



CDISC LAB MODEL PRODUCTION VERSION 1.0.1

Revision History

Date	Version	Summary of Changes	Primary Author
10/31/02	1-0-0	INITIAL RELEASE	LAB TEAM
11/26/02	1-0-0	Changes to Sections 5 & 6	Susan Bassion / Julie Evans
3/2/03	1-0-1	Minor modifications to several sections. See Release Notes for detailed list.	M. Pleshe/S. Bassion/P.Pochon

Table of Contents

1	INTRODUCTION TO CDISC	4
1.1	Background	4
1.2	The Scope of CDISC Activities.....	4
2	THE LAB TEAM	5
2.1	Team Members	5
2.2	Purpose of the Lab Team.....	5
3	THE LAB MODEL	6
3.1	Defining the problem.....	6
3.2	Designing a New Model.....	7
3.2.1	Assumptions Made	7
3.2.2	Approach Taken	8
3.3	Contents of the Model	9
3.3.1	Data Fields.....	9
3.3.2	Code Lists.....	9
3.3.3	Date and Time Format Conventions.....	11
3.4	Using the Lab Model	12
3.4.1	General	12
3.4.2	Good Transmission Practice.....	12
3.4.3	Study.....	13
3.4.4	Site.....	13
3.4.5	Investigator.....	13
3.4.6	Subject.....	14
3.4.7	Visit	14
3.4.8	Accession.....	15
3.4.9	Record Extension Type	15
3.4.10	Base Specimen.....	15
3.4.11	Base Battery	17
3.4.12	Base Test	17
3.4.13	Base Result	19
3.5	Comments on Practical Application	23
3.5.1	Use of Model Fields	23
3.5.2	Data Compression	23
3.5.3	Transmission Agreements	23
3.6	Testing of the Lab Model.....	23

3.7	Use Case Scenarios	25
4	REVIEW PROCESS	25
5	NEXT STEPS.....	26
6	DIRECTIONS FOR SUBMITTING COMMENTS	26
7	APPENDICES.....	27
7.1	APPENDIX A: Synonyms.....	27
7.2	APPENDIX B: Example Specimen Condition List	28
7.3	APPENDIX C: Index	29

CDISC LAB MODEL PRODUCTION VERSION 1.0.1

1 Introduction to CDISC

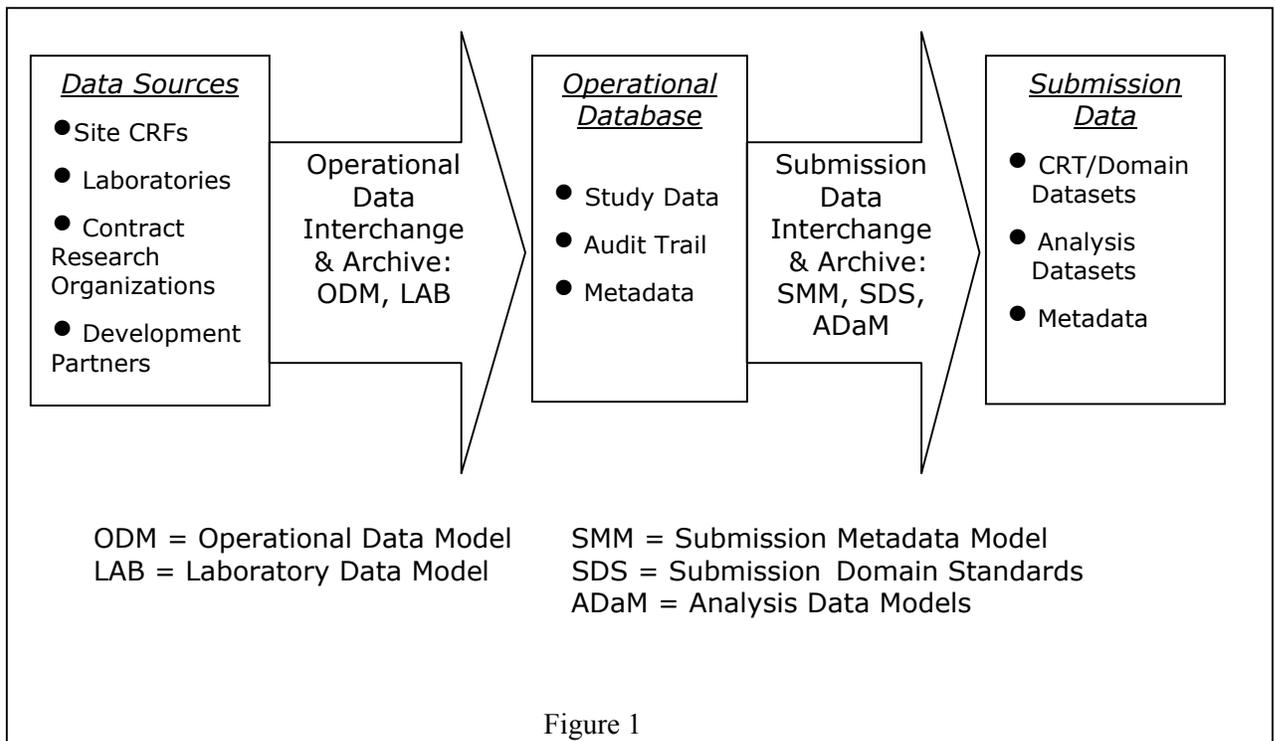
1.1 Background

CDISC is an open, multidisciplinary, non-profit organization committed to the development of industry standards to support the electronic acquisition, exchange, submission, and archiving of clinical trials data and metadata for medical and biopharmaceutical product development.

The mission of CDISC is to lead the development of global, vendor-neutral, platform-independent standards to improve data quality and accelerate product development in our industry.

1.2 The Scope of CDISC Activities

CDISC seeks to achieve its mission through the development of standard data models designed to support the end-to-end data flow of clinical trials from the data sources into an operational database and through to analysis and submission, as shown in Figure 1 below:



2 The LAB Team

2.1 Team Members

The members of the LAB team represent a broad cross-section of stakeholders from within the biopharmaceutical industry who have an interest in clinical laboratory data. The team is currently comprised of representatives from 4 pharmaceutical companies, two technology companies, one CRO and 4 central laboratories. The Lab Team has had representation from the following companies and groups during the development of the Lab Model:

ACM Labs
AstraZeneca
CRL Medinet
Covance CLS
Duke Clinical Research Institute
Eli Lilly and Company
GlaxoSmithKline
HL7
IBM Life Sciences
LabCorp
Lincoln Technologies
Merck
MRL
PharmaNet
Physician's Reference Laboratory
Pfizer
PHT
Quintiles
Quintiles Lab
Regenstrief Institute
Schering AG

The membership of the team has incorporated expertise from a variety of clinical, technical and statistical disciplines and a variety of different perspectives including academic, commercial and European.

