



## WHITEPAPER

# Facilitating faster decision making and improved productivity in the pharmaceutical industry

Santi Dominguez Vivero - Mestrelab Research CEO

Background - The Problem.....	1
Mestrelab Research.....	1
Our Solution:Automatic structure verification and spectral aware databases.....	2
Some of our customers.....	6
Case studies.....	7
Benefits.....	9



## **BACKGROUND – THE PROBLEM**

The pharmaceutical and biotechnology industries are currently facing very significant challenges which are resulting on major organizational and operational changes. The trend is for pharmaceutical companies to outsource many of their Research and Development operations to third parties with specific specialities and areas of expertise, in order to focus their own resources on areas where they are in a position to maximize value.

As a result, where many operations such as data mining, information sharing and decisions on chemical entities used to have to be carried out at an internal company level, now these same functions need to encompass the entirety of the supply chain and all partners involved in the company's R&D effort. In addition, it is desirable to standardize data formats and data handling as much as possible, to keep tight controls on quality, particularly when compounds are being transferred between organizations and to optimize the sharing of information to maximize productivity.

A further challenge, which affects, among others, the domain of Analytical Data, is the reduction in the resourcing of internal support teams. In the case of Analytical Data, this means that, whilst progress in hardware capabilities and huge steps forward in automation allow us to generate very large amounts of analytical data in high throughput, a bottleneck is being created by the capability of Analytical Departments to get eyes on data. In this situation, overseeing also the data and proposed chemical structures proposed by partners outside the organization, and typically received in large numbers, is becoming harder and in most cases has been abandoned as an achievable objective.

## **MESTRELAB RESEARCH SL**

Mestrelab Research is a Spanish software house specializing in the development of tools for analytical data processing and management, which has been providing solutions to the pharmaceutical, biotechnology and chemical industries since its foundation over six years ago. Our focus is to support open access laboratories with a set of tools which maximize their productivity and ability to make fast decisions and minimize the labour intensive efforts needed to fulfil these workflows.



## OUR SOLUTIONS: AUTOMATIC STRUCTURE VERIFICATION AND SPECTRA AWARE DATABASES

Our solutions are based on a clear set of premises. If these apply to your organization, then our software solutions should improve your productivity and results:

- Chemist time is a very valuable resource best used designing and carrying out reactions.
- Typical Analytical Chemistry groups no longer have the capability to eyeball all analytical data generated both internally and externally.
- Having two scientists do the same job twice is a waste of resources and a common drain on productivity.
- Information available somewhere in the organization should be reachable and 'minable' by all other researchers to optimize results.
- Compound identity and integrity are fundamental pillars of a productive R&D operation and are essential at all stages of the R&D process, internal and external.
- Software solutions should be used whenever possible to automate routine workflows in order to allow companies to release highly skilled human resources for more valuable activities.

In order to address the above points, we have developed a flexible set of software tools, which are focused on these premises and on the following issues:

Identification and validation of the structure of early compounds synthesized in the drug discovery process are fundamental to the success and productivity of drug development processes. Mistakes in structure identification can be very costly, generating further errors and time wasting along the line. Valuable resources could be deployed on the basis of erroneous information with disastrous consequences to productivity.



Our solutions handle these problems in full automation, by providing a fully integrated, automatic software environment ideal for the open access laboratory. Routinely acquired analytical data, such as NMR and LCMS, can be handled automatically, independently of the data format and the hardware manufacturer the data have been generated by, eliminating the need for chemists in an open access environment to go through steep learning curves for a variety of software packages. One single software package handles all.

