

# Electronic Data Submission capabilities of NCI Center for Cancer Research

## Introduction

The purpose of this document is to describe relevant architecture details of NCI's Clinical Data Management system and the electronic data submission capability from this database to external sponsors and monitors.

## C<sup>3</sup>D Architecture

The clinical trials data for NCI Center for Cancer Research (CCR) studies are maintained in the Cancer Central Clinical Database (C<sup>3</sup>D). This database is built around an industry standard Clinical Data Management application called Oracle Clinical. The C<sup>3</sup>D Application provides tools to model, standardize, and maintain clinical data collection and reporting modules. Starting with these standards, the application further helps modeling of study specific modules.

## Modeling CRFs

The C<sup>3</sup>D architecture provides a method of readily deploying study specific Data Entry Forms and Validation/Derivation Rules corresponding to the Case Report Forms (CRF). It achieves this by introducing a meta-data layer that stores the CRF definitions and layouts at the study level. The clinical data itself is stored together for all studies in common tables and can be readily accessed via study specific Oracle Views, which are described later.

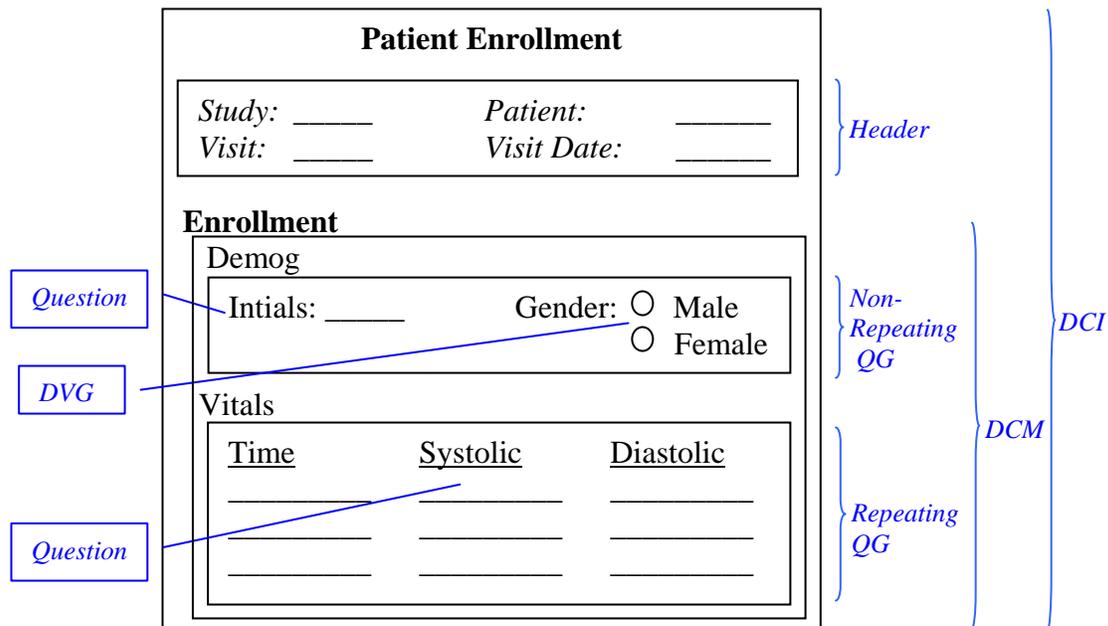


Figure 1 C<sup>3</sup>D Data Structures

Refer to Figure 1 for an illustration that would be used to introduce the terminologies relevant to this discussion.

Any Data Element to be collected on a CRF is modeled as a *Question* in C<sup>3</sup>D. Questions are based on NCI's Common Data Element (CDE) standards as maintained in the Cancer Data Standards Repository (CaDSR) (<http://ncicb.nci.nih.gov/core/caDSR>). The Questions can be modeled with data type: Character, Numeric, Date, and Time. List of acceptable responses could be associated with Questions. These lists are referred to as *Discrete Value Group (DVG)* within C<sup>3</sup>D. The DVGs are constrained by the CDE definition.

A group of logical Questions are put together in a *Question Group (QG)*. A QG that allows for multiple repeats is called a Repeating QG. A logical CRF page containing one or more QGs is called a *Data Collection Module (DCM)*. A physical CRF page that can contain one or more logical CRF pages or DCMs is called a *Data Collection Instrument (DCI)*. The hierarchy of DCMs within DCI allow for similar logical information (DCM) to be collected on various physical pages (DCI) and yet be consolidated in one place for reporting.

In the current example ( Figure 1), five questions are used namely Initials, Gender, Time, Systolic, and Diastolic. The Gender Question is associated with a DVG that allows for 'Male' and 'Female' as valid values. These questions are contained in two question groups: 'Demog' and 'Vitals'. While the Vitals QG is a repeating QG, the Demog QG is Non-repeating. The two QGs are contained in the logical CRF (DCM) 'Enrollment'. The logical CRF is part of the physical CRF (DCI) 'Patient Enrollment'.

A standard set of CRFs have been modeled in C<sup>3</sup>D for various phases of a typical oncology trial. Each of these CRFs have corresponding DCI/DCM definitions, data entry layouts, validation/derivation rules, and detailed instructions for completion of the form. While developing new studies in C<sup>3</sup>D, these standard forms are used as a starting point and modifications are made to accommodate study specific needs. Additional DCI/DCMs may be developed to capture concepts not covered by the standards.