2.2 Purpose of the Lab Team

One of the largest components of clinical trial data is laboratory data, making the development of a model to support it an important CDISC initiative.

A customer requirements survey was conducted by a focus group of the Operational Data Modeling Team to determine industry priorities for the development of standard models to support data interchange for clinical trials. Based upon these results, the highest priority for standards to facilitate data interchange was placed upon those standards that would facilitate the exchange of clinical laboratory data into central data management systems. This was followed by a close second for standards that would facilitate the exchange of data from CROs to sponsor companies.

Formed in 2000, the Laboratory Data Team (LAB) has as its mission the development of a standard model for the acquisition and interchange of laboratory data.

The mission of the Lab Team is to:

- Define requirements for improving operational laboratory data interchange.
- Develop a standard content model for the acquisition and interchange of clinical trials laboratory data
- Test the model with complex real laboratory data to assure its functionality
- Explore other opportunities to improve laboratory data processing with standards

3 The LAB Model

3.1 Defining the problem

Standard models for the interchange of laboratory data do exist already but they are very seldom used within the biopharmaceutical industry. Examples of such standards are ACDM, ASTM, HL7 and X12. The main reason standards such as these have not been more accepted by the industry is that they have limited applicability to clinical trial data and hence have limited use to central laboratories, CROs or biopharmaceutical companies.

Although each of the different standards available have their relative advantages and disadvantages and, even though they may be good standards for the interchange of healthcare information in general, they do not accommodate the specific requirements of clinical trial laboratory data interchange because they have structures and rules which are not easily applicable to clinical trial laboratory data.

Common complaints from the industry are that these standards are difficult to use, are inefficient, have inadequate field definitions (because they either ask for data that is not relevant to clinical trials or do not ask for data that *is* relevant to clinical trials), have population rules which do not match the structure and inter-relationships of clinical trial data and are unnecessarily complex.

In the absence of acceptable industry standards, each CRO and biopharmaceutical company has developed their own, specifically designed for their own particular needs. Furthermore, these standards rarely apply to all clinical trials and instead usually tend to be developed on a per-study basis. This has led to there being a plethora of standards within the industry. Large central laboratories support several hundred. This situation causes a variety of problems for everyone concerned because of the increased resources required to handle them all in terms of staff, staff training, development time, development cost, quality control, data interchange, data verification and issue resolution: everything in fact involved in exporting the data out of one clinical data management system in such a way that it can be successfully imported into another.

It is estimated that the cost to the industry per year simply of laboratory data interchange itself is at least \$150m and that between approximately 30% and 60% of that cost could be saved from the use of a standard. However it is recognized that the real—and substantially greater—value that faster and higher quality data interchange brings is in terms of time savings in an industry with running costs estimated at \$1m per day.

Despite the costs and quality issues of laboratory data interchange, the situation has been perpetuated because, in the absence of usable standards as reasonable alternatives, there has been little or no incentive for CRO and biopharmaceutical companies to change.

The LAB Team has put together the following estimates of the current time cost of doing business with the current lack of standardization:

- Central Lab estimates of time required to set up a new study with new data requirements 1.5 – 2.0 months
- Pharmaceutical company estimates of time required to:
 - Set up new study with new lab (depending on lab's technical capabilities) 2 – 12 months
 - Set up a new study with new data requirements with an existing lab (depending on the complexity of data requirements) 1 – 2 months
 - Set up new study for already supported data types with existing lab 0.5 – 1 month
- Total time estimates for set up at both sending and receiving ends 2 – 14 months

3.2 Designing a New Model

3.2.1 Assumptions Made

The assumptions made in the design of the model were based upon the analysis why the existing standards are so little used. They can be summarized as follows:

- The model should offer the industry clear advantages over other clinical trial data interchange standards.
- A successful model should be designed specifically to handle laboratory data acquired in clinical trials.
- The structure and contents of the model should be intuitive and clearly understandable to industry stakeholders familiar with clinical trial data and should have straightforward and easy to follow rules.
- The model should be sufficiently flexible that it could be applied to any clinical trial and keep pace with industry changes but not sacrifice simplicity in an attempt to cope with extreme cases.
- The first release of the model should be as comprehensive as possible to avoid the need for continual updates and revisions in the future other than those related to changes within the industry.
- The model should not be limited to any one specific implementation and so risk rejection by industry stakeholders who would otherwise be willing to embrace it but for whom the implementation chosen is incompatible with their technical infrastructures.
- The development of the model should concentrate on content first and implementation second.

- The model should accommodate as much as possible the different practices of the many central laboratories, CROs and biopharmaceutical companies within the industry.
- The model should support data interchange between any organizations in the industry and not just the classic flow from lab to CRO to sponsor. For example: reference laboratory to central laboratory to CRO to biopharmaceutical company to partner biopharmaceutical company.
- The model should use existing standards and draw upon the work of existing standards organizations wherever appropriate (for example, code lists from HL7, ISO, LOINC and NCI).
- The model should allow some controlled flexibility in the way that some data is represented to support differing preferences within the industry.
- A separate model for the interchange of reference range data should be made and be based upon the laboratory data model.
- The model should support both incremental and cumulative data interchange.
- The model should support the interchange of data from either one or more studies within a single file.

3.2.2 Approach Taken

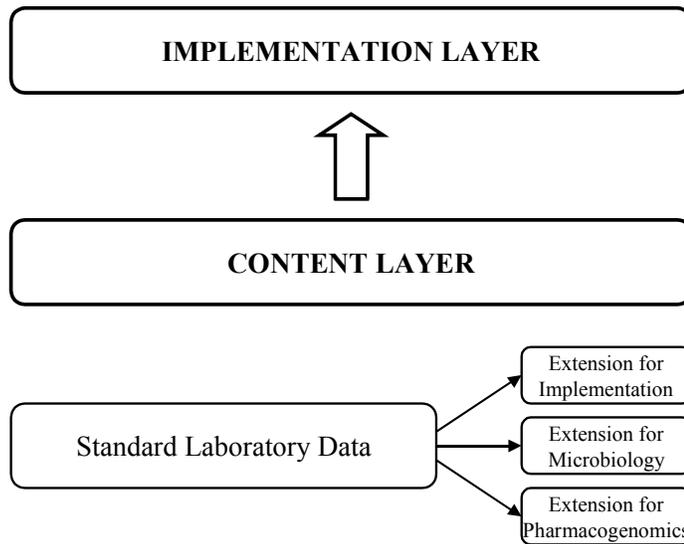
The first step was to define exactly what the industry means by ‘clinical laboratory data’ in order to build a ‘superset’ of data items that fully describes a clinical trial to the satisfaction of all the stakeholders involved in it. This superset of fields constitutes the content of the model.

