

Sol LeWitt and the CDISC Initiative: A Comparative Essay

Karin LaPann, ViroPharma Inc.

Introduction

Just who was Sol LeWitt, you may ask. Well in the art world, he is a renowned contemporary artist (1928- 2007), who had tremendous influence on minimalism and conceptual art in the 1960s and beyond. His works have been presented in installations around the world and are self-documented. In fact, his writing, photography and instructions which convey the idea behind the work, are just as important as the final product and often presented together. The importance of the ideas is a central tenet of conceptual art, and the use of simple elements is a central tenet of minimalism. His famous wall drawings are generally massive and yet evolved from very simple forms. He started with straight lines, and then added curved lines, squiggly lines and complex shading schemes, gradually evolving from black and white color scheme to primary colors.

Instead of creating a single wall mural, Sol LeWitt created instructions on how to create what he called “wall drawings” so that they would be reproducible and could be created by a collaborative team of skilled artists or fabricators. He fully explored in his works the relationship between the idea and the final product, between outlining the concept and fully defining the work. In one of his essays he says “The system is the work of art; the visual work of art is the proof of the system. The visual aspect can't be understood without understanding the system. It isn't what it looks like but what it is that is of basic importance”.¹

You might wonder, what does this have to do with the Clinical Data Interchange Standards Consortium (CDISC) and one of the CDISC standards: Study Data Tabulation Model (SDTM) standards? For those not in this industry here is a brief introduction. The pharmaceutical industry has developed a set of standards through a non-profit group called CDISC. Guidelines such as SDTM are basically a set of instructions on how to collect and present data across pharmaceutical industries so that they can be shared. The FDA who is the primary customer of the research created by pharmaceutical companies, is accepting submissions using these standards as part of the submission requirements. Below are elements generally requested by the FDA:

1. Raw datasets, preferably in SDTM specifications
2. Patient Profiles – only if no SDTM datasets sent
3. Analysis files

In this paper I will draw parallels on the relationship among the invention, execution and utilization of the LeWitt wall drawings, compared to the invention of SDTM standards, execution by individual companies of these standards, and the utilization of the final product from the research, which is a successful submission of data to the FDA.

¹ From Sol LeWitt's writings, as quoted in www.brainyquote.com/quotes/s/sollewitt355012.html

I will explore the relationship between the increasing complexity of LeWitt's work over time and the increasing amount of instructions in the data standards used to clearly define the datasets for collection of study data. I will also emphasize the fact that the perceived simplicity of the end product is actually highly complex and choreographed. I will also provide sources and resources to use these increasingly specialized instructions.

I will suggest that the SAS programmer is similar to the trained artist in the LeWitt installations; that we need to have some skilled people available in order to carry the standards out correctly. Then I will point out where to find the instructions and how the end products can be created.

Background on Sol LeWitt

One of Sol LeWitt's landmark contributions was to separate the act of conceiving a work of art from the act of executing it. His famous quote, "The idea becomes the machine that makes the art" summarizes this concept succinctly². His first wall drawing was executed in 1968 at the Paula Cooper Gallery in New York City (wall drawing #11) and consisted of four squares, each with lines meticulously drawn in the four directions: left to right, up and down, diagonal in each direction. He was exploring a system of rules and the various permutations of the system. He developed written instructions for these wall drawings for his studio assistants to execute. The instructions stated to use a contractor's blue chalk snap line, which is a very simplistic and commonly used tool. These instructions were written in lexicon that he adopted and which was easily understandable to the trained artist.

Significantly, LeWitt explored the inherent complexity of simplicity by focusing on the simplest of elements: the line. He began experimenting with primary colors of yellow, blue and red, which when overlapped would also include the secondary colors of purple, green and orange. LeWitt continued the exploration, adding new ways in which the art could be defined, such as wall drawing #766 below.



