jqcML: An open-source Java API for mass spectrometry quality control data in the qcML format

Wout Bittremieux, Pieter Kelchtermans, Dirk Valkenborg, Lennart Martens and Kris Laukens

Introduction

Mass spectrometry is widely used to identify proteins based on the mass distribution of their peptides/fragments. Unfortunately, because of its inherent complexity, the results can be subject to a large **variability** (Figure 1). As a means of **quality control**, several qualitative metrics have been defined. However, these still suffer some **limiting factors**:

- **Compatibility**: Storing and communicating of quality control data is not standardized, limiting the dissemination along with experimental data;
- **Variability**: The data can be generated by software tools of different origins, with content and definitions varying for each tool.

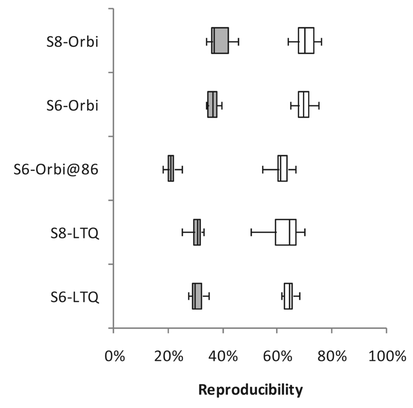

Figure 1: Reproducibility of identifications between different experiments on different instruments. Shaded boxes represent peptides, while white boxes represent proteins.[1]

qcML

The **qcML standard**[2] addresses these issues:

- **Compatibility**: XML-based file format (Figure 2; interchange format), and relational database (archival);
- **Variability**: Controlled vocabularies to unambiguously define terms.

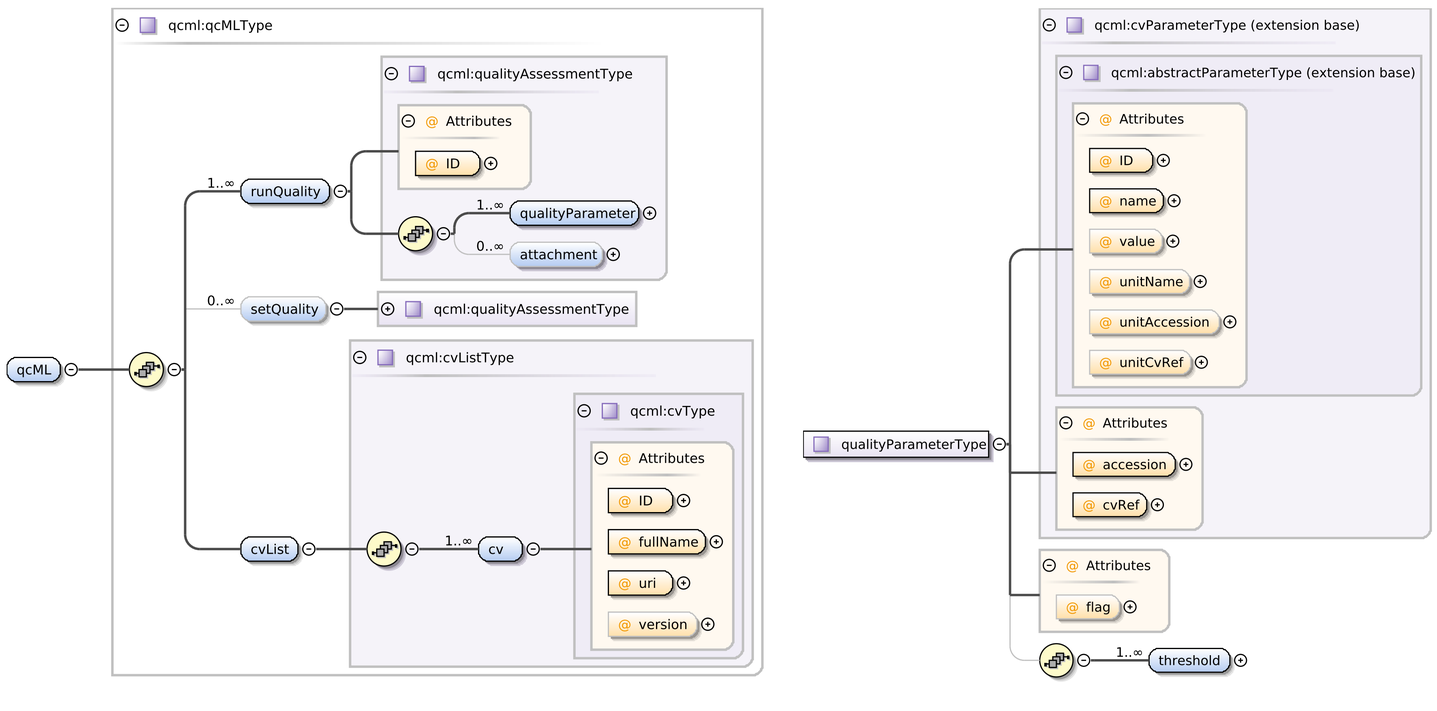

Figure 2: Core elements of the XML schema used to represent the information stored in qcML.

jqcML

jqcML is an open-source **Java API** for working with qcML data:

- Complete **object model** to represent qcML data;
- The ability to work with data from **several sources** in a **uniform manner** (Figure 3).