Next, the structural definitions of those data items were defined in terms of data type, length, default values, standards of representation, code lists and whether or not they should be optional or required. Wherever appropriate existing standards were employed. For example: ISO 8601 date and time representations and some HL7 code lists are used.

It was decided that the content should have a main core designed to handle ‘simple’ laboratory data with the classic ‘one test, one result’ data structure of testing such as routine Chemistry for example and that extensions designed to handle more ‘complex’ laboratory data such as Microbiology and Pharmacogenomic would be added later. The main core has been termed the “Base Model”. Extensions will be designated in the future by their application.

From there the notion of a ‘multi-layer’ model was developed whereby the first layer would be the content layer and above that would be an implementation layer, the idea being that the content would not change but the implementation could. The advantage of this approach is that it offers flexibility but retains control: it doesn’t make the use of the model dependent upon any one implementation and if different implementations are used the content remains the same so the standard still applies.

The design of the model is thus as follows:



3.3 Contents of the Model

3.3.1 Data Fields

The superset of data fields is separated into 12 logical levels as follows:

1. Good Transmission Practice
2. Study
3. Site
4. Investigator
5. Subject
6. Visit
7. Accession
8. Record Extension Type
9. Base Specimen
10. Base Battery
11. Base Test
12. Base Result

These levels were chosen because they follow the recognizable hierarchy of clinical laboratory data.

3.3.2 Code Lists

Since the purpose of the model is to improve the interchange of laboratory data it is important to consider not only the structure and the contents of the model in terms of the fields it contains and

how they are organized but also the population of those fields. Accordingly, for some fields, code lists have been suggested. The purpose of these code lists is to offer a higher degree of standardization and so further improve the reliability and accuracy of data interchange.

The use of LOINC codes to supplement internal test codes is recommended. A more standardized test code system will permit higher-level analysis of data and will be of assistance in the future should regulatory authorities mandate such a standardized coding system. Note that on March 21, 2003 the U.S. Department of Health and Human Services announced that all federal agencies that deal with health care data will adopt laboratory LOINC codes to standardize electronic interchange of clinical laboratory results. Some pharmaceutical companies have already made the switch from proprietary internal codes to LOINC. Although many companies will not be able to use LOINC exclusively, supplementing internal codes with LOINC should be advantageous.

Some of the suggested code lists are as follows:

Variable	Code List
Specimen Material ID and Name	HL7 V 2.4 Specimen Source Code Table 0070
LOINC Code – The corresponding LOINC code for the Test or Lab Test ID.	LOINC
Units (Reported, Conventional, SI)	HL7 V 2.4 Figure 7-9 HL7 Common ISO derived units and ISO+ extensions
Toxicity Grade	NCI Common Toxicity Criteria

A full list of recommended codes can be found in the CDISC Laboratory Data Interchange Standard Transmission Data Fields document at www.cdisc.org. The CDISC Specimen Condition code list is included in Appendix A of this document.

The recommended HL7 code tables can be accessed on the HL7 website at www.hl7.org. Use of the HL7 code mnemonic is recommended.

The LOINC[®] (Logical Observation Identifier Names and Codes) database provides a set of universal names and ID codes for identifying laboratory and clinical tests. The purpose is to facilitate the exchange and pooling of laboratory test and observation results.

Each LOINC code corresponds to a single test definition. The LOINC test code definition includes fields for specifying:

- 1) Component (analyte) — e.g., potassium, hemoglobin, hepatitis C antigen.
- 2) Property measured — e.g., a mass concentration, enzyme activity (catalytic rate).
- 3) Timing - i.e., whether the measurement is an observation at a moment of time, or an observation integrated over an extended duration of time — e.g., 24-hour urine.
- 4) The type of sample — e.g., urine; blood.
- 5) The type of scale — e.g., whether the measurement is quantitative (a true measurement) ordinal (a ranked set of options), nominal (e.g., *E. coli*; *Staphylococcus aureus*), or narrative (e.g., dictation results from x-rays).
- 6) Where relevant, the method used to produce the result or other observation.

The LOINC codes are not intended to transmit all possible information about a test or observation. They are only intended to *identify* the test result or clinical observation. Other fields in the message can transmit the identity of the source laboratory and very detailed information about the sample or test method.

The LOINC database currently includes entries for more than 30,000 tests and is available without charge from the Regenstrief Institute at <http://www.loinc.org>. The Lab Team has worked with the Regenstrief Institute to develop a list of the most commonly ordered test codes in clinical trials. This listing will soon be available using the Clinical Trials subsetting option in the LOINC search tool RELMA (see www.loinc.org). This subset should provide quick and easy identification of the LOINC code for common clinical trial tests. If a local test code cannot be matched in a search of the most commonly ordered tests, a search of the larger LOINC database may be necessary. Removing the Clinical Trials subsetting option will allow the RELMA search tool to access the entire LOINC database. Requests for additions to the most commonly ordered Clinical Trials test listing should be directed to the CDISC LAB Team who will evaluate the test and work with the Regenstrief Institute to add LOINC codes to the Clinical Trials list.

When matching your local codes to LOINC codes, one local code may have several possible LOINC codes. For most laboratory tests the LOINC database contains a methodless code and one or more method specific codes. If a clear match by method is possible, the method specific code is preferred. Conversely, variations in a method may produce several local codes that all match a single LOINC code. In such a case the data transmitter should match all such local codes to the appropriate LOINC code, and then use a separate field in the data transmission to distinguish variants of the base method. If the variations are felt to be significant, a request for new LOINC codes may be submitted to the LOINC committee.

The Regenstrief Institute (Indiana University Medical Center) maintains the LOINC database, supporting documentation and the RELMA mapping program for matching local codes to LOINC. RELMA contains a sub-setting tool that includes a subset of test codes commonly used in drug development clinical trials.

The World Wide Web URL <http://www.loinc.org> provides downloads of all LOINC and RELMA files and documentation.

The NCI Common Toxicity Criteria is available from the National Cancer Institute at <http://ctep.info.nih.gov/CTC3/CTC-Manual.htm>.

