
CoPEXplorer: A Novel Collaboration Platform for Quantitative Proteomics Research in China

Tun Lu

School of Computer Science
Fudan University
825 Zhangheng Road
Shanghai 201203, China
lutun@fudan.edu.cn

Xianghua Ding

School of Computer Science
Fudan University
825 Zhangheng Road
Shanghai 201203, China
dingx@fudan.edu.cn

Xing Huang

School of Computer Science
Fudan University
825 Zhangheng Road
Shanghai 201203, China
xinghuang@fudan.edu.cn

Guangmeng Zhai

School of Computer Science
Fudan University
825 Zhangheng Road
Shanghai 201203, China
guangmengzhai@gmail.com

Zhaocan Chen

School of Computer Science
Fudan University
825 Zhangheng Road
Shanghai 201203, China
chenzhaocan@gmail.com

Ning Gu

School of Computer Science
Fudan University
825 Zhangheng Road
Shanghai 201203, China
ninggu@fudan.edu.cn

Abstract

Quantitative proteomics, with its aim to comprehensively identify and quantify proteins in complex samples, is a relatively new research field in bioinformatics. It usually involves scientists from different domains to collaboratively deal with various artifacts such as massive data, software tools and analysis processes to achieve their research goals. However, issues such as non-standard data formats, scattered software tools, and poor interaction methods, have posed great challenges to collaborative research in proteomics. To address these issues, we developed a novel collaboration platform called *CoPEXplorer* to facilitate cooperative research in proteomics. In this position paper, we briefly introduce *CoPEXplorer* from three collaboration layers (i.e. group, software, and data) and share our initial practical experiences in this data-intensive collaboration field.

Author Keywords

Collaboration platform; Quantitative proteomics; Group collaboration; Software collaboration; Data collaboration

ACM Classification Keywords

C.2.4 Distributed Systems: Distributed Applications;
D.2.13 Reusable Software: Reuse models

General Terms

Design, Experimentation, Management

Introduction

As one of the major emerging areas of post-genome era, quantitative proteomics is of key importance to life sciences, medicine and biotechnology development. It brings hopes to explore the various phenomena of life and the nature of occurrence and development of diseases, one of the hottest research fields in the 21st century [1, 2]. Massive data analysis lies at the heart of quantitative proteomics to comprehensively identify and quantify proteins in different complex samples [3], and it often involves a group of researchers from different domains to communicate and collaborate to complete the analysis task.

However, there are many obstacles to research collaboration in quantitative proteomics [4, 5]. Firstly, it is difficult for the massive proteomics data with different structures and formats to be exchanged and shared. Secondly, various software tools for different analysis steps are scattered in geographically dispersed labs, making it inconvenient for scientists to share and reuse them, especially in dynamic analysis processes. Thirdly, for the scientists from different research domains, individual and separate communication or coordination tools are far from being adequate to meet the collaboration requirements such as collaborative awareness, collaborative annotation, etc.

To deal with these challenges, we developed a novel collaboration platform called *CoPEXplorer* to facilitate cooperative research in quantitative proteomics, under the support of the *National High-Tech Research & Development Program (863 Program)* of Ministry of Science and Technology of China. In this position paper, we will briefly introduce *CoPEXplorer* from three collaboration layers, including group collaboration layer, software collaboration layer, and data collaboration

layer; and share our initial practical experiences in this data-intensive collaboration field.

Architecture of *CoPEXplorer*

CoPEXplorer is devised and implemented in a three-layer architecture showed in Figure 1.

The data collaboration layer manages and standardizes heterogeneous data involved in quantitative proteomics research process, and maintains the consistency of the analysis results. The software collaboration layer wraps the scattered data processing software tools and publishes them as Web services, and customizes the data analysis process according to different analysis goals by using service discovery and composition techniques. The group collaboration layer models the group roles and their collaborative behavior, and provides collaboration tools for cooperative work such as collaborative awareness, collaborative annotation and group notification.

Sub-systems of *CoPEXplorer*

CoPEXplorer consists of seven sub-systems distributed in these three layers: standardization & management for heterogeneous data, consistency maintenance for analysis result data (in data collaboration layer), web service wrapping for analysis software, web service composition for analysis software, customizable web service process management (in software collaboration layer); organization and management for data analysis group, collaborative browse of experiment & research data (in group collaboration layer).

