

Automation in Analytical Biopharmaceutical Drug Development with Open-Source Software

Vahid Golghalyani
Lonza, Analytical Development IBEX

CASSS, AT Europe 2020, Lisboa, Portugal

Acknowledgements



Lonza

Viv Lindo
Nick Bond
Kasia Kozakowska
Davide Di Girolamo
Sophie Inman
Paul Taylor
Maurizio Muroi
Praveen Kallamvalli Illam Sankaran
Lisa Wong

Fabio La Neve
Matthieu Stettler
Torsten Schmidt
Valentin Sottas
Alessandro Sinopoli
Joanna Olivi
Valeria Rebecchi
Hui Wang
Eric Navarra

Overview

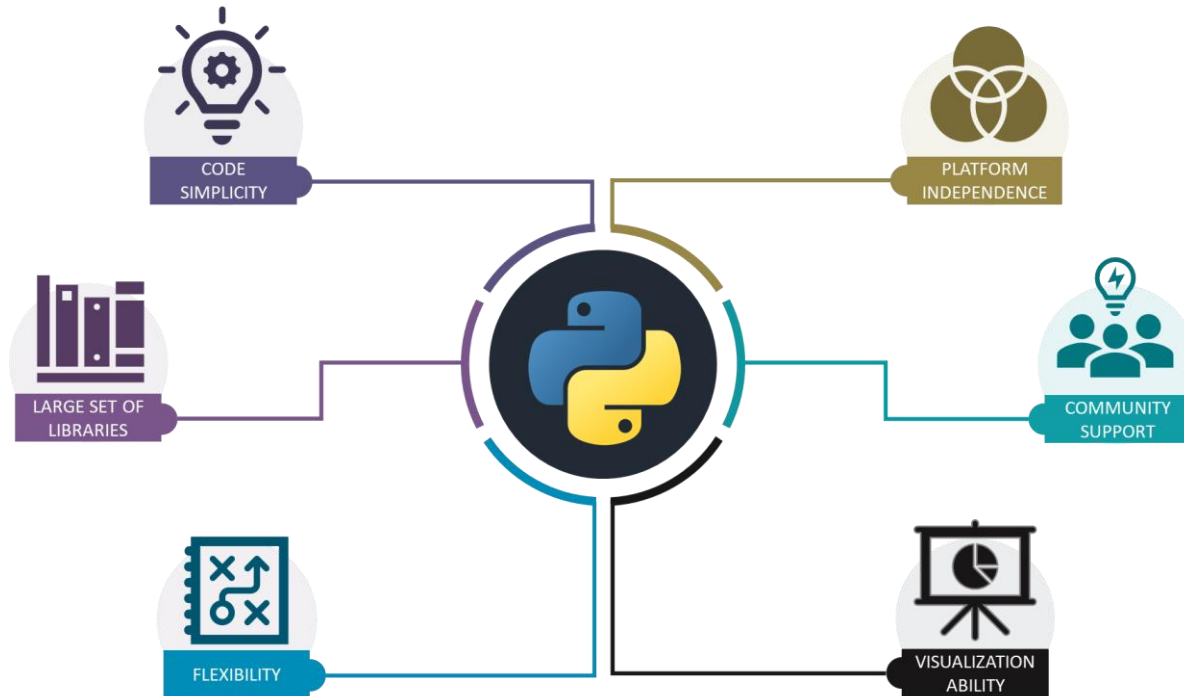


Worklist Generator

PDF data handling

MS Data Analysis

Python – Programming Language



- One of the most popular programming languages
- High Level – Human readable code
- Cross Platform – Runs on Windows, Mac, Linux
- Huge Community – Support
- Large Ecosystem – Multi Purpose, Libraries

	Redmonk (1/2019)	Tiobe (3/2019)	PYPL (2/2019)
1	JavaScript	Java	Python
2	Java	C	Java
3	Python	Python	JavaScript
4	PHP	C++	C#
5	C#	Visual Basic .NET	PHP
6	C++	C#	C/C++
7	CSS	JavaScript	R
8	Ruby	PHP	Objective-C
9	C	SQL	Swift
10	Objective-C	Objective-C	Matlab

