Agile Implementation of Automated Processes Using the BIOVIA Pipeline Pilot System

Elyette Martin, Antonio Castellon, Mark Bentley, Pavel Pospisil

PMI R&D, Philip Morris Products S.A., Quai Jeanrenaud 5, CH-2000 Neuchâtel, Switzerland

Introduction

A big part of the work performed by scientists in chemical, pharmaceutical, or cosmeceutical industry research & development (R&D), including Philip Morris International (PMI) R&D, comprises the transformation of raw measurement data into processed data, ready for analysis. For example, one of the major activities for analytical chemists at PMI R&D is to characterize and quantify the chemical composition of aerosols generated by novel heated tobacco products. A variety of methodologies with gas chromatography coupled to mass spectrometry are employed, using instruments from different vendors with different spectral deconvolution and processing software. Hence, a complex software platform containing pipelining processors that help to integrate and automate these processes was developed, thereby enhancing the speed and accuracy for correct compound identification and providing instant reporting. Many elements of this platform were developed and implemented successfully using BIOVIA Pipeline Pilot (PP; Dassault Systèmes BIOVIA, BIOVIA Pipeline Pilot, Release 2017, San Diego: Dassault Systèmes, 2017). Here we present what protocols were used to process structures found in spectral libraries (standardization, boiling point calculation), how SMART-coded rules helped to select appropriate internal standards for semi-quantification, which components were used to cluster chemicals, how the protocols were made available to scientists (web ports), and finally, how results were reported in user-defined formats. In addition, commonalities for the potential use of many of these PP protocols by the drug discovery industry will be demonstrated.

Agile Scrum project management methodology

In order to understand user needs for the automated system properly and to develop it efficiently, we used the Agile Scrum project management methodology. After a first meeting to define user needs, a product backlog is created. From this backlog, the developers and users pull out working tasks together and meet each week for 20 minutes to share their progress and adapt to possible changes in product specifications.

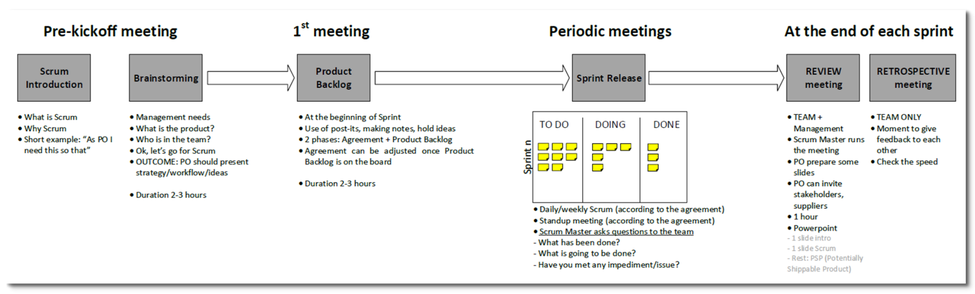


Figure 1. Scrum project management

CASI process example

To demonstrate Agile Scrum implementation, the example of Computer-Assisted Structure Identification (CASI) is shown here. CASI is a complex process developed at PMI to automate and enhance the process for structural identification of compounds present in heat-not-burn tobacco product aerosol[1]. Identification of aerosol compounds were done using the two-dimensional gas chromatography with time-of-flight mass spectrometry (GCxGC-TOFMS).