Sub-systems in data collaboration layer

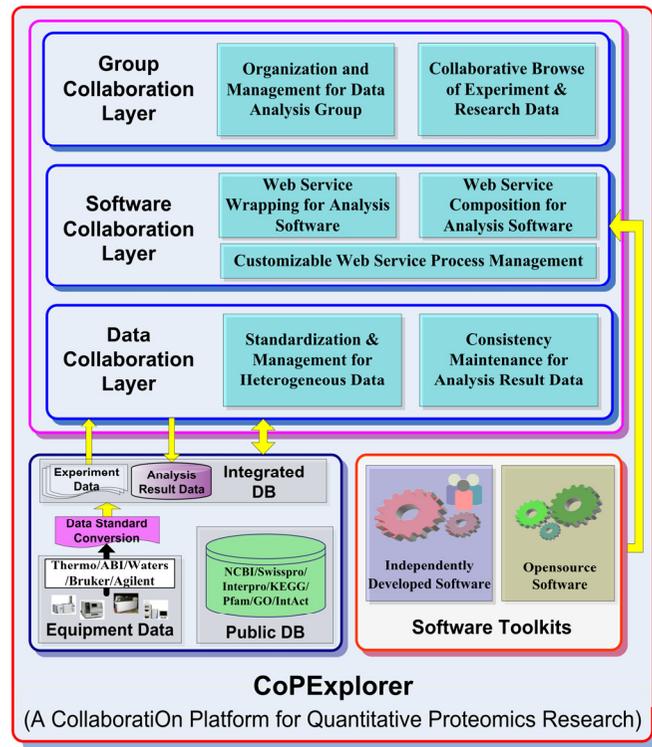


Figure 1. Architecture of CoPEXplorer

- Standardization & Management for Heterogeneous Data – To deal with heterogeneous data coming from different analysis devices & equipments, analysis process steps, and public database, this sub-system standardizes the data format and stores them in the integrated database, and provides uniform search & query interfaces for data exchange and sharing between different labs and research groups.
- Consistency Maintenance for Analysis Result Data – This sub-system maintains the consistency of

analysis results in each step to achieve the correctness of the whole analysis process in both on-line and off-line collaboration scenarios.

Sub-systems in software collaboration layer

- Web Service Wrapping for Analysis Software – This sub-system describes and wraps the analysis software by standard WSDL (Web Service Description Language), and builds UDDI (Universal Description Discovery and Integration) center to publish, discover and share these Web services.
- Web Service Composition for Analysis Software – This sub-system builds a semantic ontology repository of proteomics to model the Web services, and then supports automatic Web service composition.
- Customizable Web Service Process Management – This sub-system provides visual modeling tools, process execution engines and corresponding management tools for analysis process customization.

Sub-systems in group collaboration layer

- Organization and Management for Data Analysis Group – To address the distributed nature of analysis processes and the dynamic nature of analysis roles, this sub-system provides role-based access control and rule-based role management services.
- Collaborative Browsing of Experiment & Research Data – This sub-system integrates various collaborative tools such as collaborative awareness, collaborative annotation, group notification and instant messaging to support collaborative browsing.

A screenshot of the platform homepage of *CoPEXplorer* is shown in Figure 2.



Figure 2. Homepage of CoPEXplorer

Initial Feedback and Future Work

Quantitative proteomics is a relatively new research domain, and to our best knowledge, *CoPEXplorer* is the first platform to support collaboration research in this emerging field. As of this writing, *CoPEXplorer* has been deployed in Shanghai Center for Bioinformatics Technology (SCBIT) and Research Center of Protein Analysis (RCPA) of Shanghai Institute for Biological Sciences for a field trial. The initial feedback is

encouraging in its effectiveness in supporting collaborative research. With *CoPEXplorer* deployed in the real field, we'll explore how *CoPEXplorer* is used, what role *CoPEXplorer* plays in supporting collaborations in quantitative proteomics research, and what kinds of collaborative behavior may emerge.

References

- [1] Ong SE, Mann M. Mass spectrometry-based proteomics turns quantitative. *Nature Chemical Biology*, 1, 5 (2005): 252–262.
- [2] Bantscheff M, Schirle M, Sweetman G, Rick J, Kuster B. Quantitative mass spectrometry in proteomics: a critical review. *Anal Bioanal. Chem.*, 389, 4 (2007): 1017–31.
- [3] Quantitative Proteomics. http://www.systemsbio.org/technology/data_generation/Quantitative_Proteomics.
- [4] Sturm M, Bertsch A, Gropl C, Hildebrandt A, Hussong R, Lange E, Pfeifer N, Schulz-Trieglaff O, Zerck A, Reinert K, Kohlbacher O. OpenMS - an open-source software framework for mass spectrometry. *BMC Bioinformatics*, (2008), 9:163.
- [5] Keller A., Eng J., Zhang N., Li X. J., Aebersold R. A uniform proteomics MS/MS analysis platform utilizing open XML file formats. *Mol. Syst. Biol.* 1, (2005).0017.