Python – Deployment

- Most open-source projects deployed as command-line tools
- Less code and easier to maintain



- Proprietary software normally deployed with a user interface
- Tremendous increase of code

```
Command Prompt - pip install Pillow-3.1.0-cp34-none-win32.whl
14 Dir(s) 276,976,087,040 bytes free

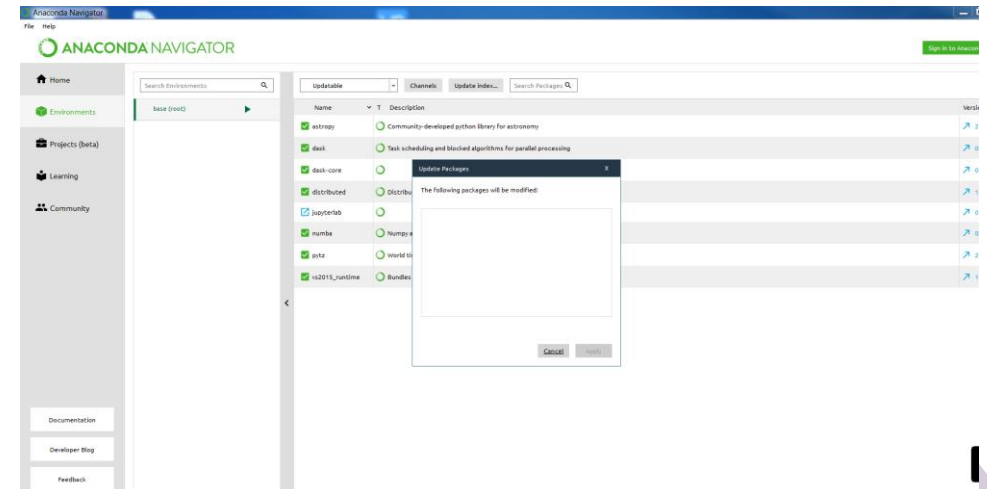
C:\>cd Users
C:\Users>dir
Volume in drive C has no label.
Volume Serial Number is 282D-00F7

Directory of C:\Users

18-04-2015  04:29 PM  <DIR>          .
18-04-2015  04:29 PM  <DIR>          ..
26-03-2015  10:47 PM  <DIR>          iet
16-01-2016  03:14 PM  <DIR>          Kavan
02-04-2015  06:23 PM  <DIR>          Public
26-03-2015  10:44 PM  <DIR>          UpdatusUser
               0 File(s)                0 bytes
               6 Dir(s) 276,976,087,040 bytes free

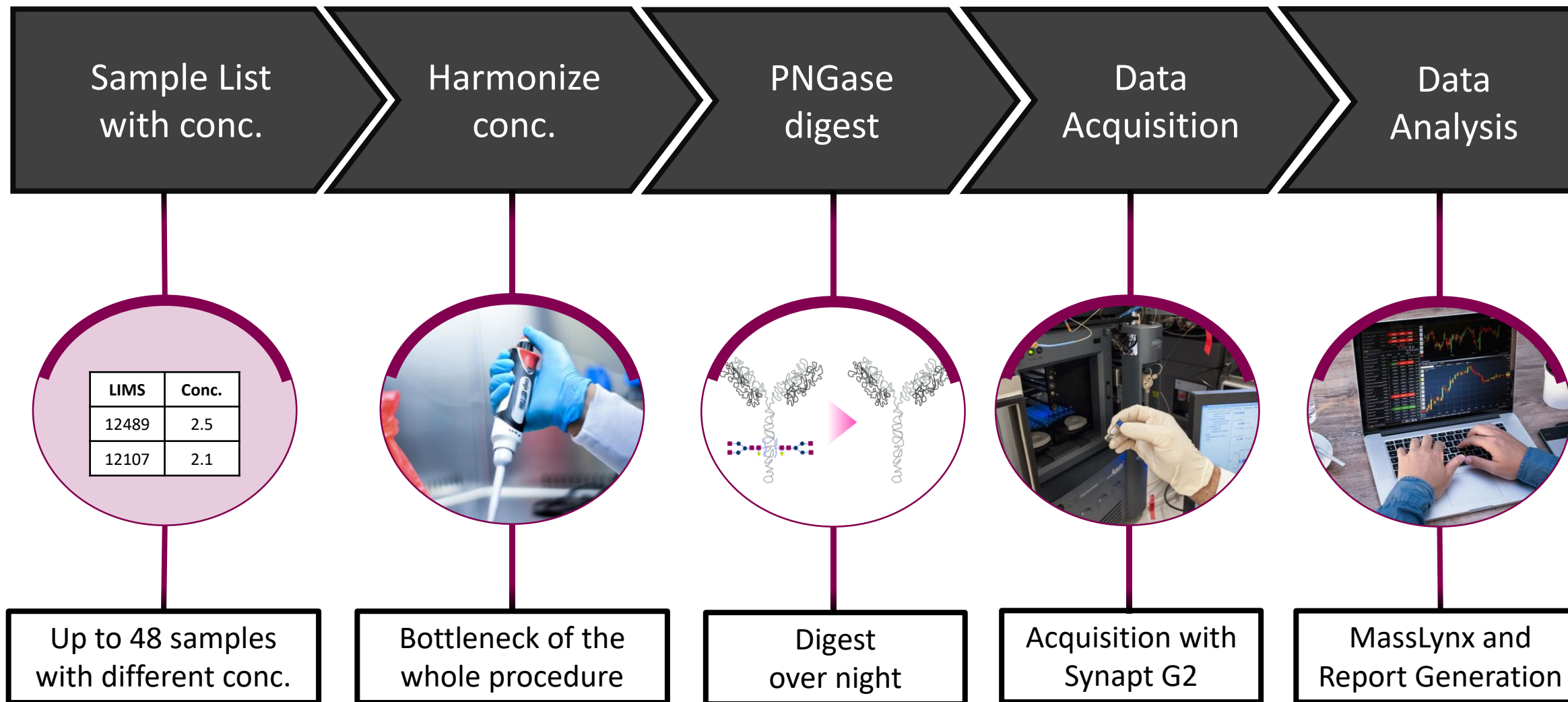
C:\Users>cd Kavan
C:\Users\Kavan>cd Downloads
C:\Users\Kavan\Downloads>pip install Pillow-3.1.0-cp34-none-win32.whl
```

