Managing complexity and enhancing knowledge retention as process understanding evolves

Martin Owen

Insight by Design

Abstract:

In industry, as processes are developed and transferred into manufacturing, scientists and engineers perform studies iteratively and develop process and product understanding. These studies are often carried out over a few years and sometimes at different sites. How can we organise our knowledge more effectively and generate a better organisational memory? This is particularly important for academic research groups where the turnover of post-graduates is high. To help teams or organizations with high staff turnover retain insights, we need to address several challenges.

How can we

- help teams compare and contrast the efficiency and effectiveness of historical experimentation?
- show what assumptions were made and identify knowledge gaps?
- help teams gain additional insight from disconnected tabulated data-sets and free text?
- discover what we need to know, when it's not always clear what we need to know?

In this paper I look at data visualisation and modelling both in the context of a learning environment and an implementation environment. I have challenged participants in both industry and academia to study the same process in a series of workshops. Teams select different factors, different ranges and perform different classical, custom and mixture designs. I then help participants compare their own output with studies carried out in previous workshops by different teams. We then use the holistic learnings to generate new ideas and actionable decisions.

The potential for JMP13 to help build and sustain Organisational Memory

Prior to launch of a new version of JMP, the Early Adopter program enables customers to download and preview the initial builds. There is an accompanying series of Webcasts and documents which are very informative and there is an opportunity to provide feedback to help improve the product and ensure that it performs reliably post-launch.

The JMP13 Early Adopter program has given insight as to how new functionality can enhance our use of prior information. This presentation will illustrate how existing functionality in JMP 12 and new features in JMP13 (e.g. Dashboard Builder, Query Builder and Text Explorer) can help us build a more accessible and structured "organizational memory" of both observational and experimentally derived data.

We can see how this functionality can then be applied to a portfolio of implemented studies in academia, so that prior information can inform future research.

Creating a learning environment

I was asked to deliver a workshop earlier this year where the sponsor wanted to know

- How reliable is our data?
- How effective is our experimentation?
- Can we make better use of prior knowledge?

These are really good questions and we wanted to create an overall workshop design which fostered an environment where workshop participants would take an active role in exploring how to answer these questions for themselves.

Stepping stones from classroom to workplace

One of the fundamental principles of our workshop design is to help participants new to experimental design take the "leap of faith" to progress from the classroom to implementation in science laboratories. Different participants will have different preferred learning styles so we usually use a mix of case-studies, simulations and "laboratory of life" practical sessions to help participants rapidly build experience and provide an environment where they are able to explore different tactical approaches and see the consequences of their decision-making.

Case studies

External case studies demonstrate tangible concrete examples of the cost and benefits of deploying experimental design, particularly if the examples chosen resonate with the participants.

Locally applied case studies can be even more effective as they are likely to create greater resonance with the workshop group. However, if the organisation department is relatively new to experimental design, then these locally examples may not yet be available. With the MSc Quality by Design students from De Montfort University, we set the expectation that the workshop participants should pool their knowledge and experience to build a portfolio of examples for themselves. These formulation-based case studies would then be available to share with the following years' cohort of students. The portfolio builds year on year to provide a knowledge repository which students can use to help them investigate new designs and extend research investigations.

However, one downside of simply presenting case studies in a workshop is that many of the key decisions have already been made and it may not be possible to explore alternative "what-if" scenarios.

Simulations

In contrast, simulations are more hands on and can offer the opportunity to try out different options e.g. comparing and contrasting different designs, models, analyses and model term selection. Having obtained a model, the "optimum solution" can be verified with additional experiments to see how reliable it is in meeting specified success criteria.

Laboratory of Life

Whilst simulations are a very effective learning tools, the choice of factors and settings has already been made. This, however, is one of the most challenging decisions when it comes to actual implementation. We have often noticed that "laboratory of life" practical examples provide an even closer step to the reality of the challenges the participants will face when applying experimental design techniques back in the workplace. We take everyday situations and encourage students to design studies to gain process understanding. One particularly effective example of this has been our tablet disintegration challenge.

Figure 1: The Tablet Disintegration challenge – assessing the impact of Tablet Drop Height

(1) A comparative study

The set-up couldn't be simpler. All the participants have to do is to drop an effervescent tablet into a glass of water and measure the time it takes for the tablet to disintegrate. The class is split up into groups and each group tests a different tablet formulation.

We collect and collate the data, look at the average time for each type of tablet to disintegrate and observe any differences in variability between groups.

(2) Full factorial design

In the next study the class is split up into groups and each group independently brainstorm factors they would like to investigate. Each group chooses three factors to investigate and the ranges or settings for each factor and performs an eight experiment full factorial design.

Again the data is collated and analysed to see which factors are statistically and/or practically significant.

(3) Classical fractional factorial

The groups are challenged to investigate between 4 to 7 factors, still using only eight experiments.

When they analyse their experiments this time, they are encouraged to examine

(a) what new information they have gained from the new study.

(b) what information they may not have gained had they performed the fractional factorial first time around instead of the full factorial.

(4) Augmentation, Custom designs and Mixture designs

Now the teams have built up some experience they are encouraged to explore other design options. The direction of exploration may be influenced by the workplace environment the participants will return to. For example:

- Formulators are likely to need to understand mixture designs which are used when the % levels of the formulation components needs to sum to 100%.
- Some participants may have automated equipment and be able to run many experiments, whereas for others the cost of experimentation may be very high in terms of money, materials or time, so for these participants, minimising runs will be very important.
- Scientific disciplines may vary with respect to the signal to noise ratios they expect to see, or whether they frequently need to investigate simple or complex curvature.

The sequence of studies outlined here is not prescriptive and can easily be tailored depending on the experience and aptitude of the participants.

Note that the simplicity of this experiment means that not all designs can be investigated using tablet disintegration scenario. For example, to study a two way split-plot design in order to investigate a two stage process may need a different (two stage) "laboratory of life" scenario like making coffee (grinding beans and extracting the flavour) or a "hard to change factor".

Within and across study learnings

With simulations, the workshop presenter knows from the outset what the underlying model is. It is easy to anticipate the sorts of models and conclusions the participants will make and what key learning objectives can be exploited.

In contrast, when I initially started running the table disintegration modules, I had no absolute "underlying model" to assess in advance what the likely outcome might be. I could only make educated guesses which factors would be important. Using common sense and logic it was often possible to predict qualitatively (but not quantitatively) of how a factor may impact on dissolution time.

By the time I've run half a dozen workshops, I've seen many different types of designs used and different factors explored in this Laboratory of Life scenario. I have seen more tablets disintegrate than you can shake a stick at. Whilst no two studies are ever the same in terms of factor and ranges choice, teams across different organisations often chose certain factors again and again. So by now I have a good expectation of the quantitative impact of most of factors that the teams are likely to choose. I also have some sense of the likely signal to noise ratios to expect, and whether factors will be detected by the statistical analysis. If several teams find a factor to be inactive and a new team

finds the same factor active, then that provokes a discussion as to whether this is a real finding or whether the result has come about through “operator error “ or transcription errors.
By observing many different combinations of choices and pooling knowledge, a more holistic picture is obtained.

