



MSDATASTREAM

CONNECTING A BRUKER MASS SPECTROMETER TO THE INTERNET

Roman Zoun¹, Kay Schallert¹, David Broneske¹, Wolfram Fenske¹, Marcus Pinnecke¹, Robert Heyer¹, Sven Brehmer², Dirk Benndorf¹, Gunter Saake¹.

¹ Otto-von-Guericke University Magdeburg
² Bruker Daltonik GmbH

18. BTW 2019

Introduction

- Mass spectrometry workflow is sequential (Figure 1)

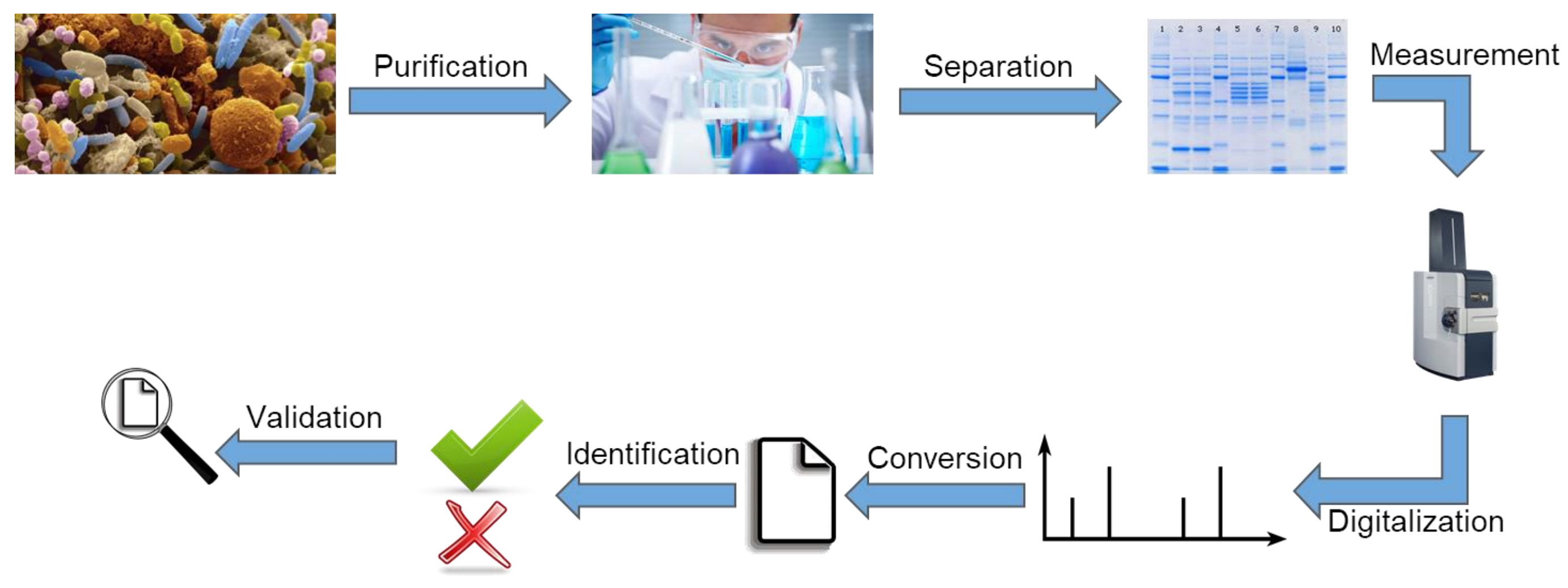


Figure 1: Sequential mass spectrometry pipeline.

