# Challenges in consistent, correct, and comprehensive characterization of R&D materials and processes

**Graham A. McGibbon, Sanjivanjit K. Bhal, Andrew A. Anderson, Joe DiMartino**

Advanced Chemistry Development, Inc. Toronto, ON, Canada

ACD/Labs

## Introduction

Pharmaceutical R&D organizations rely extensively on analytical data about materials and processes as they strive to bring innovative products to market. Medications are drugs or medicinal products; which are materials comprised of substances designed to treat or prevent disease.

To ensure a supply of appropriate quality, safe, and efficacious material, those products and substances must be well characterized by analytical data; and the processes used to prepare them must be well controlled. Quality materials are inevitably prepared in processes that instantiate the plans and protocols of well conceived experimental designs. Some consideration of general concepts and models for effective management of design, process, and material characterization data is thus important.

A general conceptual approach to relating entities via one or more unit operations for the various processes involved in producing and characterizing drugs in R&D is shown in **Figure 1**.

In this poster we will use the example of impurity data management in the development of new therapeutics, to illustrate these challenges and highlight software designed to facilitate this work.

## Methods: Managing and Preserving Data including Metadata

Across process optimization and scale-up operations, batches of material will be produced and then analyzed by a variety of techniques, each with their own proprietary data encoding. The software application Luminata™, (v2017.1) based on the ACD/Spectrus Platform, was used to manage the analytical and chemical data shown in this poster.

The architecture and forms in Luminata allow for consistent metadata describing experimental equipment, methods, materials, and processes. The features of ACD/Spectrus enables data integrity through the capture of data from a variety of instruments and techniques. In this poster are shown HPLC data from an Agilent-1200-Series with an Agilent VWD G1314B UV detector, acquiring spectra at 210nm; and an Agilent 6110 Quadrupole API-ES Mass Spectrometer. The software allowed re-processing, reporting, and re-use of the data throughout the impurity lifecycle, and for sharing between sites and organizations. User login and audit trails are available to help ensure data integrity.

## Concepts for Process, Material, and Data Operations and Relationships

Fundamentally, experimental designs should nowadays be represented via digital data and metadata. So while the drug materials are of ultimate value to patients, underlying considerations exist in data analysis, data management, and data preservation with integrity.