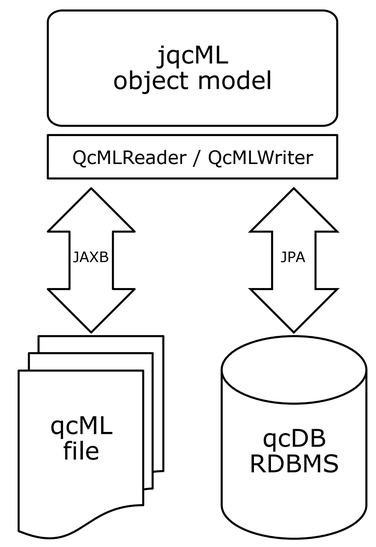

Figure 3: Simplified representation of the jqcML workflow. Through the use of common interfaces, jqcML is able to work with qcML data from several sources in a uniform way.

Example

Using OpenMS[3], qcML data was calculated for **several thousand of mass spectrometry experiments**. With the use of jqcML, specific parameters can easily be extracted in order to perform advanced **analyses between different experiments** (Figure 4).

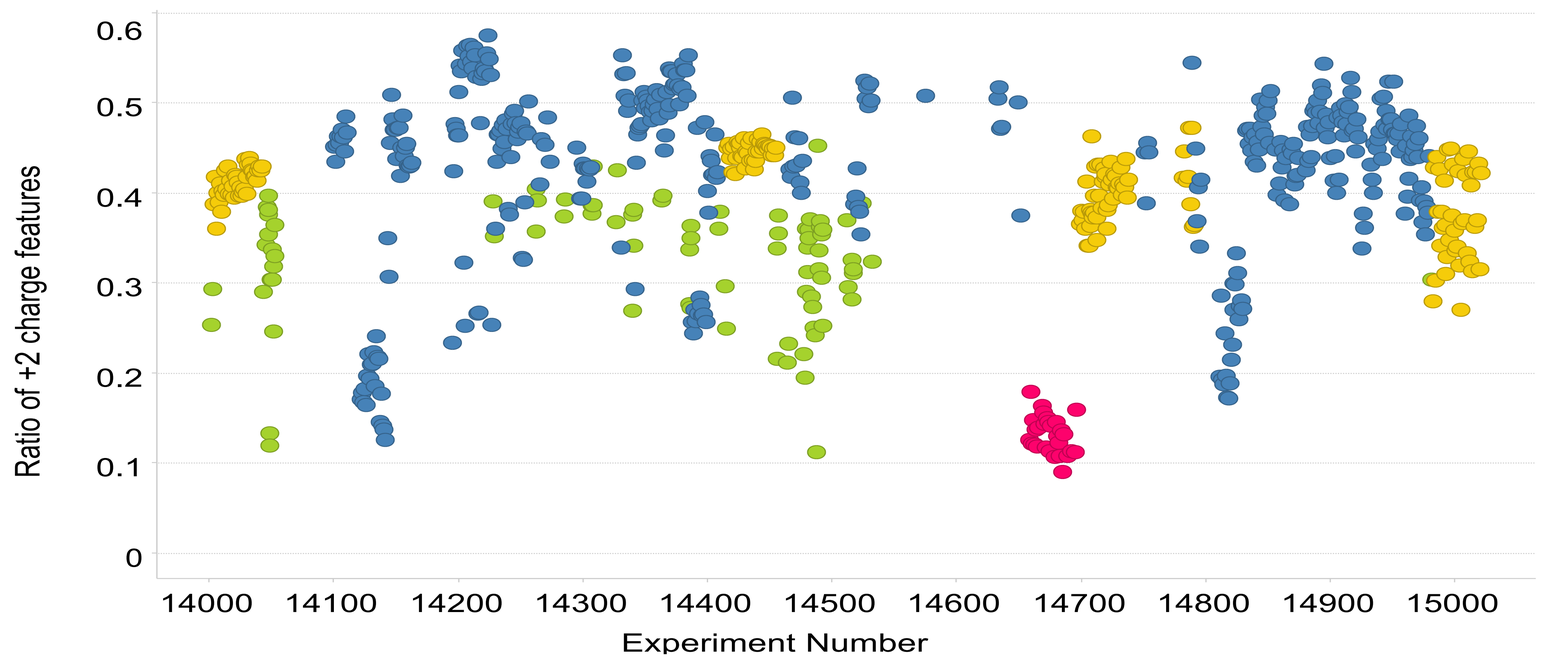

Figure 4: The ratio of +2 charged features on MS1 level for a set of experiments on a Thermo Scientific LTQ Orbitrap Velos. The data points are colored by type of experimental protocol.

Conclusion

The expressive file format and database structure defined by the qcML specification allows a wide range of possibilities in dealing with quality control data in a standardized way. The jqcML library contains all the required functionality in order to work with qcML data. As such, it will provide an invaluable tool for the adoption of the qcML standard.
Our future work will deal with performing advanced analyses and data mining tasks on qcML data.

1. Tabb, D. L. Quality assessment for clinical proteomics. *Clinical Biochemistry* **46**, 411–420 (2013).
2. Walzer, M. *et al.* qcML: an exchange format for quality control metrics from mass spectrometry instruments. [Manuscript submitted].
3. Kohlbacher, O. *et al.* TOPP--The OpenMS Proteomics Pipeline. *Bioinformatics* **23,** e191–e197 (2007).

wout.bittremieux@uantwerpen.be

Universiteit Antwerpen