Wall Drawing 766

Twenty-one isometric cubes of varying sizes, each with color ink washes superimposed. September 1994

Color ink wash

San Francisco Museum of Modern Art, Accessions Committee Fund: gift of Barbara and Gerson Bakar, Emily Carroll and Thomas Weisel, Jean and James E. Douglas, Jr., Evelyn D. Haas, the Modern Art Council, Phyllis and Stuart G. Moldaw, Robin Wright Moll, Norah and Norman Stone, Danielle and Brooks Walker Jr., and

Judy and John Webb. 2000.301 ,FIRST INSTALLATION

² Sol LeWitt A Wall Drawing Retrospective Gallery Guide, Mass MoCA

As LeWitt's art evolved, he added more colors as well as textures such as shiny versus matte paint, and structures called "Complex Forms" that look like modern day three-dimensional graphs. Also, the vocabulary to create these works had to evolve, as the drawings became increasingly complex in color choices and application. LeWitt kept these instructions very standard, developing easy-to-understand codes for color combinations, lines, and other elements. Throughout his entire career, he continued to explore his basic concept, to use writing (both captions and instructions) to make that concept accessible to the viewers, and valuing the instructions and idea over the art object itself.

The artist and his assistants had to become more specialized in specific drafting and painting techniques to create the desired effect (e.g. , masking with blue painter's tape to achieve clear separation between bands of color, using a contractor's blue chalk snap line tool to execute lines, and sanding and preparing walls with chemicals to make them more receptive to the paint). The designs continued to grow in detail as exemplified by some of his final works "Scribbles", where he returned to the graceful graphite pencil drawings using scribble lines in different spacing to create beautiful geometric forms with shading (wall drawing #1247 August 2007, #1260 July 2008, #1261 July 2008). A complete Sol LeWitt wall drawing retrospective has been installed at the Massachusetts Museum of Contemporary Art (Mass MoCA), and will be on display for 25 years, until 2033.

What is CDISC

The Clinical Data Interchange Standards Consortium, also known as CDISC, was founded in late 1997 as a volunteer group. In February 2000 it was formed as an independent, non-profit organization. Finally in December 2001 it became a global organization. Its members are primarily large and mid-sized pharmaceuticals around the world, clinical research organizations that work with pharmaceutical companies, and clinical review boards (such as the FDA). The purpose of the consortium is to standardize the means to collect, store and disseminate clinical study data to better communicate effectively. The primary market drive was the need for uniform standards to help decrease drug development time and cost. CDISC standards are vendor-neutral, platform-independent and freely available at the CDISC website.

In December 2001 the FDA became interested in standardized submissions to review clinical data. They commissioned and piloted a software application called the Patient Profile Viewer which would enable them to use the SDTM formatted data to create and view patient profiles. Then on October 20, 2003 the FDA endorsed the CDISC standards development. At this time they also began mentioning the preferred transfer method to be HL7 version 3. Once the FDA became involved the CDISC membership seems to have increased dramatically. The 10th Anniversary CDISC International Interchange held in Bethesda Maryland in 2007 was very well attended, as was the 4th CDISC European Interchange held in Switzerland of the same year ³.

³ www.cdisc.org/archives.html/

Among many of the initiatives sponsored by CDISC are:

- SDTM - Study Data Tabulation Model
- CDASH – Clinical Data Acquisition Standards Harmonization project
- ADaM - Analysis Data Model
- ODM – Operational Data Model

SDTM is the most developed of the standardization efforts, currently at version 3.1.2, although the FDA is still currently using 3.1.1. Again, this was developed by general consensus among teams of members comprised of pharmaceuticals, the government, and private technology firms. These are the standards that allow the FDA to browse for patient profiles using the Patient Profile Viewer. The set of instructions that comprise these standards are clear and scalable. Currently the SDTM datasets are mentioned in the FDA “Guidance” documents, soon to become a requirement.

CDASH is relatively new as version 1.0 was released in October 2008. These standards offer a standard for collection of the safety data for a clinical study. This data is collected using case report forms (CRF's) which are paper versions that are then entered in database systems, or electronically (eCRF).

The analysis data model (ADaM) is now on version 2.0. It is not yet required by the FDA as it is still going through significant restructuring. The FDA accepts the analysis datasets from pharmaceutical companies in their preferred structure, and often they are not required with submissions, but requested during the review process.