COMMAND LINE



USER INTERFACE

Workflow Intact Mass Analysis



Automated Intact Mass Analysis



- User interface to input sample, labware and liquid class information
- Samples can be transferred to multiple destination plates
- Execution creates a worklist

Worklist Generator

```
worklist.gwl - Notepad
File Edit Format View Help
B;
S;4;
A;Buffer Trough;;;1;;39;water wet contact intact mass
D;Digest Plate;;;1;;39;water wet contact intact mass
W;
A;Buffer Trough;;;3;;27;water wet contact intact mass
D;Digest Plate;;;3;;27;water wet contact intact mass
W;
A;Buffer Trough;;;4;;36;water wet contact intact mass
D;Digest Plate;;;4;;36;water wet contact intact mass
W;
A;Buffer Trough;;;5;;35;water wet contact intact mass
D;Digest Plate;;;5;;35;water wet contact intact mass
W;
A;Buffer Trough;;;6;;37;water wet contact intact mass
D;Digest Plate;;;6;;37;water wet contact intact mass
W;
A;Buffer Trough;;;8;;39;water wet contact intact mass
D;Digest Plate;;;8;;39;water wet contact intact mass
W;
A;Buffer Trough;;;1;;28;water wet contact intact mass
D;Digest Plate;;;9;;28;water wet contact intact mass
W;
A;Buffer Trough;;;2;;26;water wet contact intact mass
D;Digest Plate;;;10;;26;water wet contact intact mass
```

- External tools can be triggered within a Evoware Tecan script
- Generated Worklist can be loaded and executed