- Protein identification [2] in real time, see Figure 2

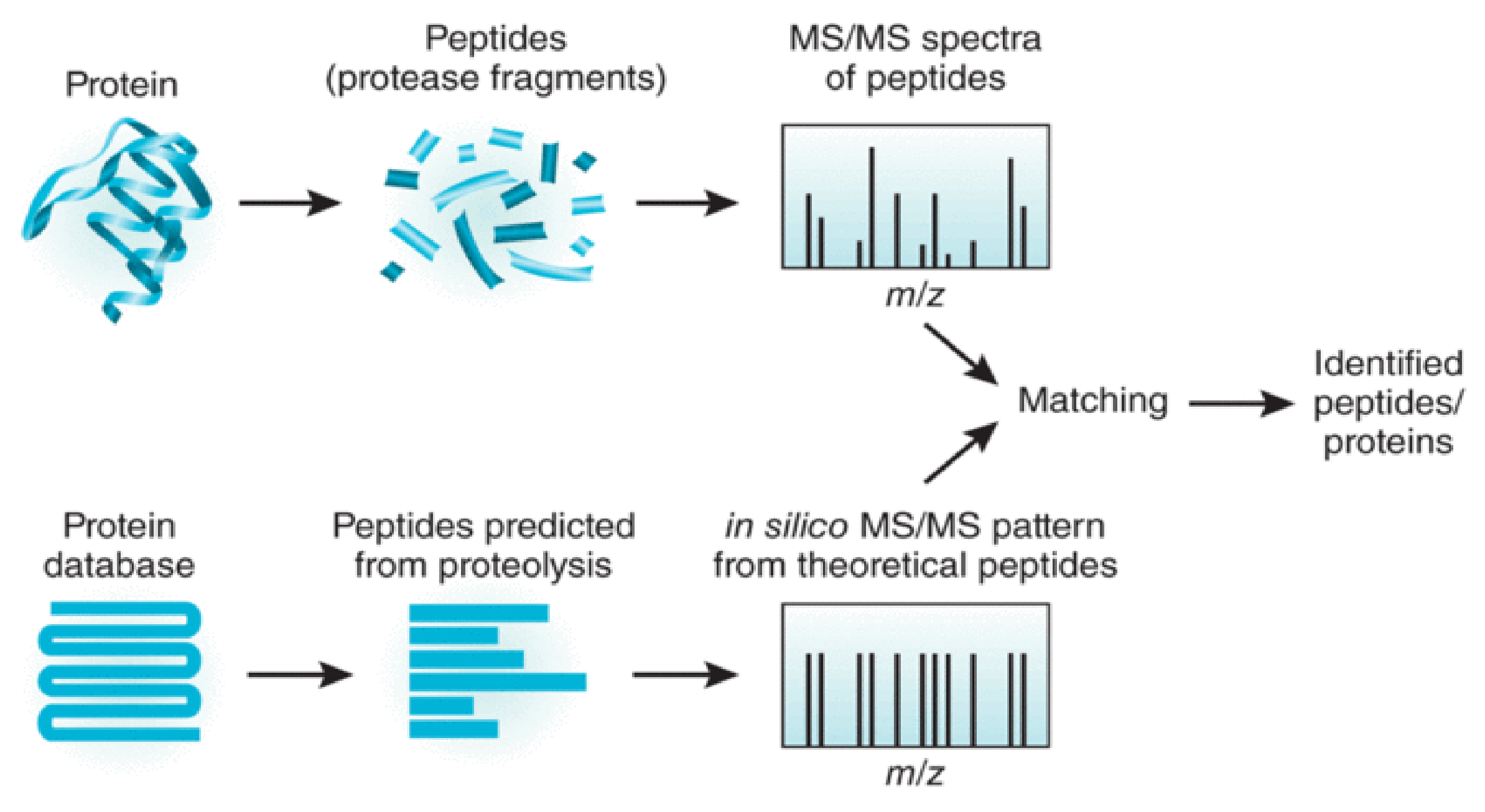


Figure 2: Principle of protein database search.
(source: <https://www.nature.com/articles/nbt0710-659>)