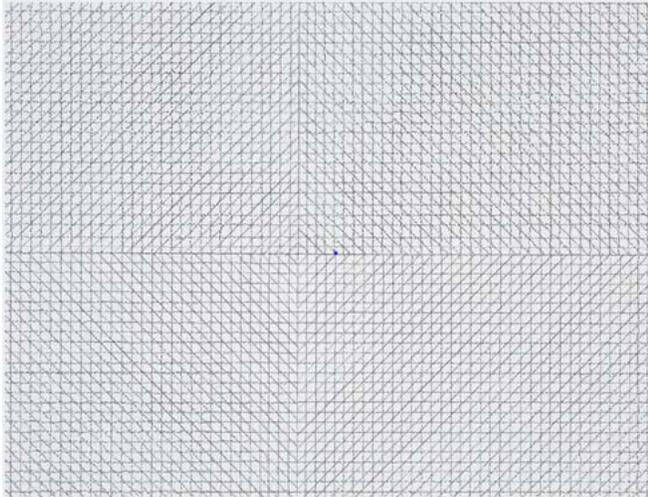
ODM is the standard that expresses the requirements for systems that read or write datasets. It includes requirements for the XML structure for such files as the Define.xml file that gives a summary of datasets being submitted⁴. ODM is currently in version 1.3.

Other activities by CDISC include harmonization efforts with other standardization groups in the United States and Europe.

First Steps

The first few works that Sol LeWitt created utilized only straight lines, which were drawn in 8H or 9H graphite sticks. The classic four squares with lines in different directions had four squares (xx by xx) with horizontal lines, vertical lines, and diagonal lines to the left and diagonal lines to the right. This simple concept explores the four basic directions that these lines could travel.

⁴ www.cdisc.org/models/odm/v1.3/final/ODM1-3-0-Final.html



Wall Drawing 11

A wall divided horizontally and vertically into four equal parts. Within each part, three of the four kinds of lines are superimposed.

May 1969

Black pencil

Yale University Art Gallery

Gift of Anna Marie and Robert F. Shapiro, B.A.

1956

First Installation

Paula Cooper Gallery, New York

First Drawn By

Jerry Orter, Adrian Piper, Sol LeWitt

MASS MoCA Building 7

Ground Floor

In the same light, four basic domains can be identified below. Domains are defined as datasets of specific definitions that share key variables and then have additional carefully selected variables. The domain name is a unique 2 letter code and used in four ways:

1. The dataset name
2. The value of the DOMAIN variable in the dataset
3. The prefix of most variables other than keys in the dataset
4. The value for RDOMAIN in relationship tables

Domains:

DM - Demographics

EX – Exposure

DS – Disposition

AE – Adverse Events

Next Steps

The LeWitt art as discussed before, evolved over time, but LeWitt still kept the basic simplicity of its instructions by developing a unique vocabulary and standardized phrases. Thus blue, yellow and red were only referred to as (B) (Y) and (R). Combinations of shadings were described by sequences of initials. New paints such as acrylics were added, and varnish to make shiny next to matte contrasts. The designs were described by both words and minimal hand drawn sketches.

Artists were trained in the taping and drafting techniques required creating the works of art at various locations around the world. These were also basic, but some training was required so that they would be carried out similarly at each installation. By using these simple instructions the works could now be duplicated around the world. The instructions became the art and are highly valued. Below is an example of one of his more elaborate works.



Wall Drawing #692, 2009
 Continuous forms with color ink washes superimposed.
 October 1991
 Color ink wash
 Collection of Martin E. Zimmerman of the LINC Group, Chicago
 FIRST INSTALLATION
 LINC Group, Chicago
 FIRST DRAWN BY
 Rebecca Schwab
 MASS MOCA BUILDING 7
 SECOND FLOOR

Finally, here is a 3 dimensional sculpture which expands the flat painting into a vertical shape extending from the ground.



Splotch #22, 2007
 Acrylic on fiberglass
 ca. 148 x 96 x 86 inches
 Virginia Museum of Fine Arts, Richmond
 The Sydney and Frances Lewis Endowment Fund, and partial gift of the artist and Pace Wildenstein in honor of Frances Lewis and in memory of Sydney Lewis
 © Virginia Museum of Fine Arts and courtesy of the estate of Sol LeWitt

Before his death, Sol LeWitt toured the Mass MoCA facility in preparation of the installation. He made decisions on where to place the various pieces, and also on keeping the exterior walls of several buildings showing, and keeping the natural light and views as part of the art experience.