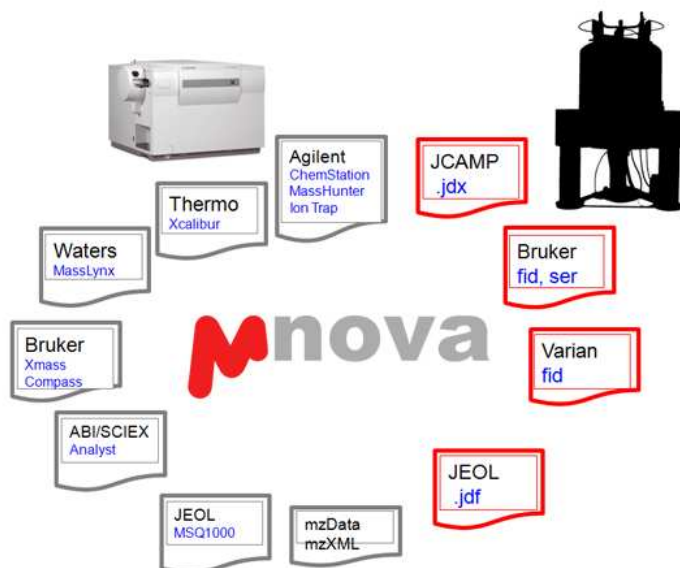


Figure 1

The software will accept a flexible array of analytical data, being able to work with any combination of 1H NMR, 2D NMR correlations and LCMS or LCMSMS data, and with single or multiple structural proposals for each set of analytical data. All information available is used to confirm, or discard, each structure proposal, and to rank multiple possibilities by their likelihood of being correct based on the available analytical data. Once this analysis has been carried out, reporting can also be handled fully automatically, resulting in more significant time savings. Purity and sample concentration or array strength are also analysed automatically and can be included in the reports.

The consequence is that, in a standard workflow where analytical services do not have the ability to check all analyses performed in open access, their resources can be focused on a percentage of the data which are flagged by the software as being inconsistent or presenting problems, whilst being able to place more reliance on the conclusions arrived to by open access users or Contract Research Organizations. Decisions can be taken faster, with errors minimized and structure integrity assured and with analysis bottlenecks eliminated.



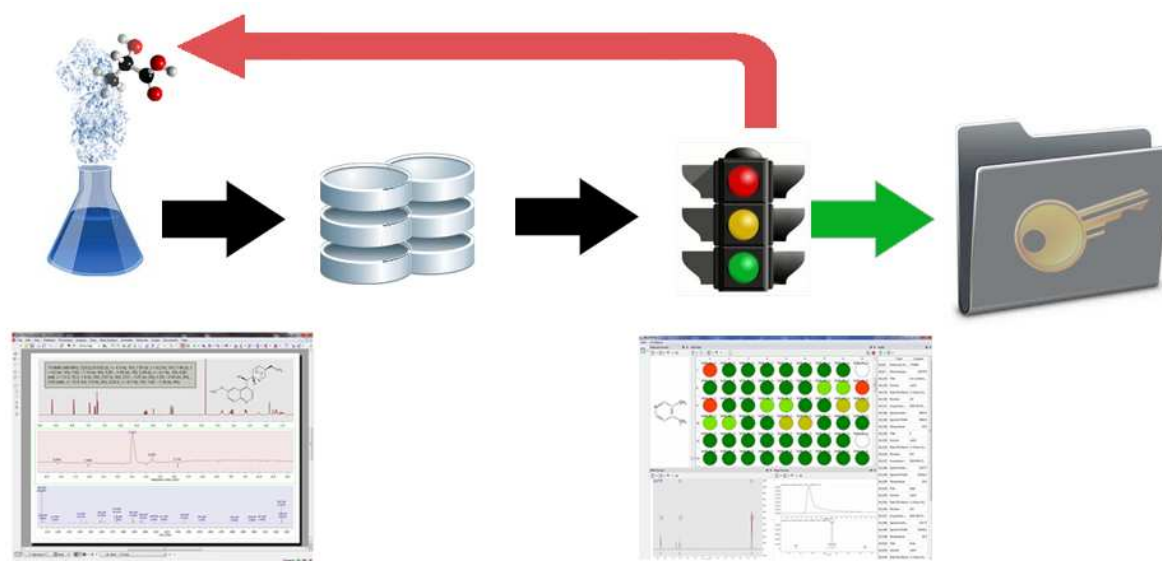


Figure 2

Chemists submit synthesized compounds for registration via our spectra aware database, which acts as a gateway. Analytical Department pick up database with all spectra run in e.g last week and run automatic verification against DB. Passes go through to registration, failures go back to chemists or are further analysed by the Analytical Chemistry Department.