3.3.3 Date and Time Format Conventions

Times are clock readings within a 24-hour period. The hour (hh) ranges from 00 to 23. The minutes (mm) and the seconds (ss) range from 00 to 59. The optional fractional part (.nnn) expresses milliseconds. Dates, times, and datetimes are to be interpreted as local clock readings at the place the data was collected. In a datetime, the +hh:mm (or -hh:mm) is the offset in hours and minutes to add to (or subtract from) Coordinated Universal Time to get the local clock reading at the time the data was collected. A special offset of -99:99 means that the relationship between the local clock and Universal Time is not known.

Examples:

- 3:14 pm on 3 January 2001 in Chicago (6 time zones west of Greenwich, standard time) would be represented as "2001-01-03T15:14:00-06:00".
- 3.5 seconds after midnight on the morning of July 20th 2001 in Chicago (daylight time) would be represented as "2001-07-20T00:00:03.500-05:00".

Note: The above formats for dates, times, and datetimes (YYYY-MM-DDThh:mm:ss±hh:mm) are compatible with ISO 8601 except for the use of the -99:99 offset.

In the CDISC Lab SAS data set implementation all date/time fields should be defined as a character variable of length 20, and contain the date, time and UTC offset as specified in the recommended ISO 8601 format. The data recipient can parse the date and/or time from such a text field into a true SAS date and/or time field (which is numeric) and apply the UTC offset as needed for any calculations.

3.4 Using the Lab Model

3.4.1 General

The field lengths given in the model are maximum field lengths.

The default implementation of the LAB Model is bar delimited ASCII. Therefore, for ASCII implementations, the fields should be separated by the vertical bar or pipe character '|

All fields must be present for the SAS and ASCII implementations. This means that all of the fields must be represented in the transmission whether it has a value or not. This first release of the LAB Model, Production Version 1.0, supports ASCII and SAS implementations.

The SAS Variable Names provided on the Data Field and Reference Range spreadsheets for Version 1.0 of the CDISC LAB Model have been matched to the current draft Version 3.0 of the CDISC SDS model.

The LAB Team will release information on XML implementations at a future date.

3.4.2 Good Transmission Practice

Good Transmission Practice data provides information about either the transmission as a whole or a particular record within it but not the study for which the transmission is being made.

Note that the Transmission Source ID identifies the organization that is the source of the transmission but not necessarily the source of the data. For example, in a transmission from a central laboratory to a CRO this would identify the central laboratory as the source of the transmission and in a transmission from a CRO to a biopharmaceutical company this would identify the CRO as the source of the transmission.

Model Version, File Creation Date and Time and Transmission Source ID must always be populated or required.

Variable	Description
Model Version	The version of the CDISC Laboratory Data Interchange Standard model indicating by definition which fields are contained within it.
File Creation Date and Time	The local date and time the data file was created at the central laboratory. This includes a Universal Time Offset plus/minus hours and minutes.
Transmission Source ID	The ID of the organization that is the source of the data transmission.
Transmission Source Name	The Name of the organization that is the source of the data transmission.

3.4.3 Study

Study data describes the study and the type of transmission being made for it. Data for more than one study can be in a data file.

Study ID and Transmission Type must always be populated or required.

Variable	Description
Study ID or Number	The ID of the study.
Study Name	The name of the study.
Transmission Type	This indicates what type of transmission the data transmission is. There are two transmission types: C – Cumulative, I - Incremental.

3.4.4 Site

Site data describes the site.

There are fields for both site and investigator identification to accommodate those situations in which there are multiple investigators at a single site. The Site ID and the Investigator ID may be either the same or different.

Site ID must always be populated or required.

Variable	Description
Site ID or Number	The ID of the site.

3.4.5 Investigator

Investigator data describes the investigator.

There are fields for both site and investigator identification to accommodate those situations in which there are multiple investigators at a single site. The Site ID and the Investigator ID may be either the same or different.

Variable	Description
Investigator ID or Number	The ID of the investigator.
Investigator Name	The name of the investigator.

3.4.6 Subject

Subject data describes the subject. The commonest method of identifying subjects in a study is by using either one or both of the Screen ID and the Subject ID. However since it is recognized that there are cases in which this is not sufficient a third, spare identifier has been provided. It is suggested that, if this field is required, its use be agreed between the data provider and the data recipient.

Screen ID and Subject ID must be populated if values exist for them. One identifier, either screen or subject ID, is required.

The race recorded for the subject should be a race that can be tied to the normal ranges employed.

Variable	Description
Screen ID or Number	The ID of the subject <i>before</i> randomization.
Subject ID or Number	The ID of the subject <i>after</i> randomization.
Spare subject level ID or Number	Spare subject level identifier. (For use with original screen IDs in cases where re-screening with new numbers is allowed, for example).
Subject Initials	The initials of the subject.
Subject Sex	The sex of the subject.
Subject Sex Code List ID	If utilized, the code list identifier and version number for the Subject Sex Code.
Subject Date Of Birth	The date of birth of the subject.
Subject Race	The biological race of the subject.
Subject Race Code List ID	If utilized, the code list identifier and version number for the Subject Race code.

3.4.7 Visit

Visit data describes the visit.

Visit ID and Visit Type must always be populated or required.

Variable	Description
Visit ID or Number	The ID of the visit.
Visit Name	The name of the visit.
Visit Type	The basic type of visit. There are 2 visit types:

Variable	Description
Visit Type Modifier	<p>S - Scheduled</p> <p>U - Unscheduled</p> <p>There are three subtypes of visit which may occur for both Scheduled and Unscheduled visit types:</p> <p>T - Early Termination</p> <p>R - Retest</p> <p>O - Physician Ordered</p> <p>The differentiation of multiple Early Termination and Retest visits can be determined by collection date and time.</p>

3.4.8 Accession

Accession data identifies the specimen collection kit and the laboratory from where it came. The convention that seems to be standard for all laboratories is that one accession number identifies one accession used at one subject visit.

Central Laboratory ID must always be populated or required.

Variable	Description
Central Laboratory ID	The ID of the central laboratory delivering the data.
Central Laboratory Name	The name of the central laboratory delivering the data.
Accession ID or Number	The ID of the accession used at a subject visit.
Last Active Date and Time	The local date and time of the last modification made to the accession record at the central laboratory. This includes a Universal Time Offset plus/minus hours and minutes.

3.4.9 Record Extension Type

Record Extension Type must always be populated or required in SAS or ASCII implementations.

Variable	Description
Record Extension Type	This allows specification of extensions (TBD) from the Base Lab Model definition. For a lab record, this would contain 'BASE'. For a Microbiology data, this would contain 'MICROBIO'.

3.4.10 Base Specimen

Specimen data describes the individual contents of a specimen collection container. In situations where the notion of specimens does not exist, the Specimen ID or Number field would be left blank and the other fields would describe the specimen collection details of the entire kit.