Worklist Generator

Labware Names →

Final concentration and volume →

Liquid Classes →

Upload xls →

Higher volume first →

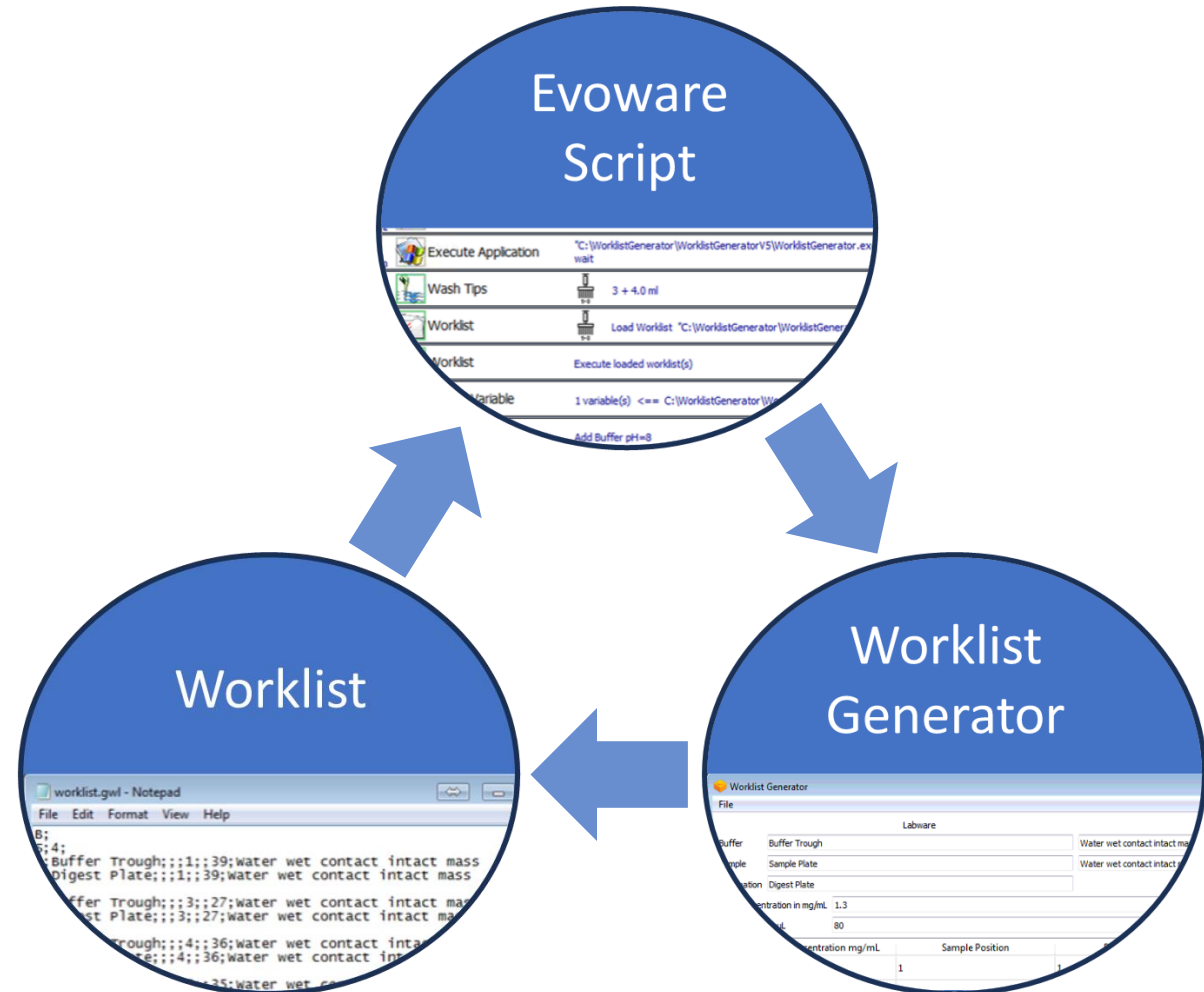
Load and save config files →

The screenshot shows the 'PlatePage 1' window of the 'Wizard' application. It contains several input fields for 'Labware' and 'Liquid Class', a 'Destination' dropdown, and numerical inputs for 'Final concentration in mg/mL' and 'Final volume in uL'. A 'Calculate' button is highlighted with a red box. Below these fields is a table with 4 columns: 'Sample concentration mg/mL', 'Sample Position', 'Buffer Position', and 'V(Sample):V(Buffer)'. The table contains 10 rows of data. At the bottom, there are buttons for 'Load Config File', 'Save Config File', 'Next', and 'Cancel'.