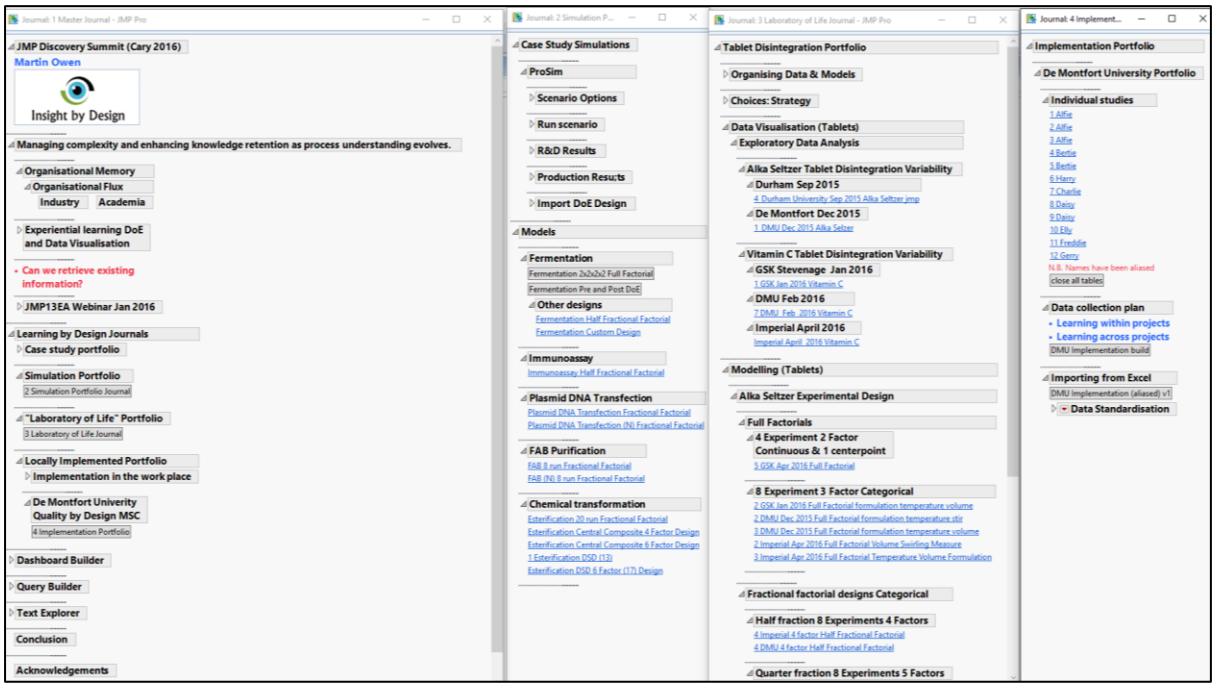
Organisational Memory

So the big question here is how can we help the individual participants see that bigger picture too? How can we organise, retrieve and communicate information from previous studies?
To retain information about a single design is relatively straightforward. The student can call the design up and summarise the findings in a report. It’s reasonably easy for an individual to appreciate how their own discrete sequential studies can build to create a greater understanding.
Enabling participants access to studies performed by others is even more challenging. How do we share data across teams? Can we utilise our prior knowledge across different products and processes, departments and even across different scientific disciplines?
But the real challenge comes when studies are dislocated across time and space, and the problem is compounded when there is organisational flux. Each year at De Montfort University, a new cohort of new students on the Quality by Design MSc, replaces a cohort leaving to progress their education or careers elsewhere.

The Journal, file management and file naming conventions

We need to be able to retrieve previously prepared data table files and recreate models using scripts. If we structure our file naming convention, that may limit our reuse of information to generate new insights. The JMP Journal functionality offers a very flexible approach to organising and reorganising data, irrespective of the original file names. As the demonstration will show, it is very easy to retrieve prior designs, compare and contrast two or more models and draw new insights.
By using a “Master Journal” we can organise information in any way we need to and reorganise easily should that be required. In Figure 2, I have used a master journal to access a portfolio of designs from simulations, laboratory of life and work place implementation. As can be seen from the different Journals, we can structure these by case study, by design type or by individual.
To reorganise the implementation portfolio by “different types of formulation”, rather than “by student” is a very simple and quick process. It is very easy to tailor a workshop or presentation by rearranging the Journal structure, rather than rearranging the underlying file structure.

Figure 2: The use of Journals to rapidly organise, access and retrieve data and models



As mentioned previously, a key part of the workshop philosophy is to enable participants to see the consequence of their decision-making. For example, in the workplace, we might face the choice between running a full factorial (20 experiments) or a half fraction (10 experiments). Using a simulation it is possible to run both designs. To make comparisons easier, it helps if we can combine both designs and view them simultaneously.

In JMP13 the Dashboard Builder provides a rapid and intuitive way to create a dashboard from any models or visualisation reports that are already open in JMP (Figure 3). Open the builder (New>Dashboard), select the template and simply drag and drop the reports into the template (Figure 3).

Figure 3: Using the 2x1 Template of Dashboard Builder

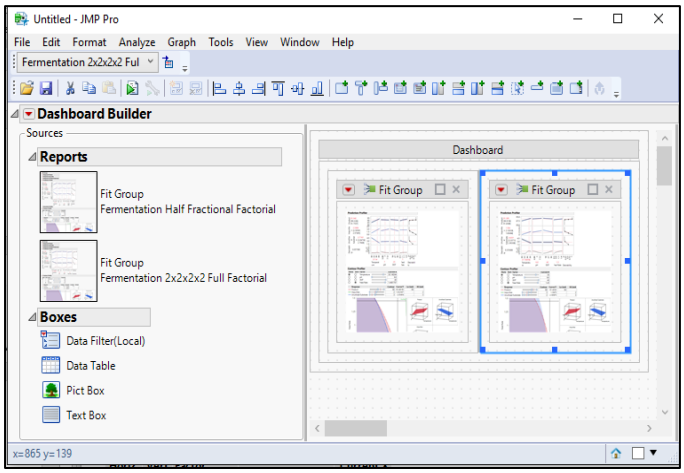
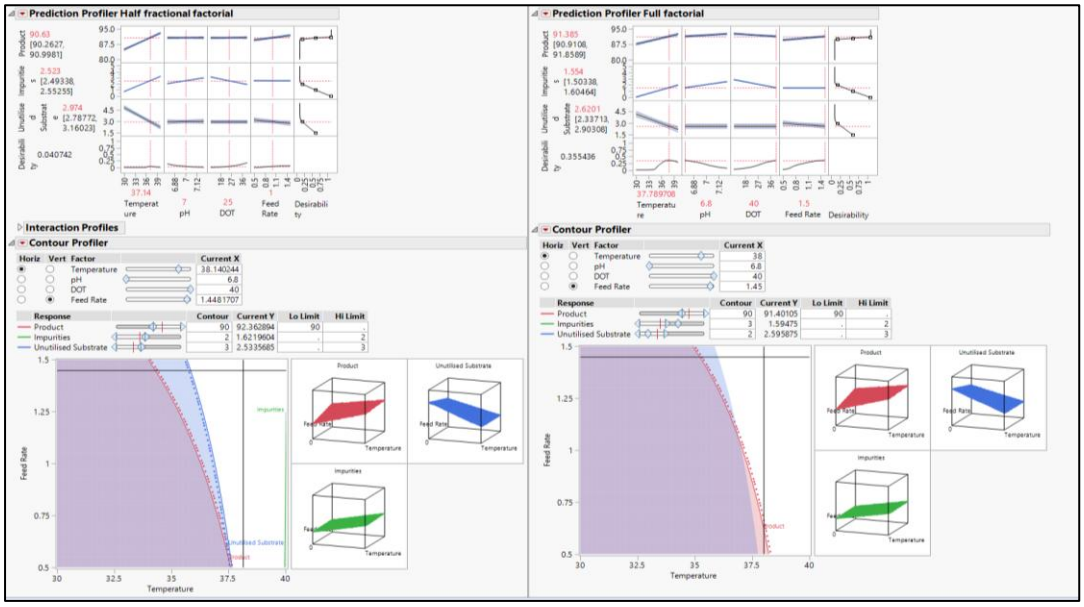