In the same manner, the most mature CDISC model, SDTM can be compared by its seeming simplicity of instructions. The instructions are detailed in two documents. The first is the Study Data Tabulation Model, currently at version 1.2, which describes the model fundamentals. It gives the general framework for describing the information collected and to be submitted to regulatory agencies. The second is the Study Data Tabulation Model Implementation Guide for Human Clinical Trials (SDTMIG) which is currently at version 3.1.2 (although previous versions are available on the website). Although it is almost 300 pages long, it has repetitive instructions for each individual domain, or dataset name. Thus, each domain could be compared to one variation of LeWitt's numbered wall drawings. The whole document can be compared to an artist's catalog of his work.

Following are some specialized terminology to describe the model. Each observation corresponds to a row in a dataset or spreadsheet. The variables correspond to columns in the dataset. Although collection and extraction of the data are software-independent, this type of dataset structure is ideal for using SAS® because SAS stores its data in flat files with rows and columns. Also, in SAS® documentation we often find the terms variables and observations. Datasets are defined by classes, which identify their purpose. The datasets themselves are called domains. A domain is a collection of logically related observations within a topic.

Therefore, as identified in the SDTMIG version 3.1.2, the special-purpose class includes the domains Demographics (DM), Comments (CO) Subject Elements (SE) and Subject Visits (SV).

The Interventions class includes Concomitant Medications (CM), Exposure (EX) and Substance Use (SU). Interventions include such topics as investigational, therapeutic and other treatments administered to or by the subject.

Findings is the largest class, and includes (ECG Test Results (EG), Laboratory Test Results (LB), Questionnaires (QS), Vital Signs (VS), Microbiology Specimen – (MB), PK Concentrations (PC), Inclusion/Exclusion Criterion Not Met (IE), Physical Examination (PE), Subject Characteristics (SC), Drug Accountability (DA), Microbiology Susceptibility Test (MS), and PK Parameters (PP). Findings captures observations of planned evaluations or tests.

The Trial Design class is comprised of Trial Arms (TA), Trial Visits (TV), Trial Summary (TS), Trial Elements (TE) and Trial Inclusion/Exclusion Criteria (TI).

The Events class captures protocol milestones. Included are Disposition (DS), Adverse Events (AE) and Medical History (MH).

The Relationship class of datasets includes Supplemental Qualifiers (SUPPQUAL in older versions, now SUPP-- with the dash representing source dataset) and Related Records (RELREC).

A sponsor should only submit the domain datasets that were actually collected in a particular clinical trial. Therefore the standards are scalable downward because less than the list above is generally included in any on trial's submission. The standards are also upwardly scalable, because if an appropriate domain does not exist, there are rules in the standards to create these domains. Over time, additional domains have been added in this manner. The SUPP-- domain for instance, was originally called SUPPQUAL. It was intended to hold all variables that were not defined in the standards but were collected within a CRF. These were identified by a set of keys to be able to merge them back in. Over time, it became clear that the SUPPQUAL dataset was unwieldy as it became too large, with too many possibilities for additions. Then the decision was made by the consortium of CDISC members to add a supplemental dataset to each domain as needed. Therefore for instance, any variable in the Demographic portion of the CRF that did not fall within the variables identified for the DM dataset can be placed into the SUPPDM dataset and merged back in as needed for analysis using the appropriate keys. The original SUPPQUAL remains as well in the standards, as the intent is to make them fully backwards compatible. Samples of merging techniques SAS® version 9.2 will be explored in the sample code provided.

Variables are also given attributes by the standards. They are classified according to its role, which gives further information about how it will be used.

- Identifier variables give information about the study, subject, domain and the sequence number of the record in that dataset.
- Topic variables give the focus of that observation
- Timing variables describe dates, times or other time periods when the observation took place
- Qualifier variables describe the results or units of that observation. These have 5 subgroups: Grouping, Result, Synonym, Record and Variable.
- Rule variables are specific to the Trial Design Model and describe the rules of the trial design.