The model does not accommodate capture of information on aliquots of specimens as it is felt that information on and audit trails of aliquots are kept inside the laboratory.

The two Planned Collection fields are for use in situations like PK visits where specimen collection is repeated at specific intervals (e.g., 1 hour post dose, 2 hours post dose, 4 hours post dose and so on). The planned collection date and time should be expressed by time elapsed in the Planned Collection Time Elapsed field (e.g., 000-03-00 for three hours) and time elapsed description in the Planned Collection Time Elapsed Description field (e.g., post dose).

Note: Collection dates and times are at this level as opposed to at the visit level because it is recognized that more than one specimen collection container may be used at one subject visit. Collection data therefore describes the container and its use and not the visit.

The two Collection End fields are for use in situations like timed urine collections where the duration over which specimens are collected is an important piece of information. The duration would be the time elapsed between the *actual* collection start date and time and the collection *end* date and time.

Note: Fasting Status is at this level as opposed to the visit level in order to allow the model to handle those situations such as repeated collections for glucose tolerance testing where fasting status may change within the same visit.

Note that Subject Age occurs at this level because what is important is not the age of the subject when the study began but the age of the subject *at each time point*.

Actual Collection Date and Time must always be populated or required.

If more than one comment is available for Lab Specimen Comments or Investigator Specimen Comments, they should be separated by the caret character '^'.

Since it is recognized that preferred ways of representing Subject Age may differ it can be expressed in either integer or decimal terms so that, for example, an age of 4 years and 6 months could be expressed either as 4.5 Y or 54 M.

Fasting Status must be populated if it is relevant to the study. If it is not this field will be left blank.

Subject Age Units must be populated if Subject Age at Collection is populated.

Variable	Description
Specimen ID or Number	The ID of an individual kit item used at a subject visit.
Actual Collection Date and Time	The Date and Time of actual specimen collection at the site.
Planned Collection Time Elapsed	The planned collection time elapsed since the dose or other key event, e.g., 000-03-00 for 3 hours. Format: DDD-HH-MM (Days, Hours, Minutes.)
Planned Collection Time Elapsed Description	The description of the planned collection time elapsed since a key event. An example description could be "post dose" or "2 hours post dose" per preference.
Collection End Date and Time	The collection end date and time of a timed specimen collection.

Variable	Description
Received Date and Time	The local date and time of specimen receipt at the central laboratory. This includes a Universal Time Offset plus/minus hours and minutes.
Specimen Condition	Free or standardized text describing the condition of the specimen e.g., Hemolyzed, Icteric, Lipemic, etc.
Lab - Specimen Comments	Comments in addition to the specimen condition from the central or performing laboratory describing the specimen.
Investigator - Specimen Comments	Comments in addition to the specimen condition from the investigator describing the specimen.
Specimen Material ID	The ID of the specimen material.
Specimen Material Code List ID	If utilized, the code list identifier and version number for the Specimen Material ID code.
Specimen Material Name	The name of the specimen material (e.g., "Blood", "Urine", etc.).
Subject Age at Collection	The age of the subject at the collection visit. This is the age that will be used to determine which age-stratified reference ranges (if any) should be applied to the subject's test results.
Subject Age Units	The units in which the age of the subject is measured. There are three subject age units: Y – Years; M – Months; D - Days
Fasting Status	The fasting status of the subject for the specimen collected in the specimen. There are four fasting statuses: Y - Yes N - No U - Unknown and blank if not relevant.

3.4.11 Base Battery

Battery data describes the battery, panel or group of tests performed upon the specimens. The Battery field is required. If a Battery, Panel or Group ID does not exist, use the Test ID from the next level.

Variable	Description
Battery ID	The ID of the battery or panel to which the test belongs. If a Battery ID does not exist, use the Test ID from the next level.
Battery Name	The name of battery or panel to which the test belongs.

3.4.12 Base Test

Test data identifies the test and the laboratory that performed it.

The test identification fields are available for both the data recipient and data provider test. This is done to accommodate those situations where a common or standardized test code is not available or agreed upon between the data recipient or data provider.

Note that the Performing Laboratory ID identifies the laboratory that actually performed the test and that is also therefore the source of the *data*. This might be either a reference laboratory or the central laboratory.

Performing Laboratory ID, Lab Test ID and Test Status must always be populated or required.

Test ID must be populated if the data recipient requests that the test codes of the data provider be mapped using the recipient's preferred code list. If the recipient does not request test code mapping this field will be left blank. In addition, separate fields are provided for LOINC codes. (See Section 3.3.2.)

If more than one comment is available for Test Level Comments, they should be separated by the caret character '^'.

Variable	Description
Performing Laboratory ID	The ID of the laboratory that performed the test.
Performing Laboratory Name	The name of the laboratory that performed the test.
Lab Test ID	The ID of the test performed as defined by the data provider.
Lab Test Name	The name of the test performed as defined by the data provider.
Test ID	The ID of the test performed as defined by the data recipient.
Test Name ID	The name of the test performed as defined by the data recipient.
LOINC Code	The LOINC code ID for the test performed.
LOINC Code List ID	If utilized, the code list identifier and version number for the LOINC code.
Additional Test Description	Additional Test Description to characterize additional aspects associated with the test. This information may be important for interpretation for the study. For example: instrument model number, lot numbers, etc...
Test Status	This indicates what the status of the test is. There are three test statuses: D - Done N - Not Performed X - Cancelled.
Test Level Comments	Free or standardized texts of all laboratory-generated comments relating to the specific test.
Testing Date and Time	The local date and time of testing. This includes a Universal Time Offset plus/minus hours and minutes.
Test Type	This indicates what the type of the test performed is. There are three test types: S - Study test (as defined in the protocol)

Variable	Description
	N - Non-study test (not defined in the protocol) U - Unscheduled study test (test defined in the protocol but not scheduled for the current visit)

3.4.13 Base Result

Result data describes the results of the tests.

It is recognized that results may be preferred in more than one unit system so Reported, US Conventional and SI result fields have been provided. The Reported results are the results reported to the investigator sites.

When a sponsor defines the units in which results are to be reported to investigator sites and/or in the data transfer to sponsor, the appropriate units sections should be populated or required. The CDISC model offers flexibility in enabling the sender and sponsor to differentiate between the results/units that were reported to the site in the reported results section and the results/units required for the data transfer. For instance, a sponsor may require that all results/units are sent in the data transfer in conventional units but at the same time be interested in the results/units that were reported to the site (could vary depending on geographic region). In this case, the sender would populate the reported and conventional result blocks. If in the same scenario it were required that certain tests be reported in SI units, the SI block for those tests would be populated and the conventional block left blank. Or the sponsor could opt (and agree with the sender) to have all three-result blocks populated and later determine in the receiving system which units should be used for each specific test.