Figure 4: Comparison of a full factorial (20 experiments) and a half-fractional factorial (10 experiments) simulated designs.



In Figure 4 we are comparing two designs where the results have been simulated. In this case, near identical conclusions can be made – the main difference is that one design uses twenty experiments, whilst the half-fractional design uses half that number. With simulations, students have the opportunity to compare and contrast alternative scenarios. In real life where experimentation is more costly, we rarely have that luxury. The dashboard can be saved to a Journal and can then be recreated with one click.

We can use other situations to create learning opportunities and build experience rapidly.
With “Laboratory of Life”, we have to deal with unexpected sources of variation, which makes setting up and analysing designs more challenging. Figure 5 shows how we can set up a dashboard to surface two different types of variation, this time using the 2x2 template.

Figure 5: Using the 2x2 Template of Dashboard Builder

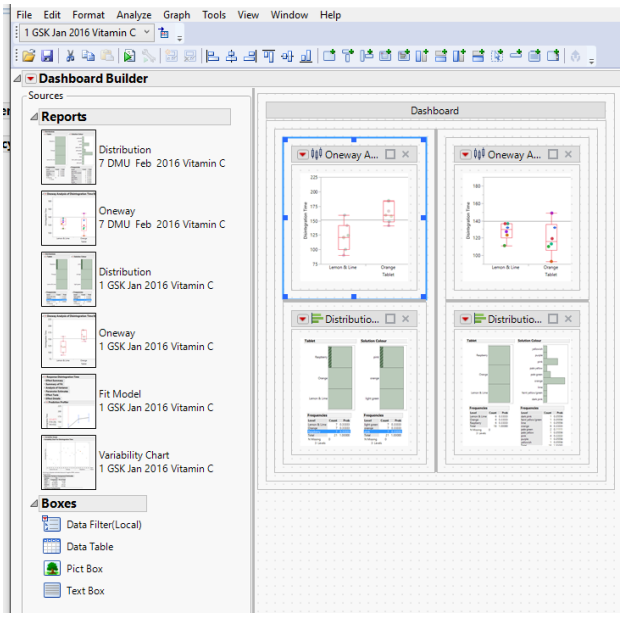


Figure 6: Comparison of Vitamin C Tablet Disintegration times across two different workshops.

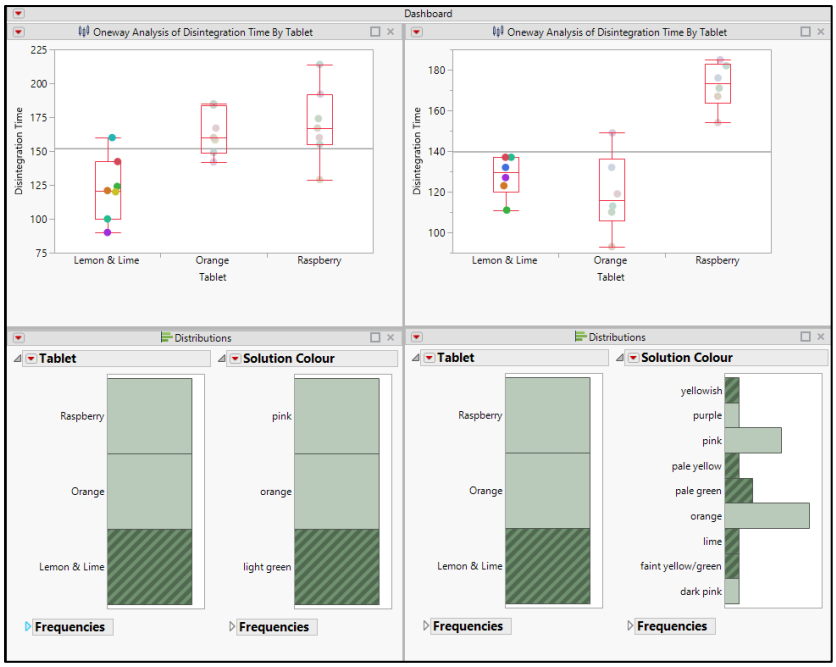


Figure 6 is a comparison of results from two different workshops carried out at different times and different locations. The top half of the dashboard compares Oneway Analysis. The overall mean of disintegration time for all three types of Vitamin C tablets appears to differ between the two workshops. Not only that, but the one-way analysis indicates teams from one workshop appear to conclude Lemon& Lime and Orange appears to be faster than Raspberry, whilst teams from the other workshop suggests Lemon& Lime is the fastest.

These findings should prompt further discussion about what else might have been changing between the two workshops, which wasn't being explicitly investigated. The discrepancy could be due to different batches of tablets, different water quality, different ambient temperature or even different practical experimentation capabilities of the teams that were investigating the orange flavoured tablets. These are all things that could be investigated in a follow-up study.

The distribution plot (bottom of Figure 6) also reveals a different sort of variation in the response measurement. On the left the analysts were in agreement that the lemon & lime tablets gave a light green solution. On the right hand-side, four different analysts described the same coloured solution in four different ways. This "unnecessary creativity" is a problem often encountered with free text fields and a pick list of response options may help in similar circumstances.

Another frequent implementation decision is where we have a fixed amount of resource and we have to decide what we can realistically explore. In one of the Tablet Disintegration modules we give the teams only enough resource to run eight experiments. How many factors should each team attempt to investigate? We encourage different teams to explore different numbers of factors. We can use the Dashboard builder again to help compare results, this time using the 2x1 template.