The variables have a strict naming convention across domains. Each variable is described in the STDMIG by class, giving the general rules and standard types of variables to be used in the domains within that class. Except for the key variables that are common across all domains, each variable starts with the two-letter code of that domain. This is often shown by two dashes in the documentation. The variables have a maximum length of eight characters in order to be compliant with older versions of SAS®. This is because the datasets are generally presented to the FDA as SAS v5 transport files, which was the last open-code format of SAS® datasets. So the rules are constant across domains, just as LeWitt imposed constant rules across new artworks.

Another instruction that is very basic in nature but rich in detail and complexity just like LeWitt's art is the representation of dates. The date format that has been chosen is the ISO 8601 extended format. This is an international standard, and provides a text-based representation of dates with or without time. It also defines intervals of time and durations of time. The date/time standard has several rules, which are described in detail on page 27 of the SDTM Implementation Guide, Version 3.1.2 which is freely available on the CDISC website.

The extended ISO 8601 format uses '-' as a delimiter between date elements, ':' between time elements and 'T' between date and time. The format is expressed at YY-MM-DDThh:mm:ss.

Therefore, New Years Day for 2009 at 10:05 am is represented as follows:

```
2009-01-01T10:05:00
```

In early versions, a variable was needed to define precision of the date. This variable has since been dropped, as the precision of the date can be computed by finding the least non-missing element within a date. Samples of coding this simple concept in SAS® version 9.2 are explored in sample 2 code attached.

Constant Improvement

The SDTM standards are constantly evolving in response to its member body. New modules can be proposed and will be added to the documentation whenever many members find one useful. These go through an exhaustive review process. Most of the work takes place within work groups of volunteers and CDISC employees, and the efforts are posted on the website at www.cdisc.org. Membership is free, unless a company wishes to become a corporate sponsor.

The CDISC Terminology Initiative was formalized in 2005 in an effort to continuously define and support the terminology needs of the CDISC models across different types of clinical trials, both on animals and people. This effort also extends to the requirements for CDASH version 1.0. There is a live webpage at [CDISC New Term Request](#). The web page is intended for new terminology requests and also changes to existing terminology.

Conclusion

The parallel I have drawn between modern art that is descriptive in nature and the CDISC consortium is that the instructions are constantly evolving to clearly and succinctly define the standardization of a concept that can therefore become global in nature. Just as Sol LeWitt's art can be and has been installed in exhibitions around the world by following a clear set of instructions, so too can the CDISC standards be used in studies across companies and continents. Therefore a simple sentence describing a line or shape in LeWitt's conceptual art can be executed by another artist at another

time and place. In the same manner, SAS programmers are the trained artists for the CDISC standards, SDTM being the most advanced example, and can therefore present collected data in this form so that it can not only be reviewed more efficiently by the governing agencies, but also can eventually become part of a bigger picture of health care and drug development.

Resources

The primary documents for the CDISC organization's work can be found at <http://www.cdisc.org>

Sol LeWitt A Wall Drawing Retrospective at MASS MoCA can be visited online by accessing: www.massmoca.org and selecting SOL LEWITT on the home page.

Bibliography

Study Data Tabulation Model prepared by the CDISC Submission Data Standards Team (Version 1.2), November 12, 2008, © 2008 Clinical Data Interchange Standards Consortium, Inc. All rights reserved.

Study Data Tabulation Model Implementation Guide: Human Clinical Trials (Version 3.1.2), November 12, 2008, © 2008 Clinical Data Interchange Standards Consortium, Inc. All rights reserved.

LeWitt, Sol wall drawings allo Studio G7, Damiani 2006, Daimani Editore, Bologna, Italy

Sol LeWitt A Wall Drawing Retrospective, Gallery Guide 2009, Mass MoCA, 1040 MASS MoCA Way, North Adams, MA 01247

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Karin LaPann

ViroPharma Inc.

730 Stockton Drive

Exton, PA 19341

Work Phone: 1-610-321-2329

Fax: 1-610-458-7380

E-mail: Karin.LaPann@viropharma.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of

SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.