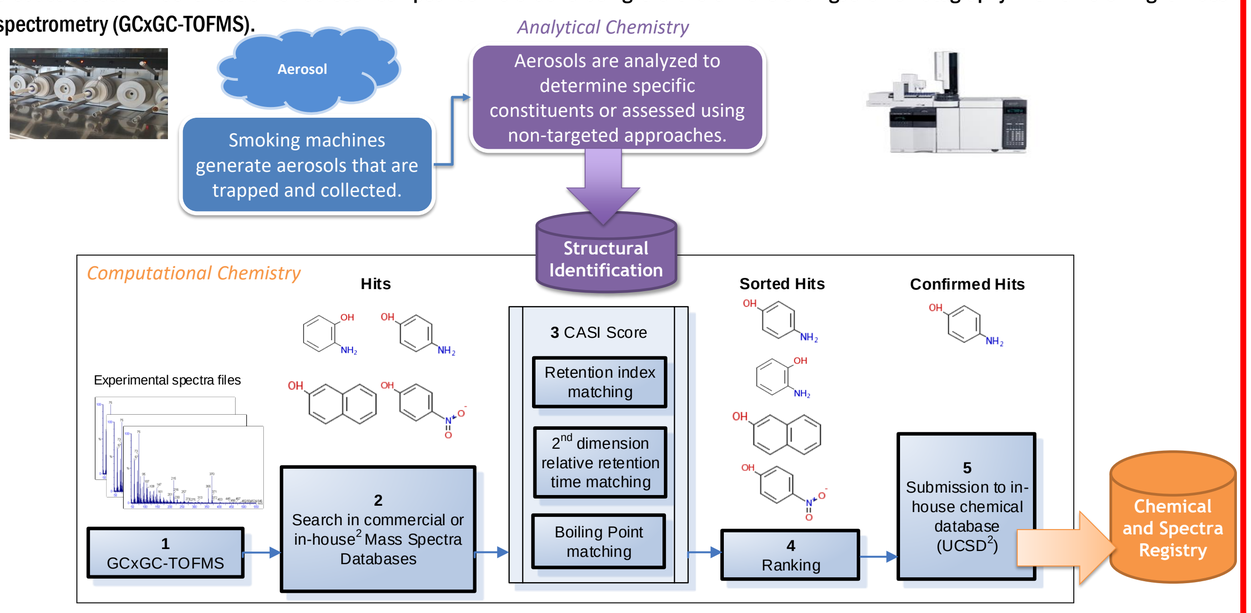


Figure 2. Enhanced identification of aerosol constituents using the CASI platform with GCxGC-TOFMS analysis.

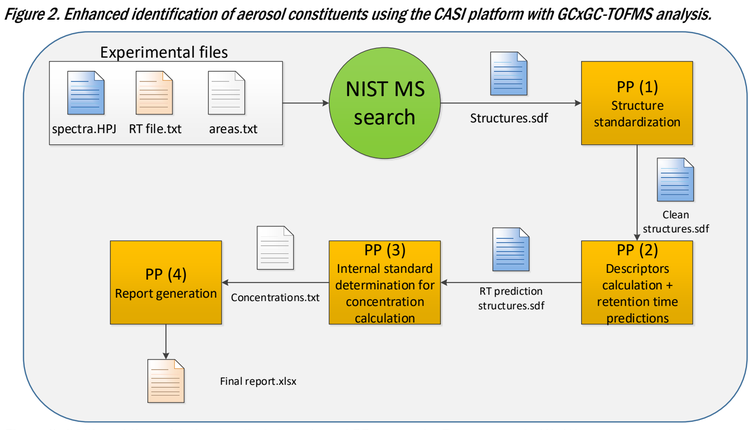
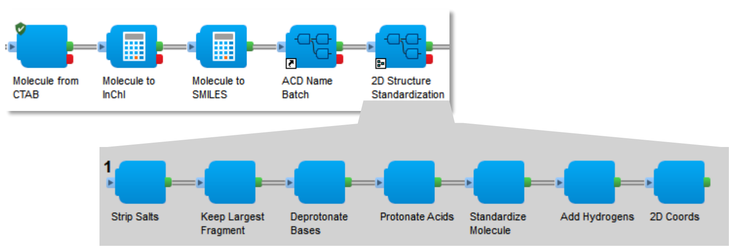


Figure 3. CASI workflow with the detailed view of used PP protocols. Four protocols (1)-(4) are presented.

