Data Quality and the Origin of ALCOA

Stan W. Woollen
Senior Compliance Advisor
Stan Woollen and Associates

If one does a Google search on the term ALCOA, virtually all of the hits lead to Alcoa Inc. According to the company's homepage, "Alcoa Inc. is among the world's top producers of alumina and aluminum. Its vertically integrated operations include bauxite mining, alumina refining, and aluminum smelting; primary products include alumina and its chemicals, automotive components, and sheet aluminum for beverage cans". Gasp!! Not one reference to data quality.

With all due respect to Alcoa Inc., which I'm sure is justifiably quite proud of the quality of its products, ALCOA, in the context of this article, is an acronym which has nothing, or almost nothing, to do with Alcoa Inc.

Many seasoned QA professionals, have heard of ALCOA used as an acronym to identify the elements of data quality. This acronym stands for **A**ttributable, **L**egible, **C**ontemporaneous, **O**riginal and **A**ccurate. The extent to which data possess these qualities determines its level of quality and thus fitness for use, particularly with respect to use for regulatory purposes.

How did the ALCOA acronym come to be used in the context of data quality? Where did these elements of data quality, for which ALCOA stands, originate? The purpose of this article is to shed some light on these two questions.

The answer to the first question of how the ALCOA acronym came to be used in the context of data quality is it bit obscure, so I'll address that first. The ALCOA acronym was first coined by me while serving in FDA's Office of Enforcement back in the early 1990's. Exactly when I first used the acronym I don't remember, but I do remember why and how it came to be.

Prior to coming to the Office of Enforcement, I served 15 years as an investigator in FDA's field office covering the Washington DC Metro area. As a regional expert investigator, and later as a supervisory investigator, my duties included occasionally, making presentations to FDA's external constituents and serving as an instructor at FDA's internal training seminars and schools. However, public speaking was just a small part of my job in the field, and not necessarily my favorite part.

When I transferred to the Office of Enforcement in FDA headquarters, public speaking was to become a much bigger part of my job and remained so during the next 14 years in various headquarters offices.

When I first came to the Office of Enforcement, Dr. Paul Lepore was the Agency's Bioresearch Monitoring (BIMO) Program Coordinator. I came to headquarters as the Associate BIMO Program Coordinator, working with Dr. Lepore.

For those readers who don't know who Paul Lepore is, you'll see his name listed prominently as the FDA contact for the GLPs in Federal Register notices announcing the final GLP regulations in 1978. Indeed, Dr. Lepore was a principal architect of the GLP regulations and for many years served as the agency's lead authority on the interpretation and application of GLPs. In that capacity, he of course spent considerable time making numerous presentations. He had years of experience speaking publically on all aspects of GLPs and was quite an accomplished speaker.

As fate would have it, shortly after I came to the Office of Enforcement, Dr. Lepore took advantage of a rare opportunity to retire early from FDA. After only a year as the Associate BIMO Program coordinator, I was immediately thrust into his former role as the Agency BIMO program coordinator and took over all of his responsibilities including serving as an agency spokesman for GLPs and FDA's overall BIMO program.

Facing the prospect of a great deal of public speaking; panic set in. After years of field experience implementing the BIMO program, I had the technical knowledge and experience with the GLPs and GCPs, but I didn't have Dr. Lepore's speaking experience. I needed to brush up on my public speaking skills immediately. This included coming up with ways to remember and speak on a variety of topics extemporaneously. I used a number of techniques to help me easily remember and organize my thoughts.

One of the techniques I used was to come up with acronyms that I could easily remember to help me organize my presentations. This is where the acronym ALCOA came in. Admittedly, this acronym was easy for me to remember, because Alcoa Inc. was a commonly known company name. On the other hand, the ALCOA acronym was not known in the context of data quality.

Ordinarily I didn't use acronyms in the actual body of my presentations. However, in preparing slides for one presentation, I ran out of space on a slide, and just inserted the acronym ALCOA as a bullet-point reminder to myself. I don't remember exactly when, or in which presentation I first used the actual ALCOA acronym. However, I do remember the consternation of at least one member of the audience, who in trying to later decipher the "government jargon" in my slide, asked what ALCOA stood for. I had to explain what ALCOA stood for on many occasions. Consequently, the acronym eventually became known in the QA community to such an extent; I could use the ALCOA acronym alone on my slides as a concise and lazy way to discuss the elements of data quality. This is how and why the ALCOA acronym originated in the context of data quality.

While I did coin the ALCOA acronym for the elements of data quality, I take no credit whatsoever for the origination of the actual elements of data quality for which ALCOA

stands. This leads to the second question in this article of where did the elements of data quality originate?

The answer to the question is a little complex because the elements of data quality have their origins in a number of FDA's quality systems regulations. Though not originally referred to as such, quality systems regulations have been around for quite sometime at FDA. The earliest of FDA's quality systems regulations include Current Good Manufacturing Practice Regulations (cGMPs) for drugs, biologics, and later, Good Laboratory Practice regulations (GLP) and GMPs for medical devices. Virtually all of these quality system regulations articulate one or more of FDA requirements for data quality covered by the ALCOA acronym.

Although the cGMPs articulate a number of the expectations for data quality, the GLP regulations, in my opinion, are the first FDA regulations which bring the ALCOA elements of data quality together in a comprehensive fashion. For this reason, this article will focus on the GLP requirements pertaining to data quality elements, particularly 21 CFR 58.130(e) which articulates virtually all the elements of ALCOA.