Figure 7: Comparison of a half fractional factorial and an eighth fractional factorial design for disintegration of Alka Seltzer tablets

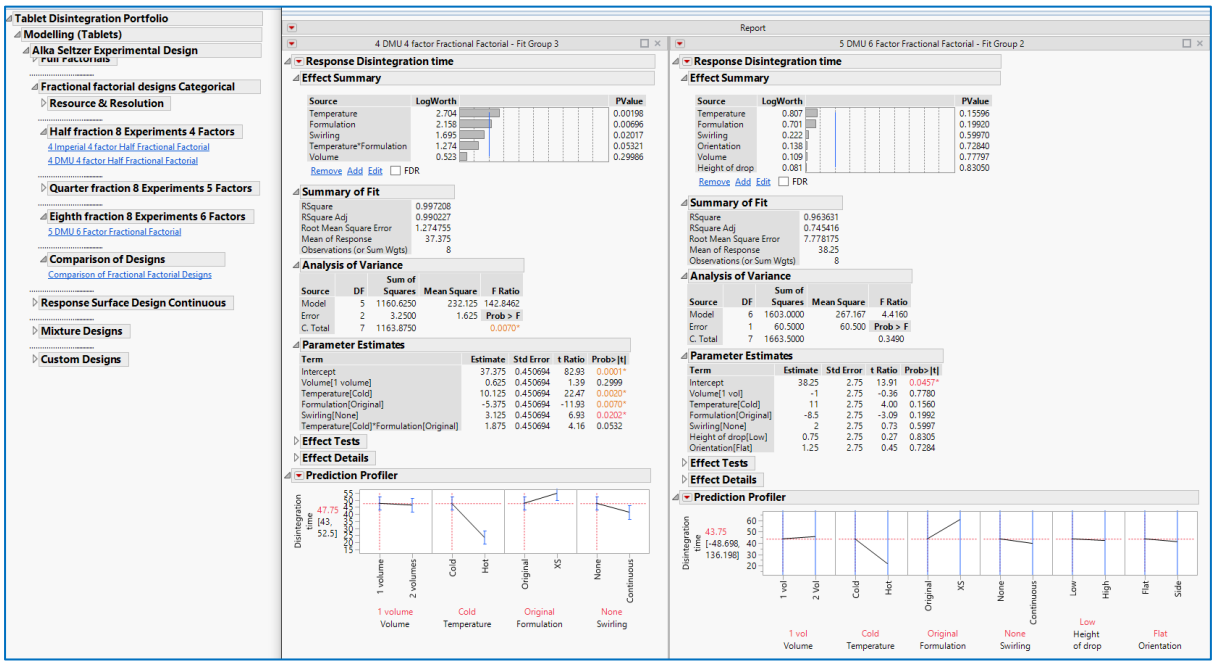


Figure 7 compares two different designs carried out in the same workshop. Both teams use eight experiments, but one team investigated four factors (a half-fractional factorial) and one team investigated six factors (an eighth fractional factorial). The team that looked at only four factors produced models with greater statistical significance. In contrast, the team that looked at six factors gained additional information on two factors which the other team have no knowledge about.

As teams across different locations perform the Tablet Disintegration studies, our overall process understanding increases and over time new factors, previously not investigated, now turn out to be important. The analogy here with what often happens in real technology transfers is startling. Unexpected findings emerge that are not due to process factors that have been studied and termed critical, but by lurking factors that emerge later on in the lifecycle as material input supplies and/or equipment changes.

The data collation challenges

Even this simple tablet disintegration experiment highlights problems of collating data. And these problems are predictable and ubiquitous across many organisations. Participants may not always name the factors or responses in the same way. They may use different units. They may use excel and merge cells and have footnotes and put text and numbers in the same cell. Left to their own devices, scientists can be notoriously creative in terms of entering and formatting data.

In both industry and academia we hit the same challenges when we collate data and attempt to derive learnings across different studies.

Early in research, factors may be chosen to investigate a specific problem. Later on new challenges may appear and new critical factors emerge for which we may not have collected the data in previous studies. But equally, we don't want to have to have the burden of collecting data we will never use.

What data should we collect and what data are we collecting unnecessarily? What additional data might we need to help retrieve different studies with different purposes or from different locations? When does it make sense to combine data across studies and when might we want to avoid doing this?

Accessing data

In the workshops, data is initially collected in many ways, on scraps of paper or in notebooks. Excel is frequently used to record data and excel tables are stored locally on laptops, or transiently on memory sticks. Whilst data preparation is a relatively unglamorous topic, establishing ground rules for data collection upfront is very important. Providing templates and examples of good data management to encourage standardised text entry reaps dividends if data is to be collated efficiently and information is to be shared within the workshop in a timely manner.

Applying these principles to help improve organisational memory at De Montfort University

As part of the MSc Quality by Design course at De Montfort University, all students have to complete a practical project related to formulation of drugs or cosmetics. As part of the coursework all students completed a pre-study report which outlines

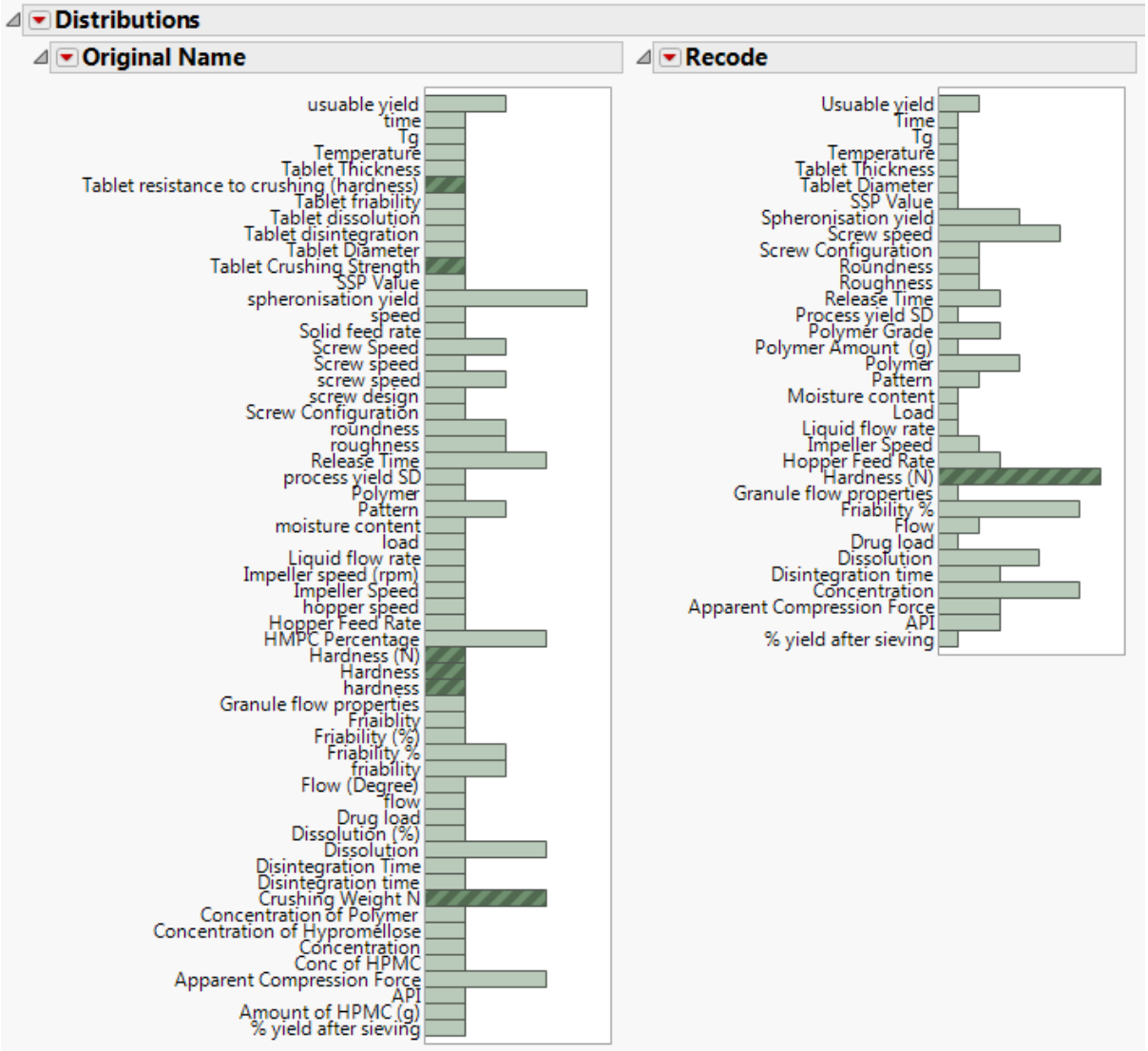
- the purpose of study
- the factors and ranges selected
- a potential experimental design in JMP.