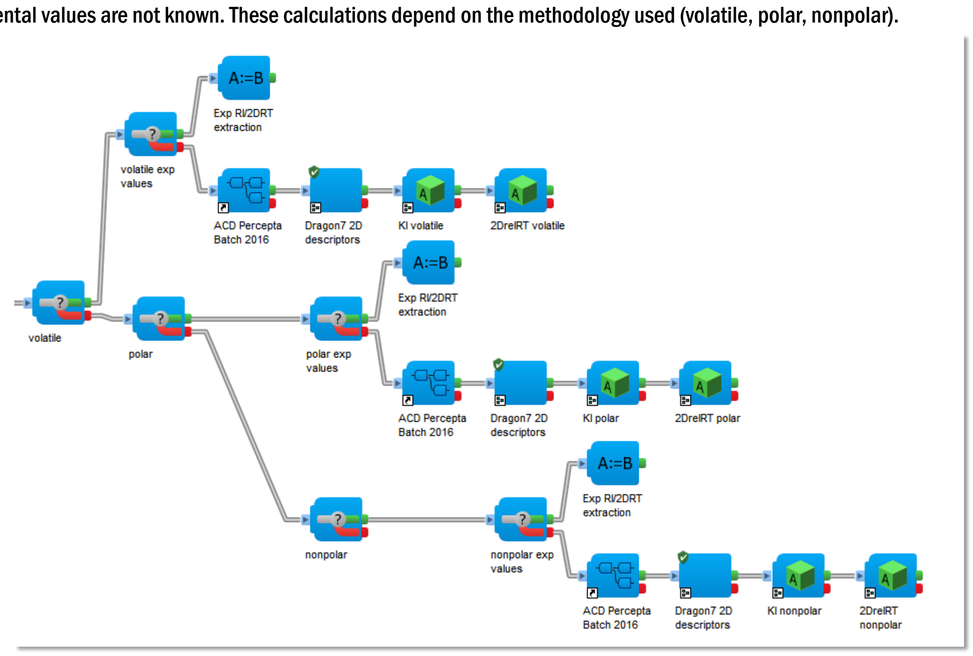
PP (1): Compound structure standardization

This protocol allows to calculate molecule identifier InChi, SMILES, or IUPAC name and to standardize the structures by removing salt, keeping the largest fragment, neutralizing charges, adding hydrogens, and cleaning 2D coordinates.



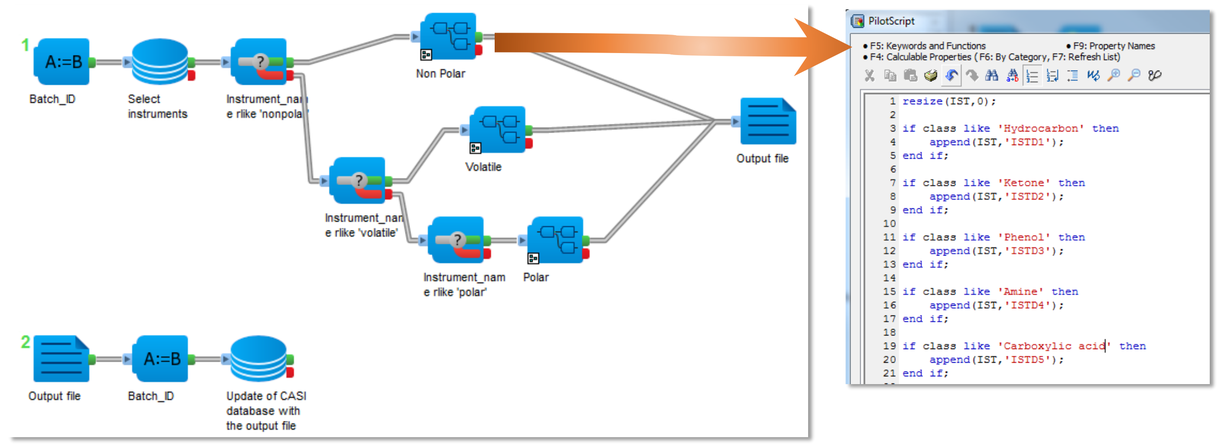
PP (2): Descriptor calculation and retention time prediction

Quantitative Structure-Property Relationship models are used to predict retention times/retention indices (RT/RI) for all constituents potentially present in the complex matrix (as is the aerosol). These predicted RT/RI may then be used to enhance the level of confidence in the correct assignment of compounds to determined mass spectra. The prediction models were done using QSAR Workbench[3] from BIOVIA and integrated to PP protocols to predict RT/RI for each potential hit. Models were created using descriptors calculated by Dragon (Kode srl, Dragon (software for molecular descriptor calculation) version 7.0.8, 2017, https://chm.kode-solutions.net) and ACD/Percepta batch (ACD/Percepta batch, version 2016.1.1, Advanced Chemistry Development, Inc., Toronto, ON, Canada, www.acdlabs.com, 2016). This protocol allows extraction of experimental retention indices (KI) and second-dimension (relative) retention times (2DrelRT) if they are present in a database or to predict them in case experimental values are not known. These calculations depend on the methodology used (volatile, polar, nonpolar).