Data mining is another crucial area when analysing samples and making structural decisions, and when trying to gain information about a chemical entity being analysed. Our solution incorporates a data storage and mining system which allows the full organization to leverage their chemical knowledge, and which can easily be shared with external partners. It is now possible, through the whole supply and research chain, to ask questions such as: Has anyone in the organization seen a compound with these analytical features before? What do we know about such a compound? Because the data mining can be driven by analytical features alone, it is not even necessary to know the structure of the compound of interest to find relevant information. Any scientist in the open access laboratory can become an analytical expert by leveraging knowledge gained by other scientists across the organization.

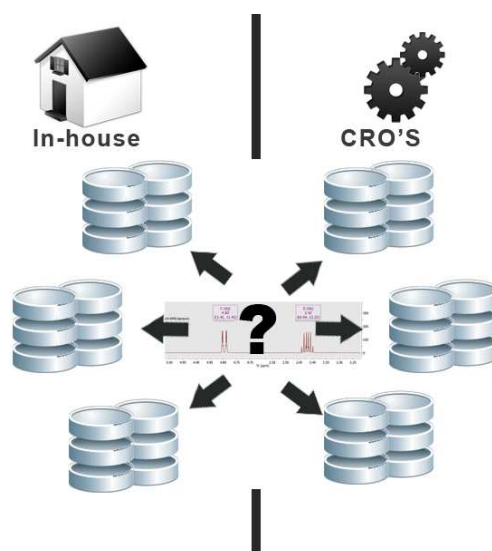
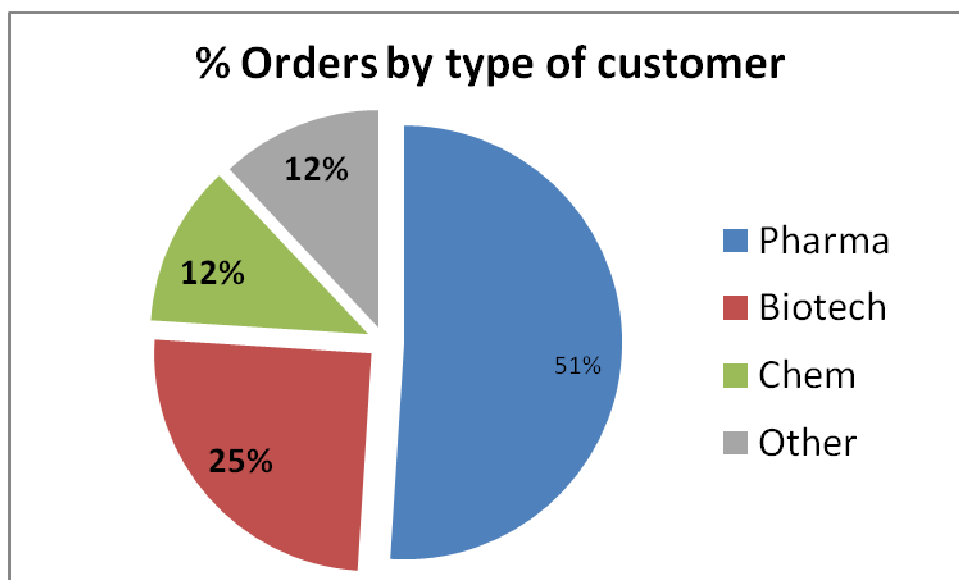
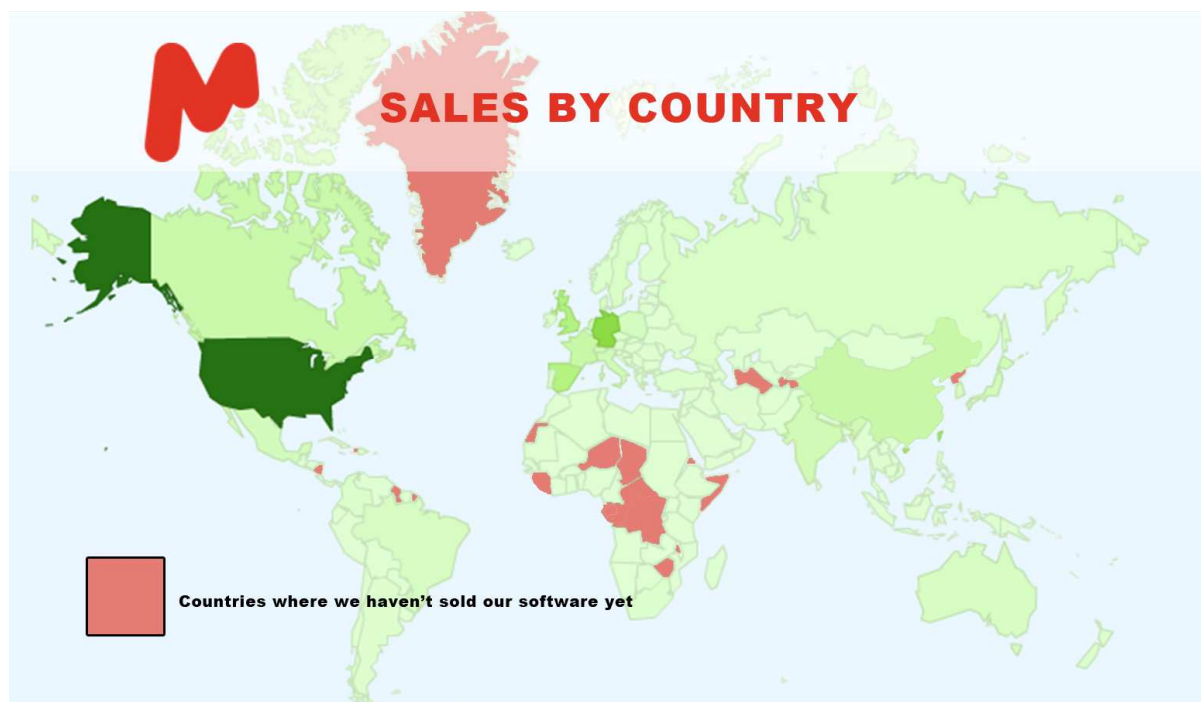


