Use Case Overview

High-Tech & Pharmaceutical LIMS & ELN with camLine's XperiDesk

Version 1.0

Dr. Dirk Ortloff

October 26, 2020

When studying high-tech & pharmaceutical R&D projects, the infrastructure often still consists of paper run cards / travelers, manual data collocation on paper or spreadsheets and storing result data on simple file servers. While this worked well for the last two decades, the current economic situation and the competitive turbo of the last years requires organizations to upgrade their infrastructure.

The current use case overview gives an outline about the situation and offers one possible solution to the challenges.

1 Introduction

High-tech is a term that is widely used to describe a device or system being state of the art. So high-tech can be applied to the cutting edge such as nanomaterials but also to some "old" industries like pharmaceuticals for example. What all these industries have in common is a production process that has multiple discrete steps or phases in reactors, many parametric settings or both. And that is where the R&D part of the title comes from. The production process needs to be developed or improved or the active substances for pharmaceuticals, dosing, forming, etc. need to be developed. Experiments need to be executed and measurements must be taken. Thanks to the increasing variability of tools to generate measurement results there are lots of images, graphs or plain measurement data to look at and evaluate. And with advances in physics there are new hard disks available so that today it is not a problem to have storage with dozens of terabytes for very little money.

We can generate measurements; we can evaluate and store the results. So is everything good? No, it isn't! The truth is, with today's IT, it is (in most cases) not a big problem to store large amounts of data. And that amount is growing every year by approximately 50% and this is the crux of it. The data can be stored. Oftentimes this is referred as the usage of a Data Lake. Many technologies and methods used are optimized for storage even if they are more than 30 years old. New paradigms for storing like Big Data and its horizontal scaling approach offer even more scaleability and promise easy and fast analysis and more insights. But this is not always the case, as stipulate in [Chemist 2019]. While that is fine for archiving results and fulfilling regulatory requirements, in R&D gaining information and knowledge out of that data is of highest importance. However, with drowning in that data and no effective way to retrieve it to generate knowledge for informed decisions, there's a problem – one that increases by 50% every year!

2 A Typical Case

Let's look at a typical medium-sized company with about 20 research engineers, pharmacists, etc. (in the following all referenced as researcher) doing experiments. The researchers work in different R&D departments and rely on the work of other researchers. There is a template (typically MS Excel) for the experiments but, because we are talking R&D, this changes regularly (and every researcher has variants thereof).

Once the first stage is completed, the data is handed over to the researcher responsible for the next phase, who copies the data into their own template. Finally a testing researcher takes the results from the experiments and does some measurements, which are then stored in a separate sheet. The copy & paste not only introduces errors but also destroys the relations between the data points, in fact the context of the data points. It is nearly impossible to find the source of the values in the new sheet or to do a commonality analysis over more than one stage of the R&D process, especially when sheets are moved or renamed.

As engineers join and leave the company and as naming conventions for partially finished experiments change, a lot of sheets are generated. When a new project starts the old data is (or should be) reviewed. However that becomes increasingly difficult. The files are stored on a file server (oftentimes even only on the laptops / desktops of the researchers) sorted by stage and author. But that isn't helpful when looking for certain material. Even if the directory structure is changed and sorting by material is done, the problem remains. So if, say, the next project needs information involving "threshold voltage limit", "ratio active substance / filling material", ... you name it, searching for it is nearly impossible in the thousands of Excel sheets already accumulated on the file server, document or content management system. In a nutshell the data, information and knowledge retrieval is experienced as shown in Figure 1. As a result studies [Barkai 2012] estimate that engineers spent 25-30% of their time just managing and rearranging data for various purposes (such as reporting and searching). To summarize, there are the following deficiencies encountered in development organizations:

- Documentation on file servers, Excel, paper notebooks, etc;
- Untraceable and undiscoverable R&D results;
- Limited formalized data available;
- R&D data not interlinked or related.

Figure 1: The search for knowledge often resembles the search for the proverbial needle in a haystack [Troscianko 2020]

At the end, experiments are repeated because the data (and information extracted from it) is not readily available.

A study of IDC Manufacturing Insights [Barkai 2012] confirms that "as much as 40% of all R&D experiments are repeated unnecessarily and often inefficiently, delaying projects and increasing costs and risk". It is summarized that "45% of the resources allocated to product development and commercialization are wasted". This is a number that no company can ignore if it wants to survive in a global market.

3 Solutions

Data is raw and it can exist in many different forms. In R&D we are looking for information and knowledge; but how to gain that? In the production area this problem is handled by a multitude of tools from categories such as Manufacturing Execution System (MES) [Wikipedia 2020a], Product Lifecycle Management (PLM) and Enterprise Resource Planning (ERP). These tools use a database as background storage to accumulate data, provide it for evaluation and to derive appropriate actions out of these evaluations. However, simply using a database instead of Excel to do a search will not help much, as contrary to production, in R&D data is changing constantly. New parameters are added daily and a big chunk of the data – such as images – is inherently unstructured and not suitable for database storage. Equally, information needs relational connections between the data and these connections are not always known at the beginning of R&D projects.