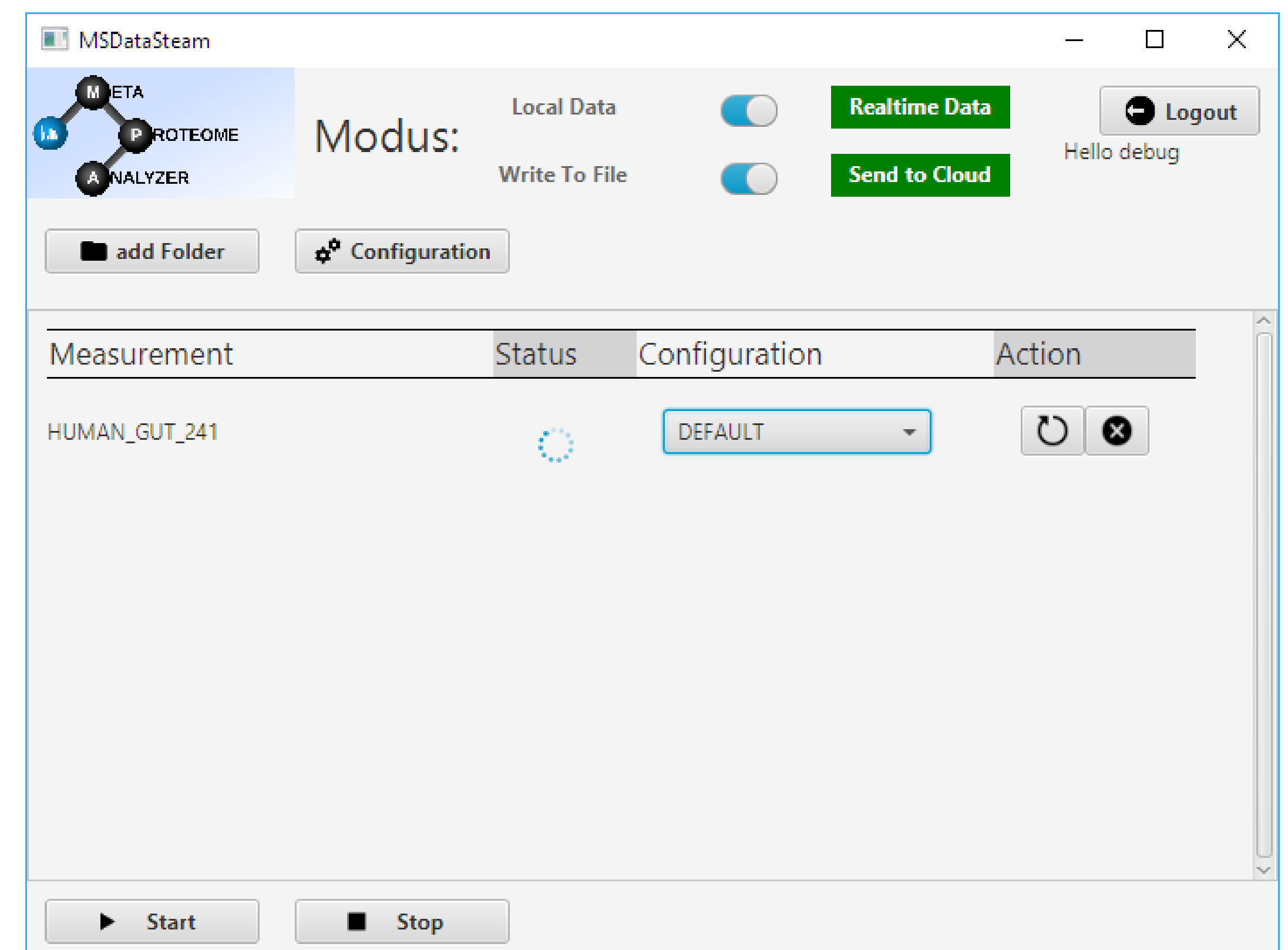


Figure 4: MSDataStream user interface

Cloud Backend Fast Data Architecture

- Fast Data architecture [3]
- Stream producer (Figure 5-1)
- Horizontally scalable scorer (Figure 5-2)
- Distributed structured protein knowledge base (Figure 5-3)
- Smart validator of peptide-spectrum-matches (Figure 5-4).