In each of these three cases (Reported, US Conventional and SI), 5 pieces of information are repeated: Text Result, Numeric Result, Numeric Result Precision, Reference Range Low, Reference Range High and Units.

Reported Text Result must be populated if the test was done unless the test result is blinded to the recipient. If a valid test result exists, it must be entered here. Therefore, the field must contain a non-null value except in the case that the test was cancelled or not performed. If a result is numeric, the Reported Numeric Result fields must be populated AND the Reported Text Result field must be populated with the numeric value formatted to the specified precision.

The LAB Team recommends consideration of coded text results if very long text result strings can be converted to coded results. When a coded value is used for the test result, the Result Code List Identifier field should be used to specify the code list (and version of that list) used.

If a result is textual, the Text Result fields must be populated and the Numeric Result fields are left blank.

For reported results that are either below or above a quantifiable limit (such as or <10000 or >50000 for example), use the G and L options from the Reported Result Type field and put the entire result value in the Reported Text Result field. The Reported Numeric Result would then be empty.

For reported results that are expressed as either lists or ranges (such as 3,5 or 3-5 for example) use the R option from the Reported Result Type field and put the entire value in the Reported Text Result field.

Reported Units must be populated if a value exists for it.

Reported Result Type must be populated unless the test is cancelled OR if the test result is blinded to the recipient. Note: ‘coded’ results are results from a code list agreed between the data provider and the data recipient. They should be put in the Text Result fields only—even if they are numeric.

Transaction Type must be populated to reflect the type of transaction for the record.

Variable	Description
Reported Text Result	Reported text result by laboratory.
Reported Text Result Code List ID	If utilized, the code list identifier and version number for the Reported Text Result code.
Reported Numeric Result	Reported numeric result by laboratory.
Reported Numeric Result Precision	Reported numeric result precision at laboratory. Example: 5,3 - representing the total size of 5 with 3 positions to right of the decimal. (12.345)
Reported Reference Range Low	Reported lower limit of reference range.
Reported Reference Range High	Reported upper limit of reference range.
Reported Units	Reported result units by laboratory.
Reported Units Code List ID	If utilized, the code list identifier and version number for the Reported Units code.

Conventional Text Result	Conventional text result at laboratory.
Conventional Text Result Code List ID	If utilized, the code list identifier and version number for the Conventional Text Result code.
Conventional Numeric Result	Conventional numeric result at laboratory.
Conventional Numeric Result Precision	Conventional numeric result precision at laboratory. Example: 5,3 - representing the total size of 5 with 3 positions to right of the decimal. (12.345)
Conventional Reference Range Low	Conventional lower limit of reference range.
Conventional Reference Range High	Conventional upper limit of reference range.
Conventional Units	Conventional result units at laboratory.
Conventional Units Code List ID	If utilized, the code list identifier and version number for the Conventional Units code.

SI Text Result	SI text result at laboratory.
----------------	-------------------------------

Variable	Description
SI Text Result Code List ID	If utilized, the code list identifier and version number for the SI Text Result code.
SI Numeric Result	SI numeric result at laboratory.
SI Numeric Result Precision	SI numeric result precision at laboratory. Example: 5,3 - representing the total size of 5 with 3 positions to right of the decimal. (12.345)
SI Reference Range Low	SI lower limit of reference range.
SI Reference Range High	SI upper limit of reference range.
SI Units	SI result units at laboratory.
SI Units Code List ID	If utilized, the code list identifier and version number for the SI Units code.

Reported Result Type	This indicates what the type of the reported result is if the test was done. There are 7 reported result types: C - Coded N - Numeric T - Text G - Greater than (quantifiable limit) L - Less than (quantifiable limit) R - Range (result) Blank (if the test was cancelled)
Reported Result Status	This indicates what the status of the reported result is and is provided for Quality Control purposes. There are two reported result statuses: P - Preliminary F - Final.
Alert Flag	The alert flag generated by the reference ranges applied and tied to the reported result. There are 9 alert flags: LP - Low Panic LT - Low Telephone LN - Low Normal N - Normal HN - High Normal HT - High Telephone HP - High Panic AB - Abnormal and blank when not used.
Delta Flag	The delta flag generated by the reference ranges applied and tied to the reported result. There are three delta flags: D+ for an increase in value D- for a decrease in value

Variable	Description
	and blank for no flag.
Toxicity Grade	The toxicity grade generated by the toxicity ranges applied and tied to the reported result.
Toxicity Grade Code List ID	If utilized, the code list identifier and version number for the Toxicity Grade code.
Exclusion Flag	The exclusion flag generated by the range applied and tied to the reported result. There are four exclusion flags: LX - Low Exclusion HX - High Exclusion EX - Excluded (for exclusions not falling under high or low) and blank for no flag.
Blinding Flag	This indicates the blinding status: S - Blinded to Sponsor, I - Blinded to Investigator, B - Blinded to Sponsor and Investigator C - Custom Blinding or blank
Reported Date and Time	The local date and time at which the result was reported to the investigator site. This includes a Universal Time Offset plus/minus hours and minutes.
Transaction Type	This indicates what type of record the data record is and consequently how it should be processed when it is imported into the study database. There are 4 transaction types: M - Remove (existing record) I - Insert (new record) R - Retransmit (existing record without changes) U - Update (or revise existing record at result record)

Within a flat file (ASCII or SAS) format, the Transaction Type applies to the record as a whole. When removing a high level entity such as a subject, every result associated with that subject should be sent as a record with a Transaction Type of “Remove”.

Within hierarchical XML, the CDISC lab model contains a Transaction Type element at each of its major levels. It is thus possible to update, for example, a subject’s initials without specifying any data below the subject level. When removing a high level entity such as a subject, the model is flexible. It allows the fully specified approach of listing all results with the Transaction Type at the Base Result level being set to “Remove”. It also supports the hierarchically efficient approach of setting the Transaction Type at the Subject level to “Remove”, implying a cascading removal of all data that belongs to that subject. When using a hierarchical XML format, the sender and receiver should define which approach will be used with the Transaction Type within the Data Transmission Agreement.

The CDISC Lab team recognizes that data recipients may wish to confirm that a file has not been corrupted or truncated during electronic transmission. Various technologies (such as check sums) exist to perform this task and do not require fields or rows be added to the data model for this purpose. The CDISC Lab team recommends the use of these technologies, and has not specified any End of File markers or row counters in its flat file implementations.