	Sample concentration mg/mL	Sample Position	Buffer Position	V(Sample):V(Buffer)
1	10.0	2	1	10:70
2	9.6	3	2	11:69
3	2.52	4	3	41:39
4	3.05	5	4	34:46
5	2.09	6	5	50:30
6	2.26	7	6	46:34
7	2.29	8	7	45:35
8	2.19	9	8	47:33
9	2.83	10	1	37:43
10	2.5	11	2	42:38

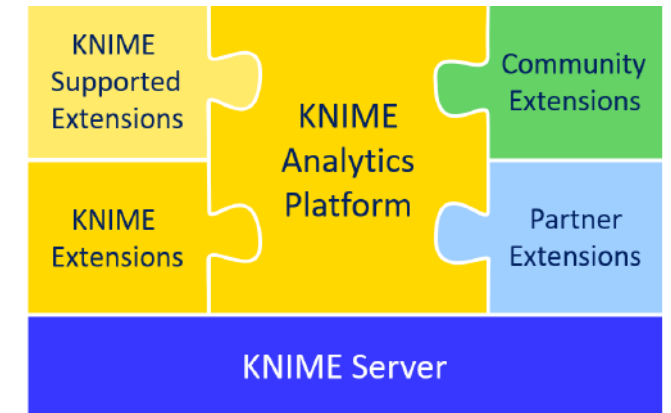
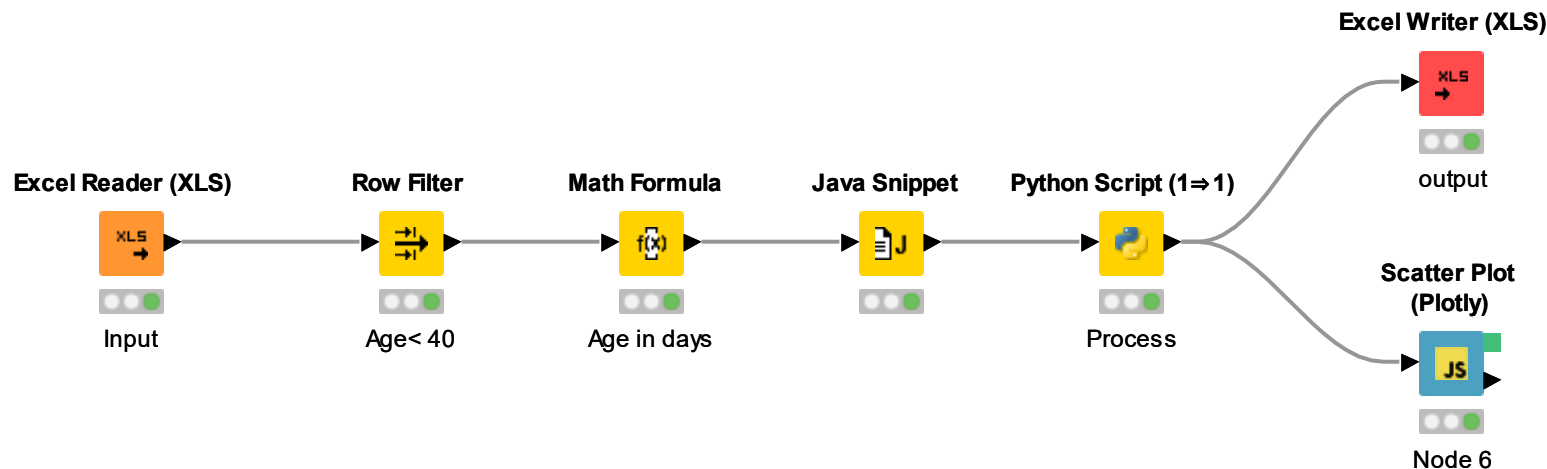
Worklist Generator - Summary

- Universal dilution method is created for Tecan liquid handling systems
- Samples can be transferred to multiple destination plates and diluted with the assay appropriate buffer
- User interface allows to define all relevant information and configuration can be saved and loaded
- Worklist has the potential to be used for other applications than dilutions, such as merging of samples from different plates to one destination plate