Figure 3

Structure elucidation of unknowns and impurity characterization can be very labour intensive activities. These data mining capabilities ensure that no information from previous elucidation efforts is lost and that all information existing in the organization is used as an input to each new elucidation effort.

## SOME OF OUR CUSTOMERS

Mestrelab has become a technology partner to many of the largest and most successful pharmaceutical and biotech companies, working specifically on the handling of analytical data in Chemistry and Medicinal Chemistry Departments, and on automating routine decision maker to release resources.



## CASE STUDIES

The software is being used in many different ways by our partners. The tools offer the flexibility to tailor them to what are perceived to be the desired work and data flows in the organization. Below, we include some examples of how these tools are being used to optimize productivity in critical processes:

### Case A: Open access and compound registration

Customer A has a large Small Molecule Chemistry department (circa 150 chemists). A full open access laboratory is operated, where each chemist submits his/her own samples for analysis and studies the results, generating a set of reports and, upon confirmation of the expected molecular structure, submits the compound for registration. The compound then proceeds further down the line for further analysis, activity studies, etc.

With the implementation of our tools the initial chemist workflow has remained unchanged. At the point of registration, however, all structure proposals are validated in batch mode, with any inconsistencies with the analytical data being automatically refused for registration and referred to the small Analytical Services department. This department can then visually inspect the data and confirm data integrity, referring again for registration. Should the visual inspection result in a negative result, the Analytical Services department uses the data mining capabilities to gain crucial information about compounds with similar structures which have been synthesized before, in order to achieve a fast structure elucidation, which is shared with the chemist.