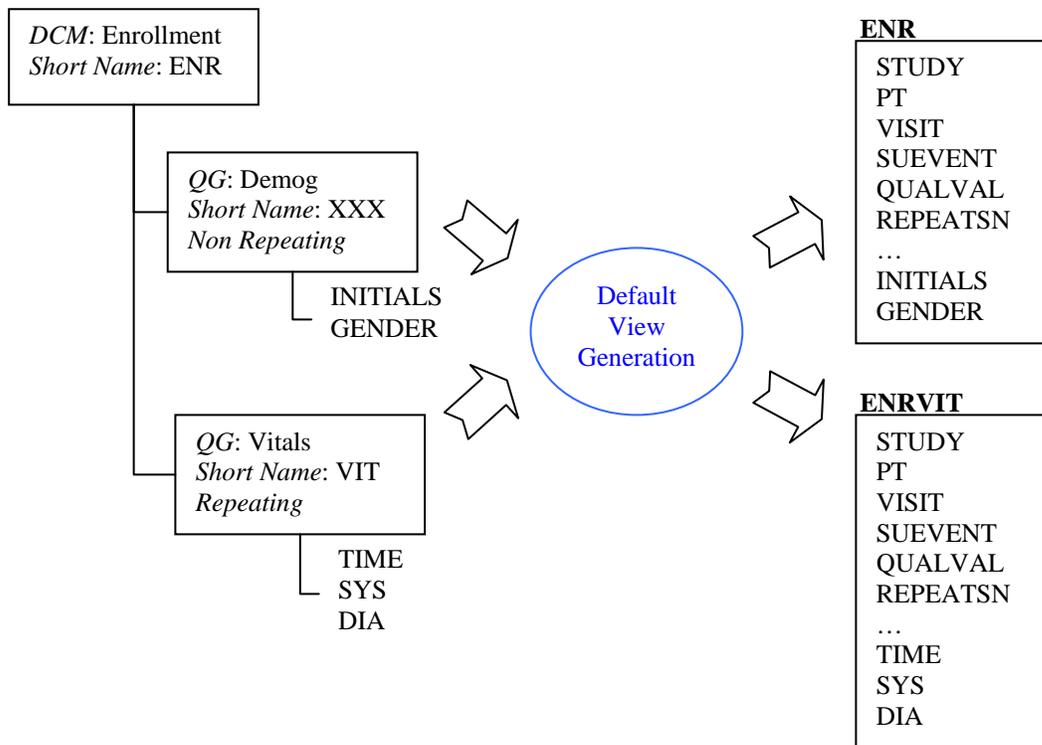
## **Study Specific Oracle Views**

C<sup>3</sup>D provides capability to generate default study specific Oracle Views from a menu driven utility. One or more Views are generated for every DCM in the study. Though the data could have been collected on various CRFs or DCIs, they can be brought together in one view provided the same DCM was used to collect the information.

C<sup>3</sup>D creates one Oracle View corresponding to all the non-repeating information in a DCM. For each repeating question group within a DCM, an additional dedicated view is generated. In the above example, two Oracle Views will be generated corresponding to the Enrollment DCM: one containing the information from the Demog QG and the other containing the information from the Vitals QG. The naming convention for the view is based on the short names assigned to the DCMs and the QGs within. See Figure 2 for an illustration based on the earlier example.

Each record in the DCM specific view is uniquely identified by standard header columns (key): Patient, Visit, Sub-event, Qualifying Value, and Repeat Sequence Number. The Sub-event number provides the capability of entering an unplanned CRF within a visit. For example an additional unplanned Physical Exam CRF could be entered in Course 1 visit using a Sub-event for Course 1. The Qualifying Value is a user defined key at the DCM level to allow for planned reuse of the DCM within a visit. For example, the same DCM ‘Checklist’ could be used to capture both the Eligibility and Ineligibility Checklist criteria in a Baseline visit using a qualifying value; two instances of the DCM will be presented at baseline and will be distinguished by a Qualifying Value indicating Eligibility or Ineligibility. The Repeat Sequence Number is to represent each row in the repeating QG.

These Key columns are presented in each DCM specific view. Apart from the unique key, additional header columns specifying information about the entered data could be made available in the View, such as Data Entry Time Stamp, Data Update Time Stamp, Data Entry Operator ID, DCI Name, Document Number, etc.



**Figure 2 View Naming Convention**

### Snapshot Views

C<sup>3</sup>D also provides the capability to create study specific snapshot oracle views. A Snapshot view is defined as a collection of the data as it appears on the paper CRF. This

type of view provides the capability to retrieve study specific data as it was during a certain data entry period of time. Since the system internally timestamps the data entry activities, these views can be generated at any point in time independent of the data entry process.

## **Full Values**

Though the Questions can be modeled as Character (with or without pick lists), Numeric, or Date values, C<sup>3</sup>D allows data to be entered in the database even if it does not adhere to the format. The system marks such data as discrepant and saves it separately as an exception value text. However, if the data adheres to the format, it is termed store as a Value Text. During the study specific retrieval process, C<sup>3</sup>D provides an option to present one of the following in the DCM Views:

- Value: Question response that has been validated to adhere to the Question definition and is blank for discrepant values.
- Exception Value: Discrepant Question response.
- Full Value: Value if present, else exception value.

NCI policy is to output both Value text and Full Value text for each Question in all views. The Full Value Question are indicated by appending ‘\_FUL’ to the Question name, e.g. SYS and SYS\_FUL.

## ***Electronic Submissions***

Data from all these study specific views can be made available electronically in various standard formats: Pipe delimited ASCII, Tab delimited ASCII, SAS dataset, etc. Along with these Data files, CRF Annotation Reports and View Description Reports can be made available as well. There are plans underway to provide HL7 based messages.

Electronic submissions can be made to the sponsor using secure standard technologies, such as secure-FTP, encrypted zip files, etc. The frequency of the submission is pre-determined to be monthly, quarterly, etc. The submissions will be cumulative in nature.

Changes to the study design that impact the View structure can also be communicated to the sponsor ahead of the change.