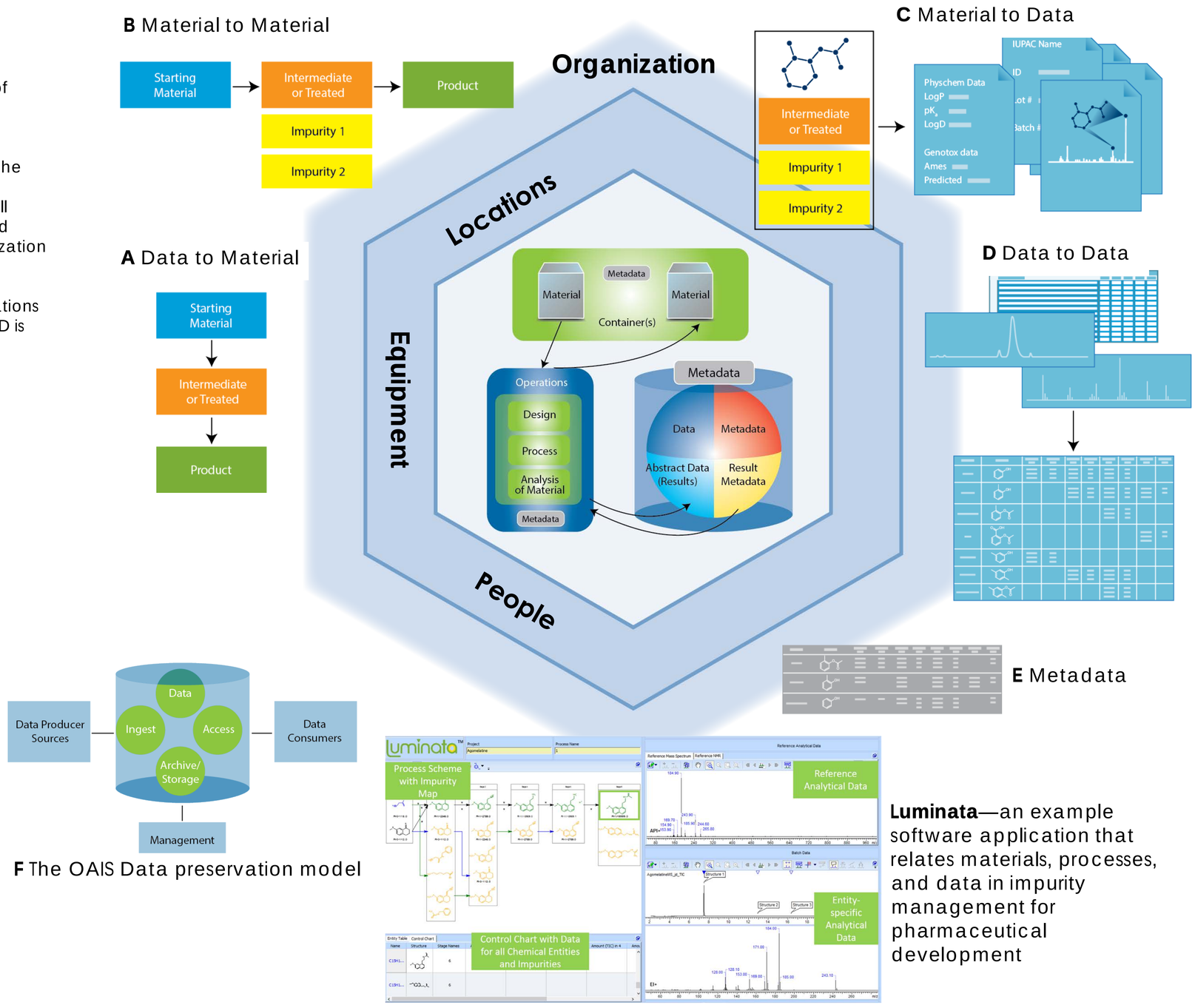
Even with digital data, a key challenge to analyze and integrate it effectively arises when assembly of results from different complex data sources is necessary for achieving consistent, correct, and comprehensive interpretation. The data of experimental designs, protocols, and perhaps equipment control information, are analogous to cake making. The recipes indicate specified quantities of ingredients and conditions in which they would be combined, manipulated, and transformed into desired new materials. This operation may be summarized (see Figure 1A) by the **data to material (DM)** relationship**.**

The subsequent physical and (bio)chemical process unit operations with materials is describable by appropriate sequential or convergent **material to material (MM)** stages (see Figure 1B) . These may commonly be chemical reactions, biological transformations, or treatments such as drying, heating, stirring, filtering, crystalizing, or using chromatographic methods of separation and isolation.

Pharmaceutical materials analyses yield characterization data of various kinds depending upon the test methods and instruments used. Analytical data is one or more measurement outputs, typically accumulated from a device into a historical artifact, for instance a file, and in raw form the data would comprise signal and noise but be captured completely, accurately, and precisely. This is denoted by the **material to data (MD)** relation, see Figure 1C.

Furthermore, to comprehensively characterize a material or process, data must not just be acquired but processed. There may often be an additional challenge in that data in various vendor formats from diverse analytical sources should be analyzed, abstracted, and assembled—while also requiring direct, statistical, or analytics comparisons—to reach confident answers. These subsequent data analyses are collectively describable as **data to data (DD) relations** in Figure 1D. A subset of these relations involve contextual rather than measurement data and these are indicated in Figure 1E by the term **Metadata**.