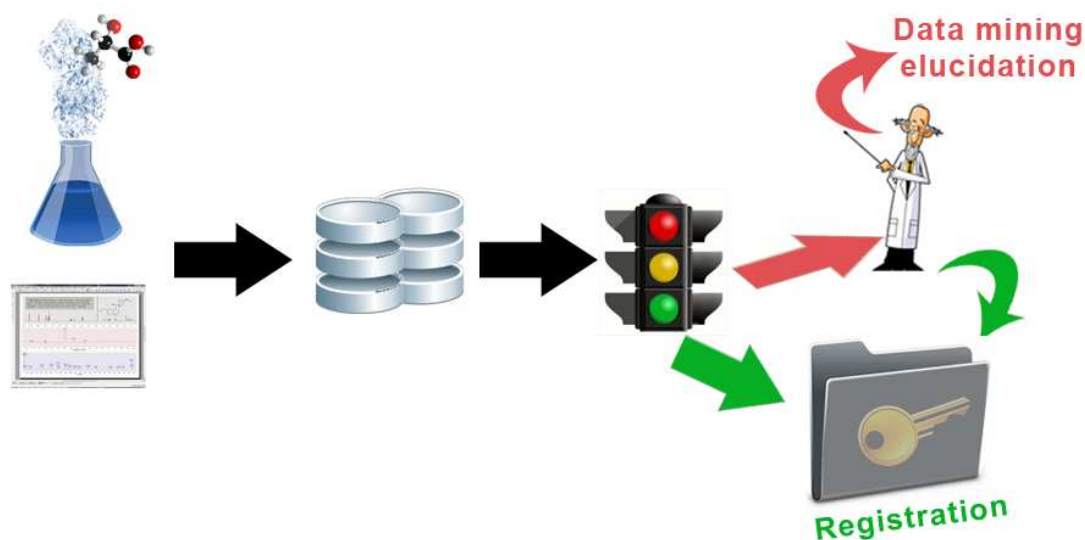


Figure 4



## Case B: Ensuring quality from external partners and managing partnerships

Customer B has a wide network of CROs collaborating in the effort to synthesize early drug candidates. Traditionally, these CROs have provided batteries of compounds with structural identity validated and supported by a PDF report for each one of the compounds. However, Customer B does not have sufficient resources with analytical chemistry expertise to inspect the PDF reports in order to validate structural identities. In addition, by using PDF for data exchange, valuable analytical information is lost and the ability to data mine analytical data is not available. Also, any potential tampering with the data cannot be identified.

A new process has been introduced by which CROs providing early leads to Customer B generate a set of reports in Mestrelab's software environment. Upon receipt by Customer B, and prior to compound registration, all analytical data are evaluated automatically in batch mode, with around 2% of inconsistencies being identified on average (this varies significantly between different CROs). In addition, purity is also evaluated automatically to ensure the compounds sent comply with the standards agreed between Customer B and the CROs.

Impure compounds or compounds of uncertain identity are then automatically referred to the Analytical Chemistry department for further analysis, including potential running of analytical data in-house. Procurement is informed so that decisions can be taken on the remuneration of the CRO for wrongly characterized or impure compounds, and information on the quality of compounds provided by each CRO on the basis of structural integrity and purity is stored as background information for future partnership negotiations.

Good quality, correctly characterized compounds are automatically registered and proceed down the R&D path.

In this setting, over 1,500 chemical entities have been analysed overnight on a given day.



Figure 5

## THE BENEFITS

The benefits afforded by these tools are many fold:

- Ability to handle compounds and samples seamlessly, both internally and with external partners, in one single software environment.
- Quality assurance and guaranteed data integrity throughout the Discovery and early Development process.
- Leveraging of spectral knowledge across the organization and its partners, optimizing the flow of information to the decision maker and ensuring faster, more accurate structural decisions and more reliable information further down the line.
- Elimination of analysis bottlenecks and ability to handle, with higher quality, a much larger volume of compounds or leads with full information.
- Visibility across the organization with a new, analytical data centric perspective, based on grouping and mining chemical entities by their spectral characteristics.
- Optimal productivity through automation, including the ability to fully automate data and sample flow throughout the organization on the basis of unaided decisions taken at crucial points