The first **"A"** in ALCOA stands for **<u>Attributable</u>**. Simply put, FDA expects data to be linked to its source. It should be attributable to the individual who observed and recorded the data, as well as traceable to the source of the data itself. (e.g. study, test system, analytical run, etc.) The applicable GLP requirements pertaining to attribution of data are found in 21 CFR 58.130 (c) and (e).

The requirement for attribution of data to the individual who collected it is found in 58. 130(e). According to the regulation, *"All data entries shall be dated on the date of entry and **signed or initialed by the person entering the data**"*. The same is true for automated data. The regulation states, "*. . . In automated data collection systems, the **individual responsible for direct data input shall be identified at the time of data input**. . .*" Not only does this concept of attribution apply to the collection of original data but also to any changes made to the data. Changes made to data must be signed and dated by the individual making the changes.

An example of a requirement for attribution of data to its source, is illustrated by 21 CFR 58.130(c) which requires study specimens to be identified by test system, study, nature, and date of collection.

The **"L"** in ALCOA stands for **<u>Legible</u>**. Quality data must also be legible if it is to be considered fit for use. The concept of legibility means that data are readable. This of course implies that data must be recorded permanently in a durable medium (e.g. pen and ink on paper). 21 CFR 58.130(e) addresses this directly by requiring that, *"data shall be recorded directly, promptly, and **legibly** in ink"*. The concept of legibility of data also extends to changes made to data. For example, 58.130(e) requires that changes be made so as not to obscure the original entry, thereby maintaining its legibility.

The requirements for legibility of electronic data may present technical challenges and take on new meaning, with respect to recording data permanently on a durable medium. However, the underlying concept of legibility/readability is the same. If one consults FDA's Electronic Record; Electronic Signature rule (21 CFR 11), many of the traditional ALCOA data quality elements are addressed. For example, with respect to legibility of data, 21 CFR 11.10 (b) requires that compliant electronic systems have, *"The ability to generate accurate and complete copies of records in both human readable and electronic form suitable for inspection, review, and copying by the agency."* This requirement clearly establishes the expectation that electronic data must be readable (i.e. legible).

The **"C"** in ALCOA stands for **Contemporaneous.** This element of data quality refers to the timing of data collection with respect to the time the observation is made. In short, the more promptly an observation is recorded, the better the quality. Data should be recorded at the time the observation is made (i.e. contemporaneously). The GLPs address this at 21 CFR 58.130(e) as discussed above. Specifically the regulation at 21 CFR 130(e) states, *". . . data shall be recorded directly,* ***promptly****, and legibly. . ."*

The requirement that data be contemporaneous is also implied in the regulations that require the date of data entry to be recorded. For example, 21 CFR 58.130 (e) also requires "All data entries shall be ***dated on the date of entry*** and signed or initialed by the person entering the data". The longstanding and virtually universal requirement in FDA regulations for dating record entries is intended to assure, or at least document, the extent to which data is recorded contemporaneously with the observation being made.

The **"O"** in ALCOA stands for **Original.** Original data is generally considered to be the first and therefore the most accurate and reliable recording of data. The terms source data or raw data embody this concept of the first recording of data, and are sometimes used interchangeably. Source data is the term generally used in the context of Good Clinical Practices (GCP), while GLP enthusiast use the term raw data as it is officially defined in the GLP regulations at 21 CFR 58.3 (k). The term source data, although defined in guidance, is nowhere to be found in FDA regulations. On the other hand, the GLPs were the first and only place the concept of raw or source data is actually put explicitly into FDA regulations. Indeed, the GLP definition of raw data is the foundation upon which the term source data is defined in a number of FDA guidance documents on GCPs.[1]

The definition at 21 CFR 58.3 (k) states in part *"Raw data means any laboratory worksheets,* records*, memoranda, notes, or exact copies thereof, that are the result of* ***original observations*** *and activities of a nonclinical laboratory study. . ."* Although the GLPs and GCP do provide for the substitution of certified copies of source/raw data in lieu of the original record, the concept that the original recorded data is of the highest quality is retained. The concept of originality being an element of data quality is further reinforced in 58.130(e) which states ". . . data shall be recorded directly. . ."

---

[1] See ICH E6 Consolidated Guide for Good Clinical Practices and Computerized Systems Used in Clinical Investigations

The last **"A"** in ALCOA stands for **<u>Accurate.</u>** Accuracy is an implied element of data quality under the GLP regulations. The <u>Merriam- Webster Dictionary-</u> defines accurate as

1: <u>*free from error*</u> **especially as the result of care <an** *accurate* **diagnosis>**
2: <u>*conforming exactly*</u> **to truth or** <u>*to a standard*</u>**: <u>EXACT</u> <providing** *accurate* **color>**
3: able to give an accurate result <an *accurate* **gauge>**
synonym see <u>CORRECT</u>

Accuracy is probably the most intuitive element of data quality. The most direct reference in the GLPs to the expectation of accuracy is found in 58.35 (b) which requires the QAU to assure the final report **<u>accurately</u>** describes the study conduct and that the reported results **<u>accurately</u>** reflect the raw data.

The first two definitions of "accurate" above are also implicit in the GLP regulations at 58.130(a) and (b). For example, under definition two, accuracy involves conforming exactly to a <u>*standard.*</u> For the conduct of a nonclinical study, the product standard is the protocol. 58.130 (a) requires that a study must be conducted <u>in accordance with</u> the protocol. Likewise 58.130 (b) requires Test systems to be monitored in <u>conformity</u> with the protocol.

While there are continuing discussions and consideration of what constitutes data quality, those ALCOA elements of data quality which have their origins in FDA quality systems regulations continue to form the basic foundation upon which data quality rests.

.