Data are acquired for a primary purpose, typically the characterization of a material or process. However, the data may be compared with other data to establish secondary patterns of materials, processes, and the analyses themselves, as well as equipment usage and failures. Furthermore, if data are of sufficient quality, have contextual metadata, and are stored appropriately, they may become a future 'Data Asset' for conducting Data Science to uncover new and unanticipated trends and information. Key for this is some kind of **data preservation** while maintaining data integrity. One model for that is the ISO 14721:2012 open archival information system (OAIS) reference model for digital preservation as summarized in Figure 1F.




**Luminata**—an example software application that relates materials, processes, and data in impurity management for pharmaceutical development

**Figure 1**. A model for relating data and materials in a variety of R&D pharmaceutical operations. Operations are classified into four general types: Design, Process, Test/Measure, Analyze. Transport is an operation that is not indicated. Operations provide relationships of (A)Data to Materials, (B) Materials to Materials, (C) Materials to Data, (D) Data to Data including (E) the subset of Data denoted as Metadata. Examples of each are described in the text.

## Data Integrity

OAIS refers to Data Integrity as data being: "*complete and unaltered in all essential respects*"; and applications designed to maintain integrity aim to "*ensure data is recorded exactly as intended, and upon later retrieval, ensure the data is the same as it was when it was originally recorded".* Unintentional changes to data are to be avoided, and responsible strategies put in place to detect unintentional changes and react as appropriately determined. However, digital preservation efforts may necessitate modifications to content or metadata through responsibly-developed procedures and by well-documented policies. Organizations or individuals may choose to retain original, integrity-checked versions of content and/or modified versions with appropriate preservation metadata."

It is absolutely essential to ensure that the acquisition and preservation of data are conducted with the awareness that the practicalities of acquiring good quality data and the primary value of that data for providing clear information about the material or process of interest, are not lost or demoted to inferior status in the hype of having completely open data; and of data science with its less immediately tangible, but undoubtedly useful in some future state, benefits.

### Applying data management principles to process development and impurity data

First it is important to take care about data integrity and data access for primary use. The ALCOA principles of data being Attributable, Legible, Contemporaneous, Original, Accurate, complete, consistent, enduring, and available offer guidance on making data appropriate and useable. Then principles like FAIR should be incorporated to the extent feasible without compromising the initial objective.

As can be seen from the panels of Figure 1 A-E, the impurity management solution Luminata attends to the aspects of relating contextual metadata about data, materials (for example ID, Name, chemical composition), and processes (physical conditions like temperature or chemical reaction details). Instrument data are captured accurately and completely into the solution and are directly visible.

The data and metadata are retained in a database and access protections and audit trails are available to ensure that data integrity can be maintained. The relationships DM, MM, MD, and DD included within the Luminata solution allow a 'line-of-sight' to consumer from producers, or in other words, from results-to-data-to-experiment with the original intent evident.

## Conclusions

- Complex relationships between materials, processes, and data require a clear, defined model for preservation and future exploitation, which are incomplete in many organizations.
- A high volume of analytical data is collected for the characterization of processes and materials but lack of formalized relationships and software systems designed to manage its unique complexities means that re-use is challenging, even when the data is accessible.
- Maintaining data integrity means that consistent, correct, and comprehensive information can be accessed for future studies, both those relating to the initial purpose for data collection and broader big data-type applications.
- Organizations must formalize the relationships of their data in order to be able to manage it effectively and preserve for future studies, IP protection, regulatory submissions etc.
- Software, such as Luminata for impurity data management, must be capable of managing the complexity of the data while enabling re-processing, reporting, and re-use throughout product lifecycles.

## References

Luminata http://www.acdlabs.com/solutions/luminata/resources.php
OAIS https://www.iso.org/standard/57284.html
ALCOA https://www.agilent.com/cs/library/slidepresentation/Public/PT0514_Data_Integrity.pdf
FAIR https://www.force11.org/group/fairgroup/fairprinciples