Figure 8: The 2016 DMU MSC Quality by Design Project Portfolio

Formulate **ibuprofen** sustained release tablet	Use **electrostatic noise sensor** to analysis powder flow behaviour from a twin screw feeder
Use of electrostatic powder flow sensors on-line to monitor **powder flow**	Influence **of cyclodextrin** in the solid dispersion of dapsone in PVP to enhance the API solubility
Screen material attributes factors affecting **cosmetic skin care emulsions.**	Investigate mpact of formulation variables and process parameters on ibuprofen granules prepared by the **high shear wet granulation** method and by wet granulation method using **twin screw extrusion**
Efect of **screw configuration** on quality of granules produced	Investigate **the effect of polymers and/or /surfactants** on increasing solubility of stable forms of *FFA-NIC and FFA-TP* pharmaceutical co-crystal formulations
Improve **Solubility and Stability** of Poorly Water Soluble Drug with Mesoporous and Polymer Materials	Optimise **insulin nanoparticle production** and quality for coating stainless steel microneedles
Optimise **fibre drug production** and quality for microneedle coating	Enhance the Solubility **of Itraconazole** using **Continuous Hot Melt Extrusion**
Enhancing the Solubility of **Piroxicam** using Continuous Hot Melt Extrusion	
Evaluate the effects of polymers and surfactants (as crystallisation inhibitors) on the phase transformation of **cocrystals** into its parent drug **Flufenamic acid**.	

Twelve experimental designs were actually subsequently executed in the laboratories and the data was collated. Initially there appeared to be a total of 58 columns of factors and responses. On closer inspection there were many instances of inconsistencies of names relating to the same factor or response. For example, the response Hardness (N) had six different names (see Figure 9). This makes it more challenging to combine datasets. Once these names had been aligned (using **Cols>Recode**), the total number of columns dropped from 58 to 34 (15 factors and 19 responses).

Figure 9: Distribution plot showing before and after recode



Standardisation of column headings means that the meta data and data from the studies could now easily be collated into one table using **Table>Concatenate.** The collated data was then interrogated or visualised in a number of ways (Figures 10 and 11). Understanding what studies have been performed previously and what constitutes a good outcome or unfavourable outcome is very useful in helping design future research investigations.

Figure 10: Six out of the twelve studies investigated a variety of ranges of the screw speed process parameter.

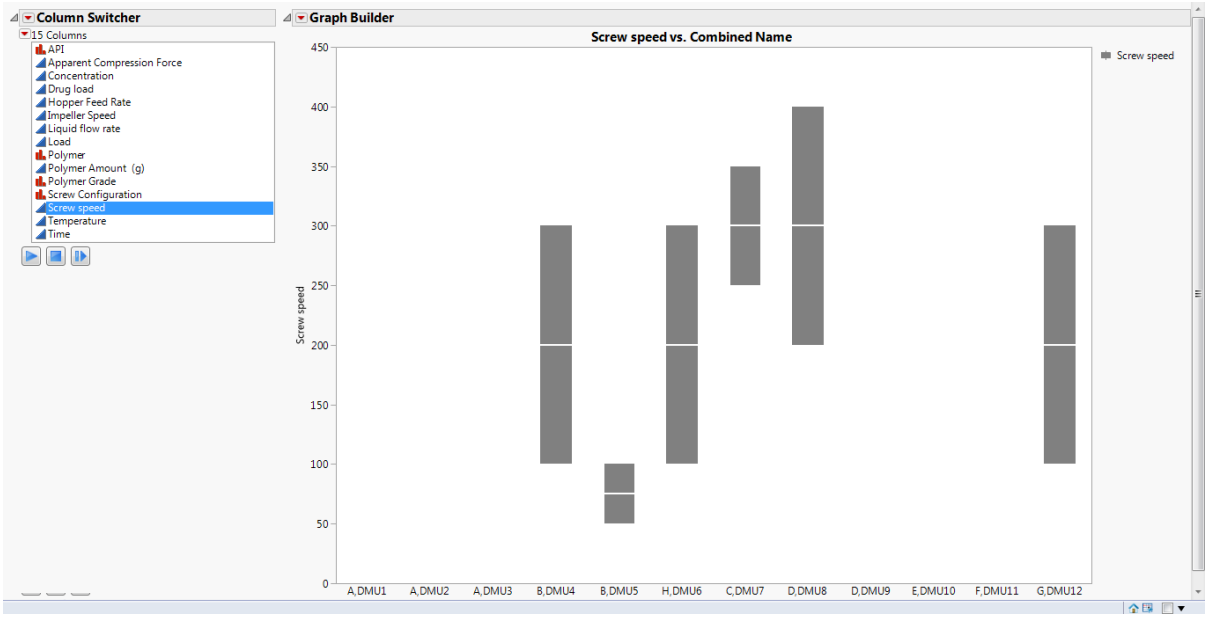
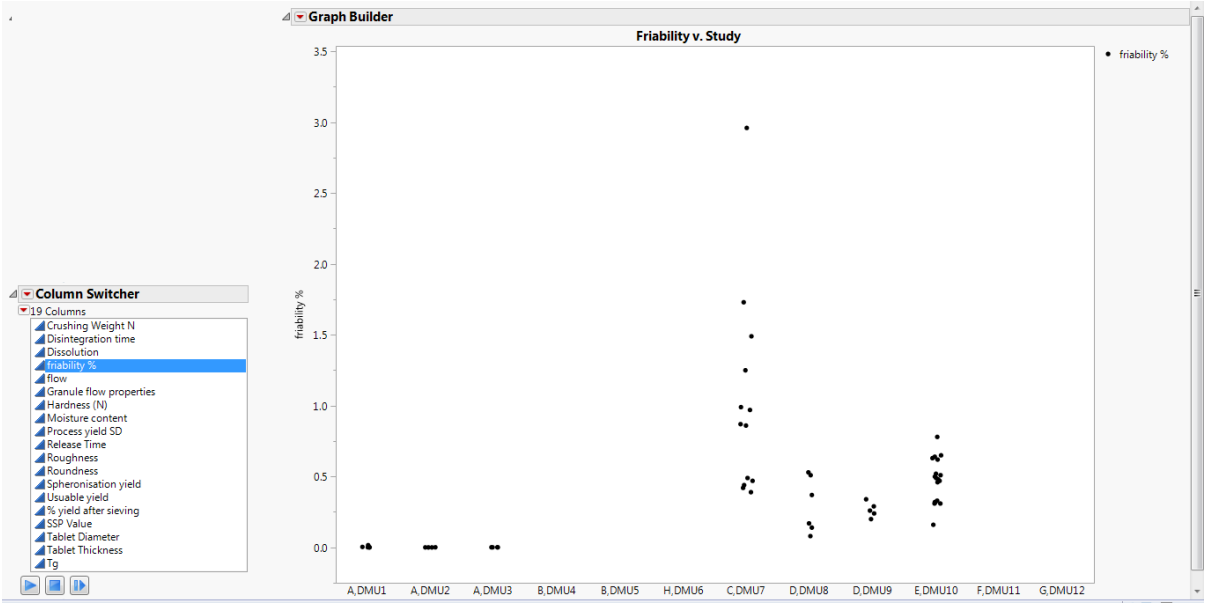