A system to manage R&D data must be a comprehensive repository for structured and unstructured data. It must provide multi-dimensional access possibilities. This is necessary to cater for multi-disciplined working environments as pharmaceutical, mechanical, electronics, electrical and other researchers need to be able to look at the same data from different perspectives. Furthermore, graphical navigation through historical data / information must be provided to enhance retrievability.

Several industries, like pharmaceuticals, are even regulated in terms of which data, which properties, which changes, who has access etc. need to be documented. As an example, for the pharmaceutical industries the FDA enforced via CFR21 Part 11 [**FDA** 2019] that certain data needs to be recorded and signatures (physically or electronically) captured. This way they even enforce to keep a full audit trail of who recorded, changed, ... what and therefore enforce that pharmaceutical and medical device company's need to be able to travel back in time, not only for production but even for engineering.

Other important requirements allow for data discovery are sophisticated searching capabilities. These need to allow searching structured as well as in unstructured data such as documents. While text-based unstructured data/files can be searched relatively easy using advanced index services, structured data management needs to be equipped with physical awareness. This means that searching for a 'voltage' only searches voltage parameters to comprehensively retrieve the information. Both means of searching must be able to be combined and the system must also be able to use freely definable relations in the searches.

However, two of the most important parts of such a system are import and export. A data management system for R&D must provide means to import data from existing data sources (Excel sheets, file servers, SQL databases, etc.) and also means to export, e.g. search results into other tools (Excel, statistical software, ERP, MES, etc) for further processing. It must enable engineers to spend more time evaluating rather than managing data. (see Figure 2)

PDES (Process Development Execution Systems) [**Wikipedia** 2020b] like XperiDesk from camLine GmbH aim to fill this role. They provide a centralized platform to collect, evaluate and export data in a multidisciplinary research facility. PDESs are geared to cope with the ever changing structured and unstructured data that is seen in R&D organizations.

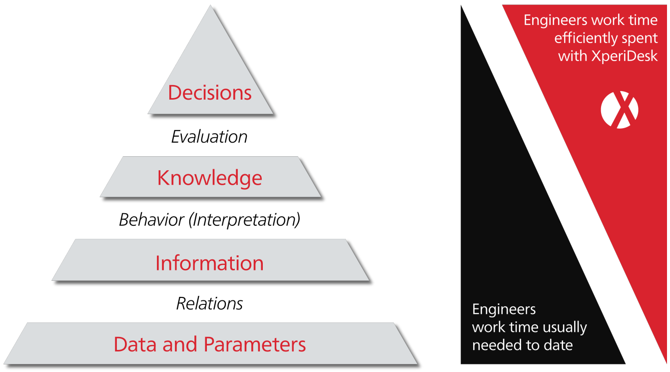

Figure 2: The knowledge pyramid and the distribution of engineering time before and after the introduction of a PDES

4 Scratching the Surface

This article just scratched the surface of the problem. There are many more issues to consider, such as data security, regulatory requirements and IP protection – in short, the complete issue of information governance for R&D facilities [Ortloff 2011]. The important message though is that companies must manage their R&D data and processes with the same rigor as they do for production, supply chain and sales&marketing. The initial investment in training and establishing of new procedures is nothing compared to the potential gain in reduced costs and improved time-to-market and might be the deciding factor to make a company ready for the development challenges of tomorrow.

References

Barkai, Joe (2012): "Accelerating Science-Led Innovation for Competitive Advantage". In: **url**: https://www.tradepub.com/free-offer/idc-manufacturing-insights-accelerating-science-led-innovation-for-competitive-advantage/w_aaaa2562 (cit. on p. 3).

Chemist, Manufacturing (Aug. 7, 2019): *Manufacturing misconceptions: The difficulties of tackling big data analysis*. In: **url**: https://manufacturingchemist.com/news/article_page/Manufacturing_misconceptions_The_difficulties_of_tackling_big_data_analysis/156965 (cit. on p. 2).

FDA (Sept. 19, 2019): *Code of Federal Regulations Title 21*. In: **url**: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=11 (cit. on p. 4).

Ortloff, Dr. Dirk (Aug. 2011): *Information Governance in MEMS / NEMS process development*. In: **url**: https://www.researchgate.net/publication/270507843_Information_Governance_in_MEMS_NEMS_process_development (cit. on p. 5).

Troscianko, Jolyon (Oct. 26, 2020): *Needle in the haystack, credits to Jolyon Troscianko*. In: **url**: http://www.jolyon.co.uk/illustrations/basic-vision/ (cit. on p. 3).

Wikipedia (June 26, 2020a): *Manufacturing execution system*. In: **url**: https://en.wikipedia.org/wiki/Manufacturing_execution_system (cit. on p. 3).

Wikipedia (June 26, 2020b): *Process Development Execution System*. In: **url**: https://en.wikipedia.org/wiki/Process_development_execution_system (cit. on p. 4).