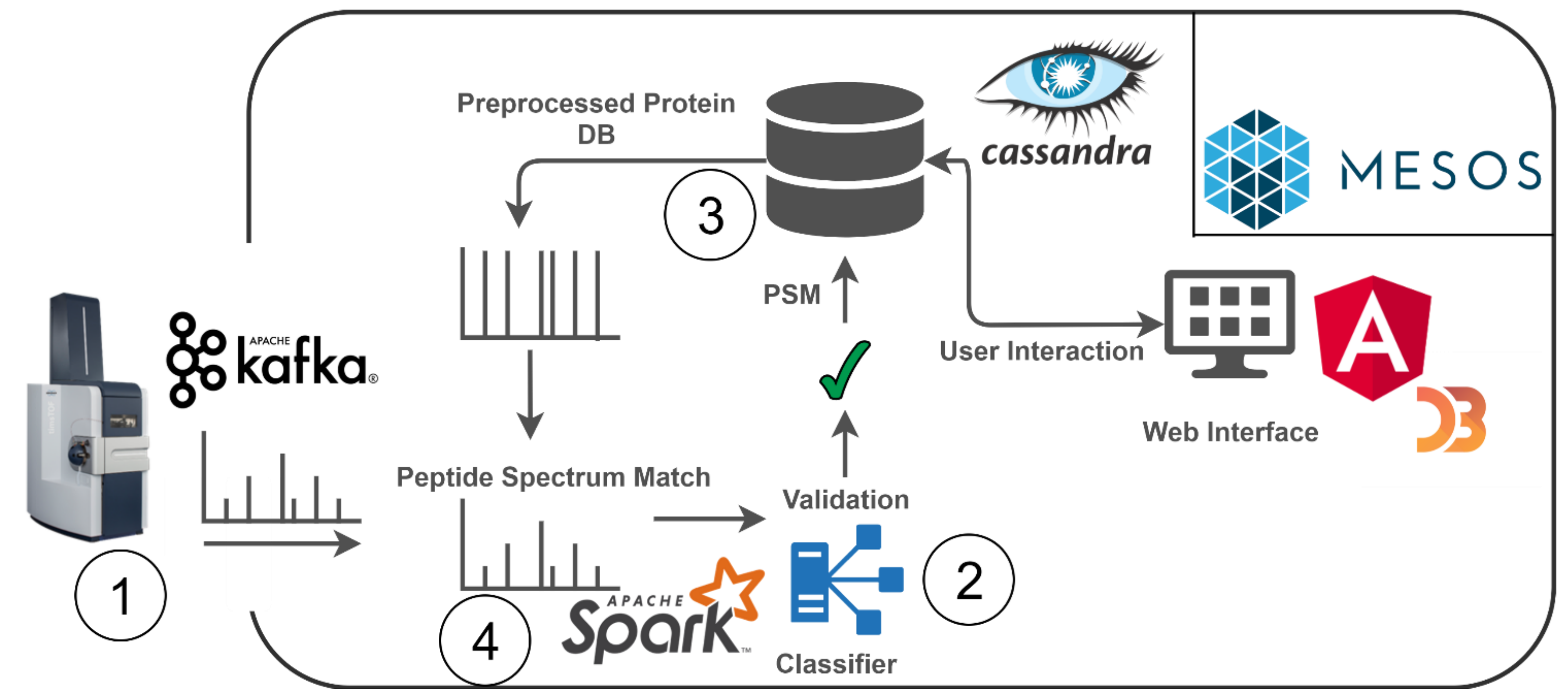


Figure 5: Fast Data architecture

MSDataStream

- Collaboration with Bruker Daltonik GmbH.
- MSDataStream reads single spectrum data during measurement
- Sending the data to the cloud
- Enables real time diagnostics on the cloud

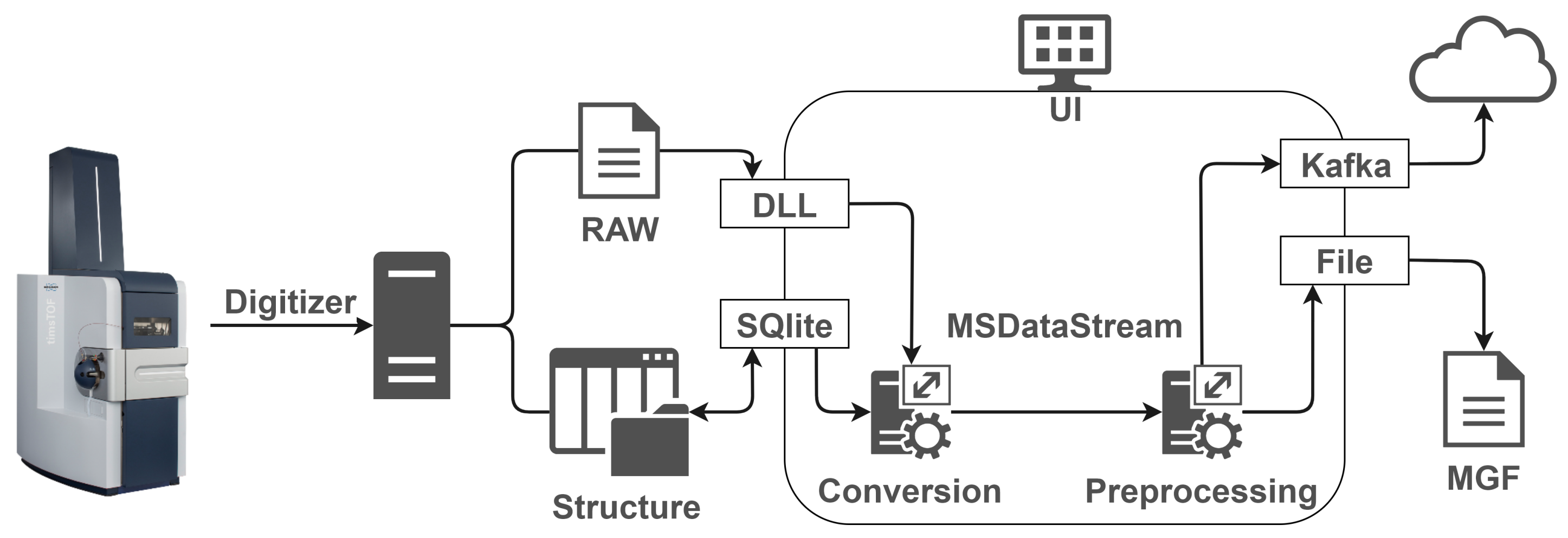


Figure 2: Architecture of MSDataStream

References

[1] "TANDEM: matching proteins with tandem mass spectra."; R. Craig, R. C. Beavis (2004); Bioinformatics, 20, pp 1466-1467;
[2] "Fast Data Architectures for Streaming Applications"; D. Wampler (2016); 1st ed. 1005 Gravenstein Highway North, Sebastopol, CA 95472.: Oreilly Media, sep.