Figure 11: Seven out of the twelve studies measured % tablet friability



Generating an Excel Workbook from JMP Data Tables

As an aside, another new feature of JMP13 is the ability to save a collection of JMP data tables into a single Excel Workbook.

Open all the tables that need to be saved with JMP. The first thing that you will need to do is to make sure that all of the tables that you wish to save are open within JMP. By default, all of the visible data tables will be included in the list. If you wish to exclude a table from the Workbook, just uncheck the leftmost check box.

Then select "**View -> Generate Excel Workbook**". A dialog appears which enables customisation of the Excel Workbook that will be created. For example each Excel worksheet can be renamed by clicking in the **Worksheet Name** edit field and change the name to something fit for purpose.

Query Builder

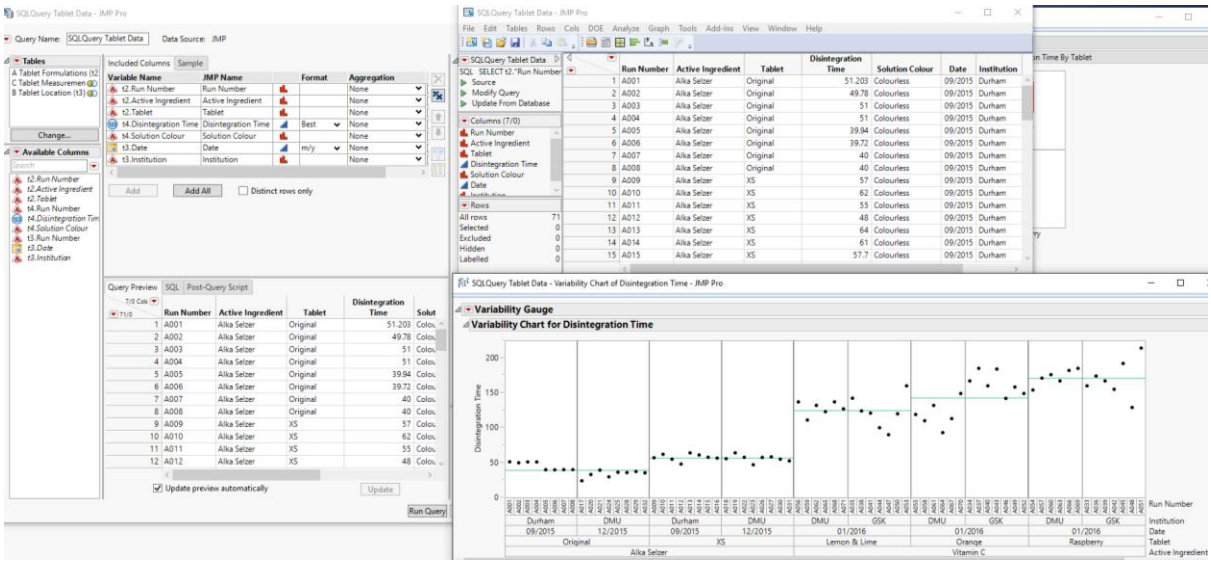
In JMP13 Query Builder provides new functionality to build a query containing one or more JMP data tables (**Tables>JMP Query Builder**).

In the example here I have opened three sets of data that I wished to join. I have selected **Tablet Formulation** table as the primary table and Tablet Measurement and Tablet Location as the secondary tables.

Use **Preview Join** to identify any duplicated entries or unwanted columns.

Now **Build Query.** It is now possible to add the desired columns to the query in the required order and deselect any duplicate columns. **Save Query** and **Run Query** to produce the output table (Figure 12). If I then construct a graph, I can save the script to **Post-Query Script**

Figure 12: use of Query Builder to build JMP tables and automatically crate graphs

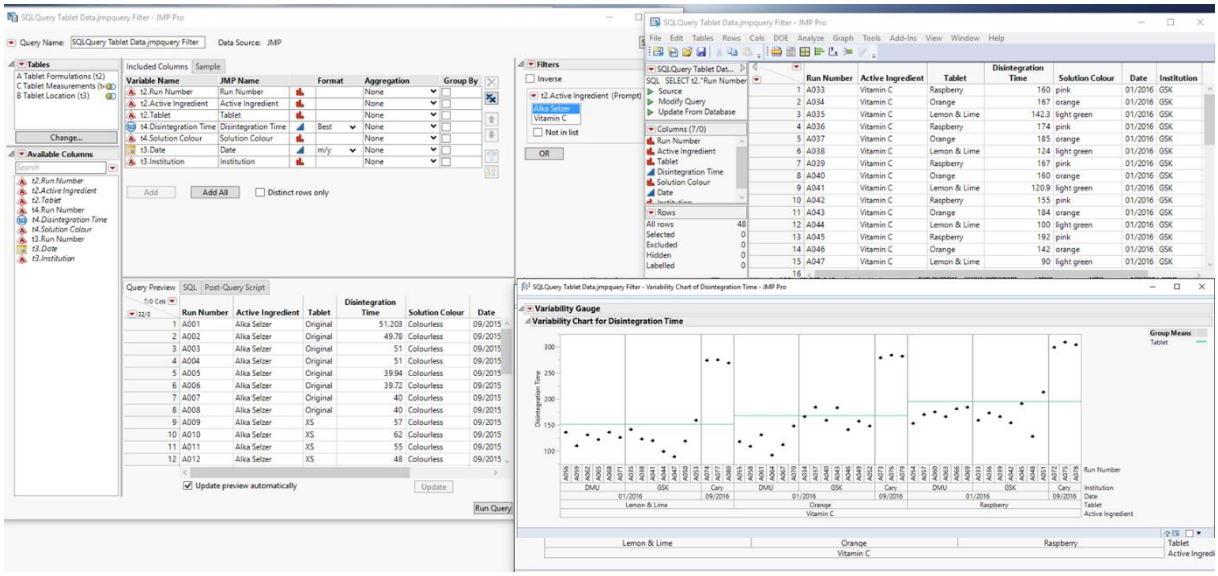


Note that in the new data table that is created the table panel on the left hand side now contains two table scripts, **Modify Query** and **Update from Database.**

I can add a filter to the script to allow me to choose between Alka-Seltzer and Vitamin C and save as a new script.

