX"/>	<input type="text" value="IE"/>

As shown above, each DCM represents each page of data entry.

The advantage of knowing Oracle Clinical Database

The understanding of Oracle Clinical database helps the SAS programmers to link the relationship between CRF and extracted raw data. SAS programmers will be able to understand how many data sets are created for each CRF page and how many events or visits should be obtained in each data set.

It also helps to understand what each variable in raw data means. The Oracle Clinical creates more data than entered. Some are standard and some derived. When merging more than one raw data set to create analysis data sets, SAS programmers will be able to find better variables to merge with.

SAS programmers and OC developers can also work together to automate the process so that SAS programmers do not need to spend too much time in programming. For example, both can decide to use the consistent variable names so that SAS programmers do not need to rename the variables to create analysis data sets. This process will help to save the time and efforts in both sides.

SAS programmers can have better communications and working relationships with data management teams. The understanding of OC database helps us to have a better understanding on the data management side - their issues, problems and even shortcomings.

Therefore, it helps us to be better SAS programmers. Our job descriptions go beyond programming. We interact with other team members. We need to understand how the whole study team works. The SAS programming is a part of process. Knowing more process makes us better SAS programmers.

Conclusion

SAS programmers need to understand that the Oracle Clinical is mainly used for the data management while SAS for statistical analysis. The OC database is built based on this idea of helping the data management. SAS programmers extract the raw data sets from OC database and generate the analysis data sets. In other words, SAS programmers convert the idea of data management to that of statistical analysis. The understanding of OC database helps to narrow the gaps between OC and SAS environment.

Contact Information

Kevin Lee
Adolor Corporation
Exton, PA
(448) 595-1907
Email: klee@Adolor.com, Kevin.kyosun.lee@gmail.com