3.5 Comments on Practical Application

3.5.1 Use of Model Fields

The model is designed to support the requirements of many different organizations within the industry and many different types of clinical trials. Accordingly, while many of the fields it contains apply to all clinical trials there are some, which do not because they have special uses. This means that for any given study it should be expected that, while most of the fields certainly will be used, some certainly would not be.

3.5.2 Data Compression

For the purposes of physically transferring and storing data files made using the model, tests have shown that typically compression rates of approximately 95% can be achieved.

3.5.3 Transmission Agreements

The CDISC LAB Model is applicable for use for data interchanges between sender and receiver but is not a “plug-in” that covers all data requirements. Implementation of the LAB Model requires adoption of a TRANSMISSION AGREEMENT between sender and receiver.

A transmission agreement would cover the following types of information:

- Implementation type (e.g., bar delimited ASCII, SAS, XML DTD, XML schema)
- Use of standardized text fields
- Code lists
- Local test codes
- Result codes
- Interpretation of reference ranges (e.g., inclusive or exclusive of upper and lower limits)
- How to represent subject age (e.g., years, months or days)
- In the XML implementations of the Lab model, the Transaction Type field is found at several levels (e.g. Investigator, Subject, Visit, etc.) and not just at the Result level. Within the Data Transmission Agreement, the sender and receiver should define how the "Remove" option is to be used. If a full audit trail of all affected values is desired, the "Remove" option should be allowed only on the Result level. If the "Remove" option is allowed at any other level, it carries the implication that all data below that level is also to be removed. Thus, if a "Remove" occurs at the Subject level, all visits, accessions, samples, panels, tests and results for that subject are implicitly removed.
- Mechanism for updating and deleting records other than result records

3.6 Testing of the Lab Model

In order to gauge the usefulness of the model that it is developing, the LAB team conducted tests amongst its own members by partnering central laboratories and pharmaceutical companies together and having them interchange real clinical trial laboratory data. Testing was conducted at more than three points in the development of the Model and each testing cycle consisted of at least three pharmaceutical company-central laboratory pairs. Of particular interest was the feedback from the central laboratories regarding the use of the model to create data files and from the pharmaceutical companies regarding the use of the model to process those data files.

The participants unanimously reported that the model is very easy to use because the structure of the different levels of laboratory data within it is clear and logical.

From the central laboratory perspective, this made the population of the data fields straightforward and unambiguous because there were clear relationships between what the model requires and where the data resides in clinical data management systems. The ease of use of the model both in terms of understanding what fields are required and how those fields should be populated made the implementation of programs to create data files a very straightforward process with no requirements for complex programming. Most notably this made significant savings in terms of time and resulted in very fast implementations with no sacrifice in quality.

The most significant advantage to the central laboratories of the broad acceptance of the model would be that once the initial development work has been done, additional development work would not be required. A single program could be reused and the only additional work necessary would be the relatively minor updates needed each time to support the differences that are always to be expected with different studies such as visit sequences and testing requirements. Far fewer resources therefore would be required for each new study so costs would be reduced and more resources could be spent on ensuring the quality of the data interchange.

The pharmaceutical companies reported that the data was easily extracted from the files and that the content of the model was certainly adequate for their requirements. It was specifically noted that the logical organization of the data and the ease of its extraction would allow very straightforward translations into existing technical infrastructures and applications.

The most significant advantage of the model to the pharmaceutical companies is that it would allow them to streamline the acquisition and processing of data since having a single program to handle all files would make data import time and effort trivial.

All the participants agreed that no matter how intuitive and easy to use a model may be good supporting documentation is essential in order to avoid any possibility of misunderstanding. For the purposes of industry review of the model the preceding section (section 3.4 – ‘Using the Model’) should be used as that documentation.

3.7 Use Case Scenarios

A biopharmaceutical company is conducting a clinical trial of a new drug XYZ. The company has hired a central laboratory to perform the specimen testing and a contract research organization (CRO) to collect, clean and compile the study data.

The biopharmaceutical company, the CRO and the central laboratory all have different database management systems (DBMS) either proprietary or commercial. Each DBMS is designed and structured differently and consequently each manages data in a different manner.

The central laboratory needs to transfer laboratory data to the CRO and the CRO needs to transfer data to the biopharmaceutical company. Despite the differences inherent in their separate database management systems the transfers will go smoothly because all three will use the CDISC Laboratory Data Interchange Standard.

Before the advent of this standard the central laboratory, the CRO and the biopharmaceutical company would have spent up to 6 or 8 weeks working to set up a new data transfer. Each data provider would have spent time trying to accommodate the requirements of their client's specifications and each data recipient would have spent time trying to make allowances for any requirements that their suppliers could not in fact accommodate. Even after all that time they may still not have been able to resolve all the inherent differences in their separate database management systems leaving additional work to be done with each data transmission.

Now, however, each has been able to more efficiently dedicate their resources to being able to fully comply with the single CDISC Laboratory Data Interchange Standard. From that basis, incorporating new data transfers for new clinical trials into their systems takes much less time, requires far fewer resources and meets all the requirements of the data transmission specifications.

4 Review Process

After internal testing and revision, the LAB Model was sent to a 65 member Lab Review Committee. The Review Committee consisted of industry experts who had expressed an interest in the work of the LAB Team. The LAB Team considered comments received, responded to them and made appropriate revisions to the model. The Model was then posted for public comment on the CDISC website. Comments received from the public posting were also considered, responded to and appropriate revisions made to the model. The result is Production Version 1.0 of the CDISC LAB Model.

All CDISC models follow the CDISC Standards Development Process, which includes multiple levels of review by internal CDISC teams, focused external groups, and the public. For more details of this process, see the CDISC Standards Development Process document (CDISC-COP-001).

5 Next Steps

The CDISC LAB Team has posted a draft version of an XML Schema for the LAB model as well as a draft microbiology extension. Comments on these draft versions are welcome. The Team will next work on an extension to handle pharmacogenomics data.

The LAB Team is also very interested in working with implementers of the LAB Model to collect information and metrics from the implementation process. An Implementation Case Study Document template is now posted on the website and implementers are encouraged to work with the LAB Team in collecting data. After collection and evaluation of approximately six case studies (and periodically thereafter), metrics and other information on implementation will be posted to help guide users and new implementers. The implementation information will be posted in the CDISC Members-Only area and will be available to CDISC members as well as those who contribute implementation information. It is hoped that posting in the Members Only area of the website will encourage companies to become CDISC members. The LAB model itself as well as all supporting documentation is available at no cost to all interested parties, members and non-members.

6 Directions for Submitting Comments

Comments about this model can be submitted by using the Public [Discussion Forum on the CDISC website](#).