If more data is entered into the original tables, then by clicking on **Update from Database** table script, the data table and associated graph can be updated with ease (see Figure 13).

Figure 13: Updating using Query Builder when new data becomes available



Text Explorer

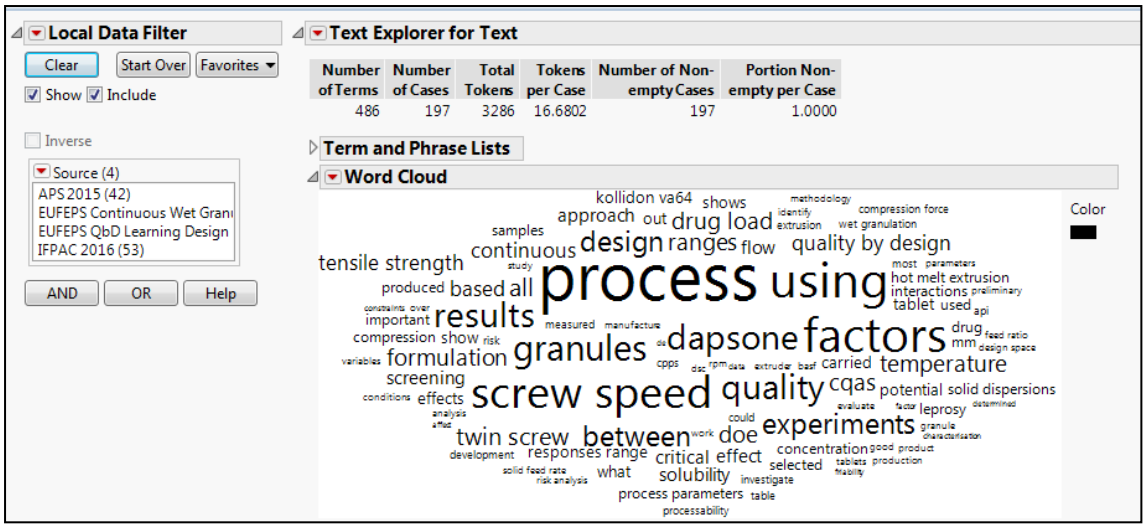
Not all information can be gleaned from tables and models. There is often a huge amount of associated contextual information that is captured in word documents and PowerPoint presentations. The free text has not always been easy to analyse in the same way we can analyse spreadsheets.

Figure 14: External presentation posters from De Montfort University

Figure 14 shows four posters describing hot melt extrusion and continuous wet granulation processes investigated at De Montfort University by undergraduates.
In order to illustrate this for this presentation, I have cut and pasted the free text from the four posters into a JMP table (via Excel) and analysed the overall findings using the new Text Explorer incorporated into the JMP13 Early Adopter.
Each sentence or individual body of text forms a “document” and the entire text data set is referred to a “corpus”. The Text Explorer then identifies “terms” (the basic unit of analysis – individual words and phrases) and orders these by frequency or alphabet. It is possible to throw out low frequency or irrelevant terms by assigning them as stop words.

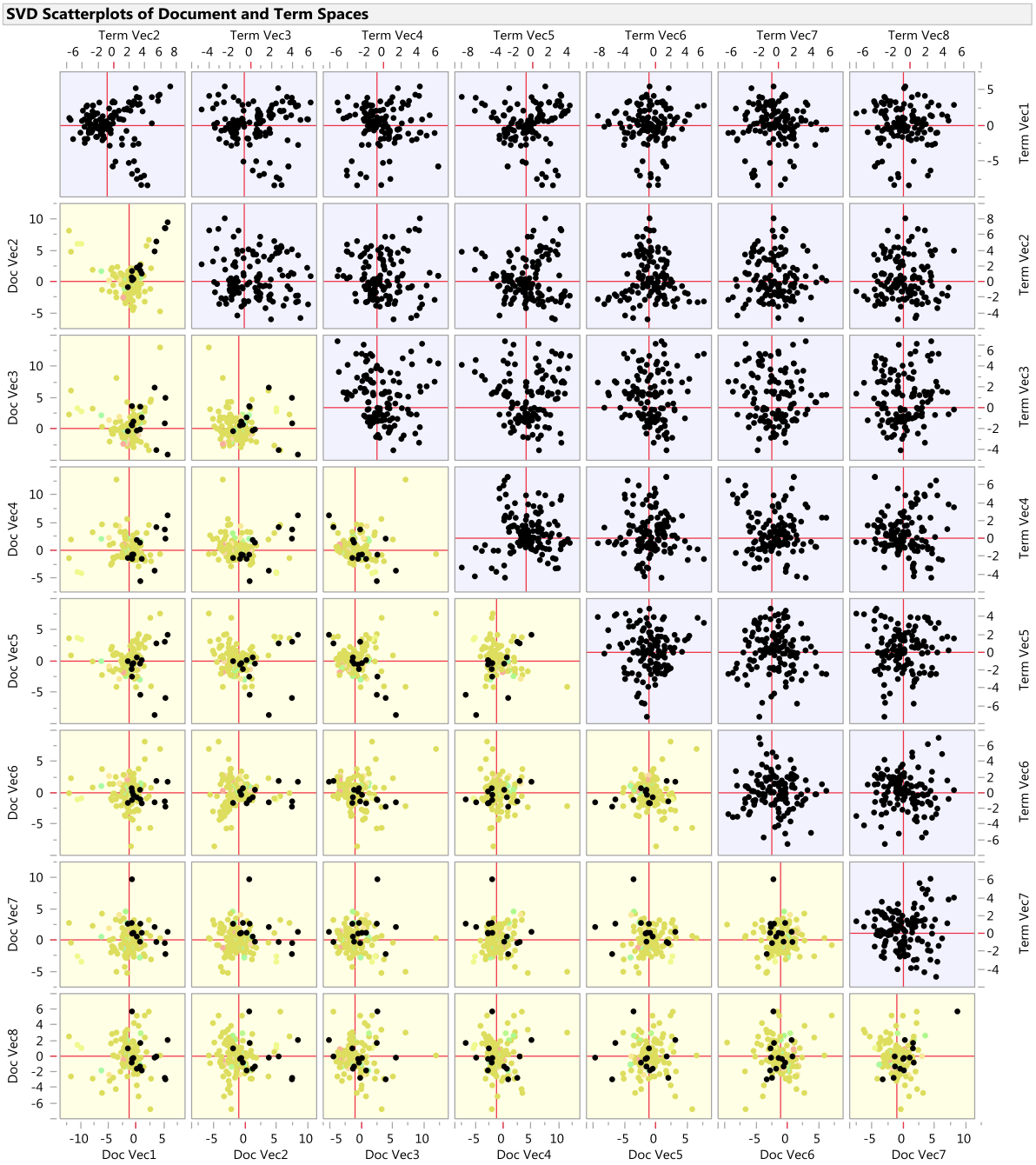
We can then produce a word cloud of all four posters. If we then add a filter we can rapidly summarise and contrast the content of each poster, as the cloud is updated dynamically (Figure 15)