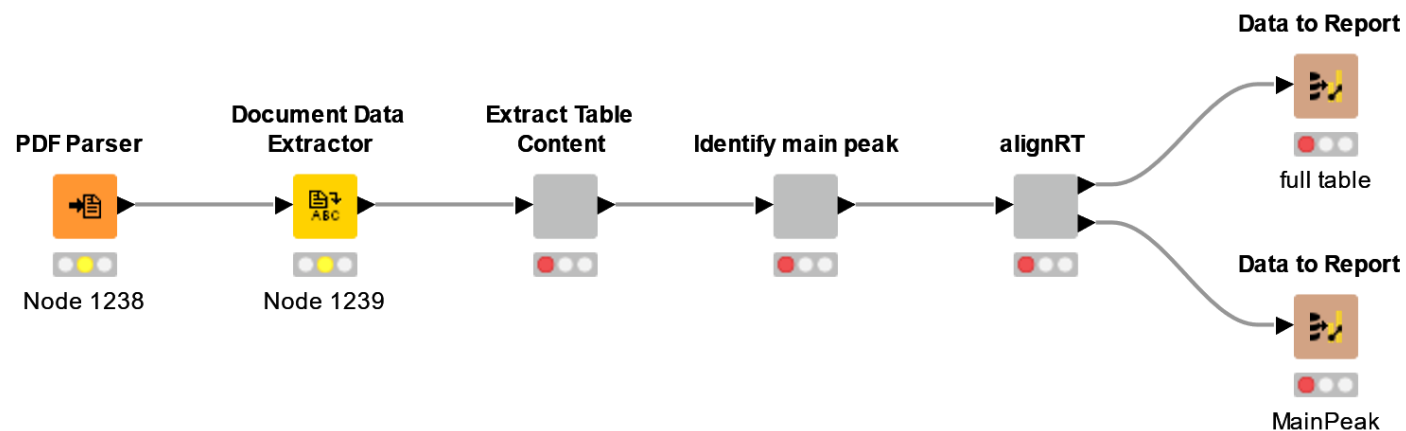
KNIME – Analytics Platform

- KNIME.com GmbH has multiple sites in Europe and one in the US
- Main product is KNIME analytics platform (open-source)
- Based on the graphical programming paradigm – workflows created of nodes
- Approximately 1500 nodes exist, huge fraction supplied by the global community



Automated PDF Data Handling

- Many assays result in a pdf report and analysts have to copy this data to process it further
- KNIME provides text processing nodes, that can extract text information out of pdf files
- Integrated Python script identifies lines of importance and creates a table
- Assignment of single rows can be done fully automatically or require the review of the analyst
- Results can be released into a report