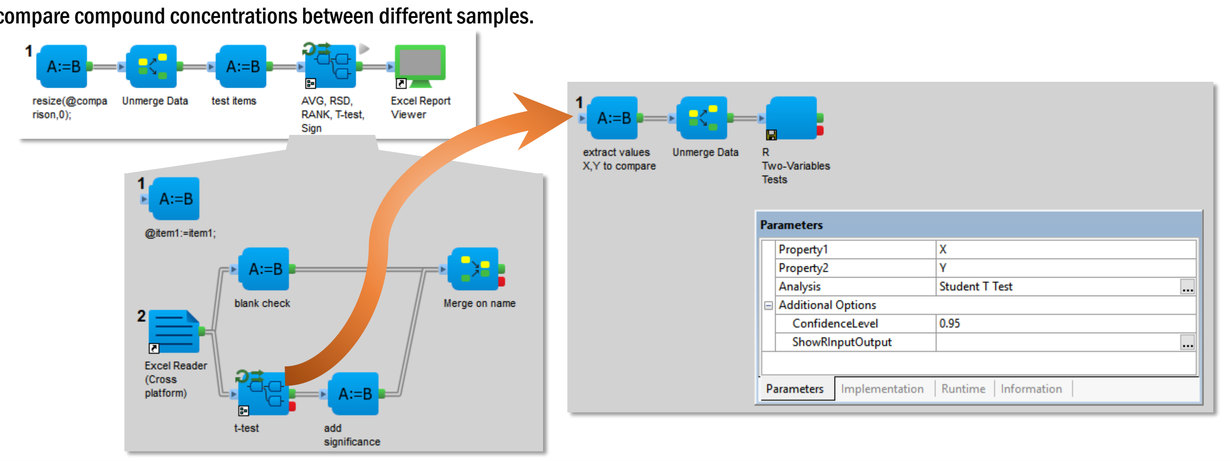
PP (3): Internal standard determination for concentration calculation

Once a structure is determined, it must be associated to an internal standard to calculate its concentration in the complex matrix. SMART rules were coded within PP to determine which internal standard to use for each structure. These rules depend on the method used for the analysis.



PP (4): Report Generation

Last step consists in generating an excel report in which samples are compared. Student's *t*-tests are computed using the R component of PP to compare compound concentrations between different samples.



Conclusions

- Use of Agile Scrum project management methodology allows rapid development of computational projects and ensures that the product is what users need.
- PP is a great tool to develop scripts to automate manual curation and calculate many chemical compound properties.
- Using this Agile Scrum approach and PP, we also developed many other in-house processes, and we recommend this approach to other industries.

References

1. Knorr A., Monge A., Stueber M., Stratmann A., Arndt D., Martin E., Pospisil P., "Computer-Assisted Structure Identification (CASI) - An Automated Platform for High-Throughput Identification of Small Molecules by Two-Dimensional Gas Chromatography Coupled to Mass Spectrometry", Anal Chem. 2013, 85(23):11216-24.
2. Martin E., Monge A., Duret J.A., Gualandi F., Peitsch M.C., Pospisil P., "Building an R&D chemical registration system", Journal of Cheminformatics 2012 4:11
3. Cox R., Green D., Luscombe C., Malcolm N. , Pickett S., "QSAR Workbench: automating QSAR modeling to drive compound design", J. Comput. Aided Mol. Des., 27, 321 (2013)

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL
www.pmiscience.com

BIT's 16th Annual Congress of INTERNATIONAL DRUG DISOVERY SCIENCE AND TECHNOLOGY

Boston, USA

August 16-18, 2018

Competing Financial Interest

The research described in this poster was funded by Philip Morris International