We invite those interested in the model, implementing the model for the first time or using the model in production to post their comments, questions or suggestions. We hope to create an active forum used to foster discussion, to provide assistance and to share experiences. Posted comments will be reviewed on an ongoing basis by members of the LAB team. The discussions will not only assist model users but will guide the LAB team in updating the model as needed and in creating training materials and educational programs.

7 APPENDICES

7.1 APPENDIX A: Synonyms

Different organizations within the industry sometimes use different terms for the same things. Below is a table of the specialist terms that have been used within this document and some examples of what they are also commonly known as.

Term	Also known as...
Incremental (of a data transmission)	Transactional
Study	Protocol
Site	Center
Subject	Patient
Unrandomized subject	Zero subject
Sex	Gender
Visit	Subject Event or Assessment
Specimen	Sample
Test	Analysis, Assay or Observation
Text (data type)	Alphanumeric, Character or Subjective
24 hour clock	Military Time
Early Termination	Early Discontinuation or Premature Withdrawal
Quantifiable Limit	Limit of Quantification
Mapping	Substitution or Translation

7.2 APPENDIX B: Example Specimen Condition List

‘Cells Present’
‘Clotted’
‘Hemolysis, Moderate’
‘Hemolysis, Severe’
‘Hemolysis, Slight’
‘Hemolysis’
‘Hemolyzed, Partially’
‘Icteric, Partially’
‘Incorrect Sample’
‘Inhibiting Substance’
‘Lab Accident’
‘Lipemia, Moderate’
‘Lipemia, Severe’
‘Lipemia, Slight’
‘Lipemia’
‘Lipemic, Partially’
‘Quantity Not Sufficient’
‘Questionable Result’
‘Serum Separation Inadequate’
‘Test Tube Broken’
‘Thawed’
‘Unspun, Partially’
‘Unspun’

7.3 APPENDIX C: Index

A

Accession ID or Number · *See* Accession
Actual Collection Date and Time · *See* Base Specimen
Additional Test Description · *See* Base Test
Alert Flag · *See* Base Result
Alphanumeric, Character or Subjective · *See* .
Synonyms
Analysis, Assay or Observation · *See* . Synonyms

B

Battery ID · *See* Base Battery
Battery Name · *See* Base Battery
Blinding Flag · *See* Base Result

C

Center · *See* . Synonyms
Central Laboratory ID · *See* Accession
Central Laboratory Name · *See* Accession
Collection End Date and Time · *See* Base Specimen
Conventional Numeric Result · *See* Base Result
Conventional Numeric Result Precision · *See* Base Result
Conventional Reference Range Low · *See* Base Result
Conventional Text Result · *See* Base Result
Conventional Text Result Code List ID · *See* Base Result
Conventional Units · *See* Base Result
Conventional Units Code List ID · *See* Base Result

D

Delta Flag · *See* Base Result

E

Early Discontinuation or Premature Withdrawal ·
See .Synonyms
Exclusion Flag · *See* Base Result

F

Fasting Status · *See* Base Specimen
File Creation Date and Time · *See* Good Transmission Practice

G

Gender · *See* . Synonyms

I

Investigator - Specimen Comments · *See* Base Specimen
Investigator ID or Number · *See* Investigator
Investigator Name · *See* Investigator

L

Lab - Specimen Comments · *See* Base Specimen
Lab Test ID · *See* Base Test
Lab Test Name · *See* Base Test
Last Active Date and Time · *See* Accession
Limit of Quantification · *See* . Synonyms
LOINC Code · *See* Base Test. *See* Code Lists
LOINC Code List ID · *See* Base Test

M

Military Time · *See* . Synonyms
Model Version · *See* Good Transmission Practice

P

Patient · *See* . Synonyms
Performing Laboratory ID · *See* Base Test
Performing Laboratory Name · *See* Base Test
Planned Collection Time Elapsed · *See* Base Specimen
Planned Collection Time Elapsed Description ·
See Base Specimen
Protocol · *See* . Synonyms

R

Received Date and Time · *See* Base Specimen

Record Extension Type · *See* Record Extension Type
Reference Range High · *See* Base Result
Reported Date and Time · *See* Base Result
Reported Numeric Result · *See* Base Result
Reported Numeric Result Precision · *See* Base Result
Reported Reference Range High · *See* Base Result
Reported Reference Range Low · *See* Base Result
Reported Result Status · *See* Base Result
Reported Result Type · *See* Base Result
Reported Text Result · *See* Base Result
Reported Text Result Code List ID · *See* Base Result
Reported Units · *See* Base Result
Reported Units Code List ID · *See* Base Result

S

Sample · *See* .Synonyms
Screen ID or Number · *See* Subject
SI Numeric Result · *See* Base Result
SI Numeric Result Precision · *See* Base Result
SI Reference Range High · *See* Base Result
SI Reference Range Low · *See* Base Result
SI Text Result · *See* Base Result
SI Text Result Code List ID · *See* Base Result
SI Units · *See* Base Result
SI Units Code List ID · *See* Base Result
Site ID or Number · *See* Site
Spare subject level ID or Number · *See* Subject
Specimen Condition · *See* Base Specimen
Specimen ID or Number · *See* Base Specimen
Specimen Material Code List ID · *See* Base Specimen
Specimen Material ID · *See* Base Specimen
Specimen Material ID and Name · *See* Code Lists
Specimen Material Name · *See* Base Specimen
Study ID or Number · *See* Study
Study Name · *See* Study
Subject Age at Collection · *See* Base Specimen
Subject Age Units · *See* Base Specimen

Subject Date Of Birth · *See* Subject
Subject Event or Assessment · *See* . Synonyms
Subject ID or Number · *See* Subject
Subject Initials · *See* Subject
Subject Race · *See* Subject
Subject Race Code List ID · *See* Subject
Subject Sex · *See* Subject
Subject Sex Code List · *See* Subject
Substitution or Translation · *See* . Synonyms

T

Test ID · *See* Base Test
Test Level Comments · *See* Base Test
Test Name · *See* Base Test
Test Status · *See* Base Test
Test Type · *See* Base Test
Testing Date and Time · *See* Base Test
Toxicity Grade · *See* Base Result. *See* Code Lists
Toxicity Grade Code List ID · *See* Base Result
Transaction Type · *See* Base Result
Transactional · *See* . Synonyms
Transmission Source ID · *See* Good Transmission Practice
Transmission Source Name · *See* Good Transmission Practice
Transmission Type · *See* Study

U

Units (Reported, Conventional, SI) · *See* Code Lists

V

Visit ID or Number · *See* Visit
Visit Name · *See* Visit
Visit Type · *See* Visit

Z

Zero subject · *See* . Synonyms