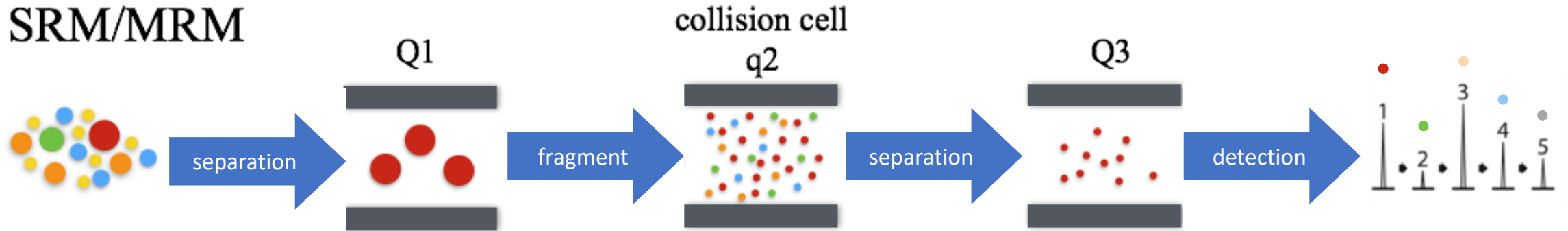
Automated PDF Data Handling



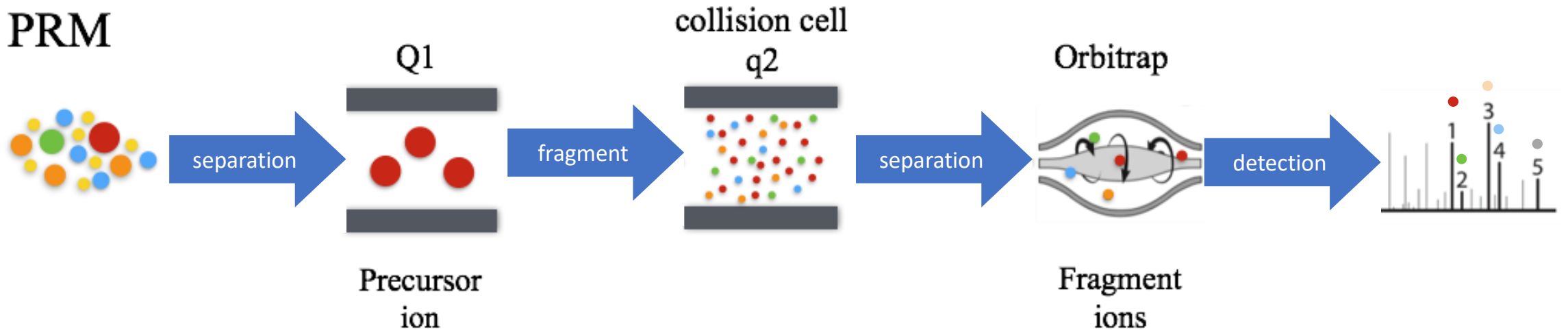
Video will be shown at the presentation.