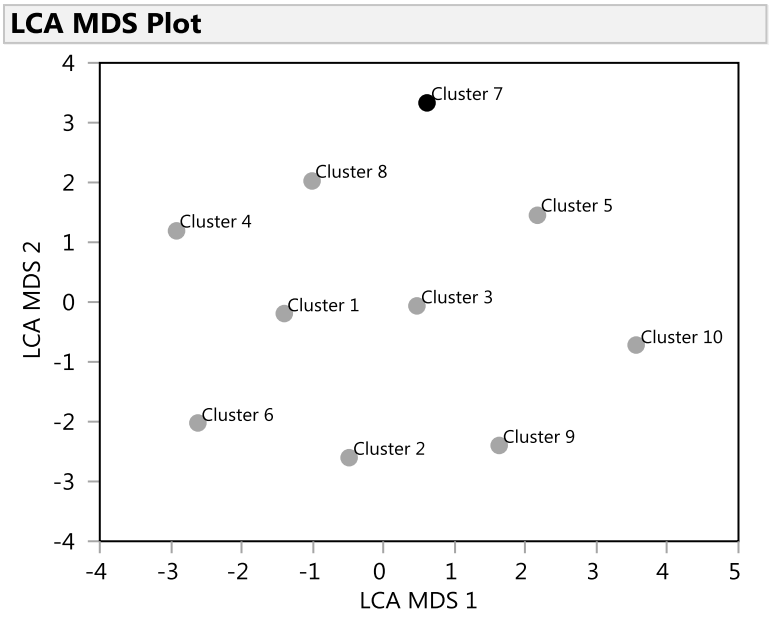
Figure 15: Word cloud summarising Text in poster presentations



The number of terms used in this example is relatively small, but it is still possible to do multivariate analysis and clustering using Latent Semantic Analysis (Figure 17).

Figure 16: Visual representation of word clusters





Cluster 7 selected rows that are to do with investigating the solubility wet granulation of Ibroprofen

The aim of this study was to investigate wet granulation manufacturing process by using twin screw extrusion (TSE) for the production of Ibuprofen granules using science and risk based approach [24]

Flow process of wet granulation tablet manufacture. [26]

DoE is the best approach to investigate the extrusion process with as few experiments as possible. . [45]

The methodology applied throughout this project (based on quality by design principles) has enabled the researchers to learn and use science and risk based approach to efficiently develop and manufacture pharmaceutical granules using extrusion process in a very short time. [62]

J. Vercruysse, D. Córdoba Díaz, E. Peeters, Fonteyne, U. Delaet, I. Van Assche, T. De Beer, J.P. Remon, C. Vervaet. Continuous twin screw granulation: Influence of process variables on granule and tablet quality. European Journal of Pharmaceutics and Biopharmaceutics, 2012; 82(1), 205-211. [63]

A formulation of ibuprofen 200 mg was evaluated using Quality by Design principles to study the effect of Twin Screw Wet Granulation (TSWG) factors on granule quality and characterise the final compressed tablet. [71]

It is clear that the compression force is the most significant factor that influences the 3 responses. process factor and would have been better included as a quality output measurement. [84]

The tabletability of the granules was studied further to examine the formulation and process effects. [86]

Statistical analysis of the data showed that compression force had the greatest effect upon tablet quality and had perhaps masked the smaller effects of the process settings for this formulation. The formulation was surprisingly robust and would allow extension of the design space for future work. [100]

For contrast, Cluster 6 selected rows that are to do with investigating the solubility of Dapsone using a polymer and hot melt extrusion.

The aim of this study was to investigate wet granulation manufacturing process by using twin screw extrusion (TSE) for the production of Ibuprofen granules using science and risk based approach [24]

Flow process of wet granulation tablet manufacture. [26]

DoE is the best approach to investigate the extrusion process with as few experiments as possible. . [45]

The methodology applied throughout this project (based on quality by design principles) has enabled the researchers to learn and use science and risk based approach to efficiently develop and manufacture pharmaceutical granules using extrusion process in a very short time. [62]

J. Vercruysse, D. Córdoba Díaz, E. Peeters, Fonteyne, U. Delaet, I. Van Assche, T. De Beer, J.P. Remon, C. Vervaet. Continuous twin screw granulation: Influence of process variables on granule and tablet quality. European Journal of Pharmaceutics and Biopharmaceutics, 2012; 82(1), 205-211. [63]

A formulation of ibuprofen 200 mg was evaluated using Quality by Design principles to study the effect of Twin Screw Wet Granulation (TSWG) factors on granule quality and characterise the final compressed tablet. [71]

It is clear that the compression force is the most significant factor that influences the 3 responses. process factor and would have been better included as a quality output measurement. [84]

The tabletability of the granules was studied further to examine the formulation and process effects. [86]

Statistical analysis of the data showed that compression force had the greatest effect upon tablet quality and had perhaps masked the smaller effects of the process settings for this formulation. The formulation was surprisingly robust and would allow extension of the design space for future work. [100]

The Text Explorer has the potential to be a really valuable mechanism for identifying additional meta data to be more formally collected in structured spreadsheets.

Conclusion

Building an organisational memory is a key advantage to retain knowledge and exploit previous data-derived information, within and across organisations. The more data we have and the more different sources of data we have, the more we require more efficient methods of collating and retrieving information. Whilst the Tablet Disintegration study appears at first sight to be a trivial example, it highlights the same challenges and opportunities associated with much more complex and multi-stage processes. We have shown similar approaches can help organisations like De Montfort University build an accessible portfolio of experimental design studies used to support Quality by Design ways of working and help students access prior data more easily.

Why stop here? What if we could compare the De Montfort formulation designs and data with those from other academic institutions in the UK. And what about world-wide?

This should enable us to use prior information more effectively and generate new knowledge more efficiently. Imagine what we could achieve if we could be more efficient at retaining what we do know and become more effective at researching what we don't know!

Acknowledgements

I would like to thank the following for allowing me to share insights for organisational learning.

- Dr Walkiria Schlindwein and the MSc Quality by Design students at De Montfort University
- Prof Patrick Steel (Durham University), Dr Dennis Lendrem and Matt Linsey (Newcastle University), Prof Dave Woods (Southampton University) and Dr Stephen Chambers (SynbiCITE- Imperial College)
- I'd also like to thank David Burnham of Pega Analytics, Paul Nelson of PrismTC and Iain Grant of Phruitful for their advice and support.