Parallel Reaction Monitoring

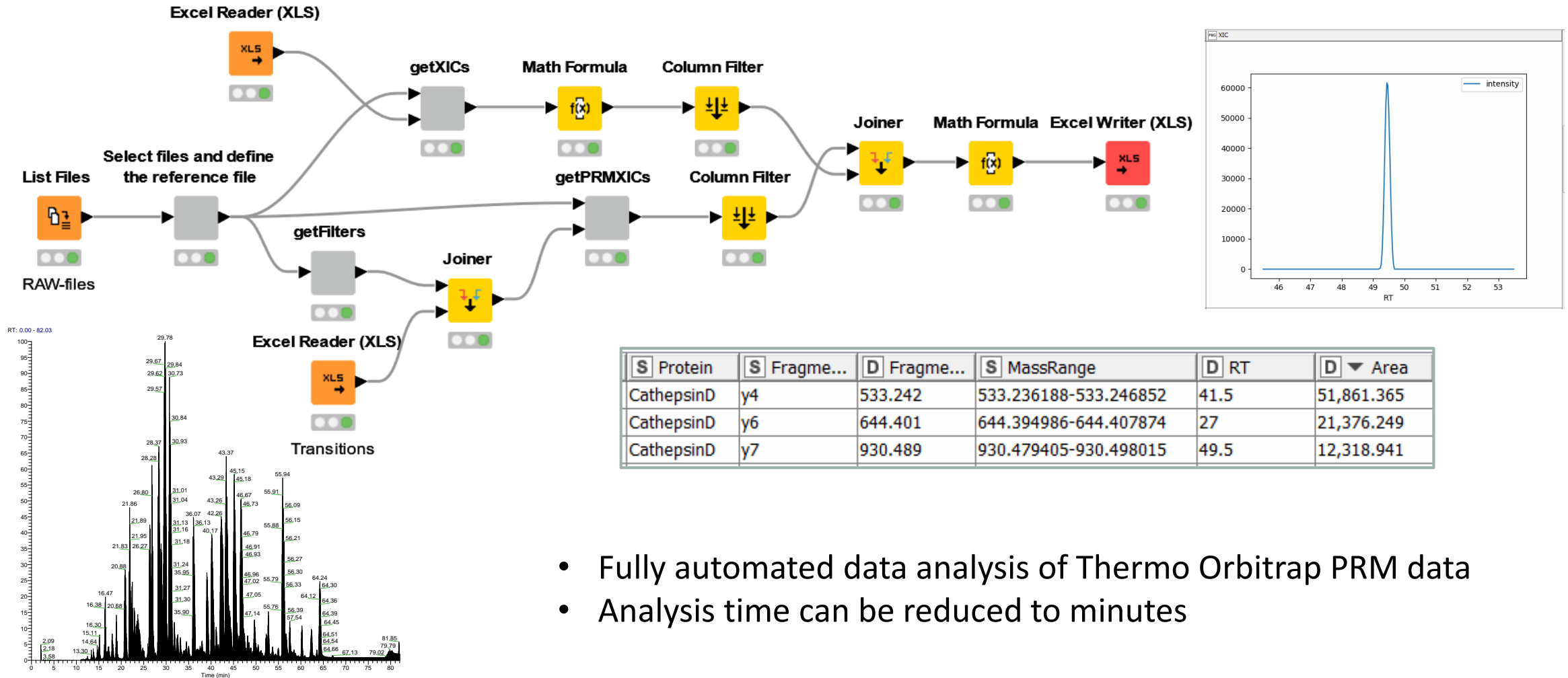
SRM/MRM



PRM



Automated PRM Data Analysis

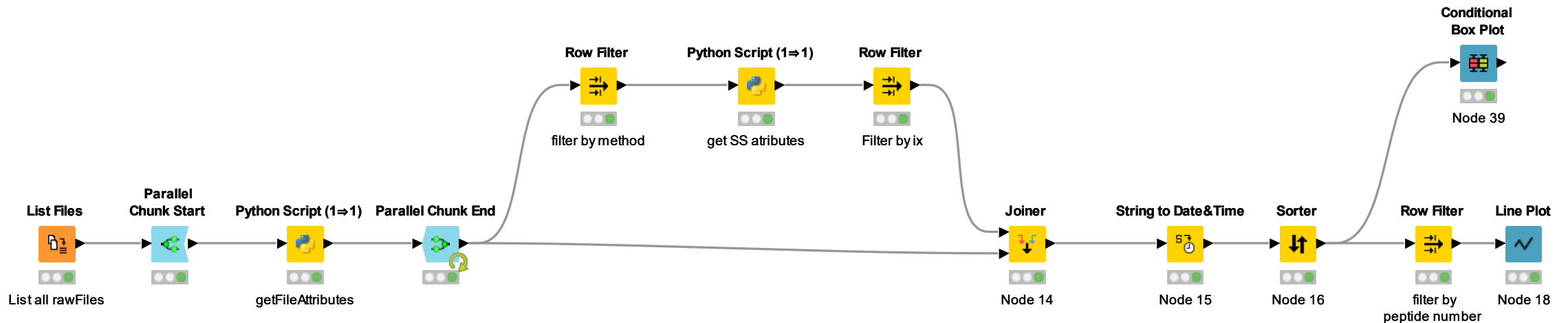


S Protein	S Fragme...	D Fragme...	S MassRange	D RT	D ▼ Area
CathepsinD	y4	533.242	533.236188-533.246852	41.5	51,861.365
CathepsinD	y6	644.401	644.394986-644.407874	27	21,376.249
CathepsinD	y7	930.489	930.479405-930.498015	49.5	12,318.941

- Fully automated data analysis of Thermo Orbitrap PRM data
- Analysis time can be reduced to minutes

Automated MS Data Analysis

- MSFileReader to extract data from Thermo RAW files and enables the creation of XICs and TICs
- Many Python libraries available to analyse mass spectrometry data like pyOpenMS, pyProteome, PYQMS, pyMzML, Multiplierz, mMass and Unidec
- KNIME workflows exist for system suitability trending, UV-overlay creation and report creation for BioPharmaFinder peptide mapping results



<https://github.com/Roestlab/PythonProteomics>

Summary

- Introduction into KNIME and Python, case examples and how these tools can be used to create customized software solutions
- Worklist Generator, UI-base software written in Python and can be utilized for a universal dilution method and/or to run multiple assays on the same set of samples
- With KNIME workflows data can be extracted out of PDF reports, analyzed and results can be released in a customized report
- KNIME workflows have been created to analyze MS data such as PRM and peptide mapping data