

This item is the archived peer-reviewed author-version of:

The European Bioinformatics Community for Mass Spectrometry (EuBIC-MS) : an open community for bioinformatics training and research

Reference:

Bittremieux Wout, Bouyssié David, Dorfer Viktoria, Locard-Paulet Marie, Perez-Riverol Yasset, Schwämmle Veit, Uszkoreit Julian, Van Den Bossche Tim.- The European Bioinformatics Community for Mass Spectrometry (EuBIC-MS) : an open community for bioinformatics training and research
Rapid communications in mass spectrometry - ISSN 0951-4198 - (2021), e9087
Full text (Publisher's DOI): <https://doi.org/10.1002/RCM.9087>
To cite this reference: <https://hdl.handle.net/10067/1772740151162165141>

The European Bioinformatics Community for Mass Spectrometry (EuBIC-MS): an open community for bioinformatics training and research

*Wout Bittremieux^{1,3}, David Bouyssié^{1,4}, Viktoria Dorfer^{1,5}, Marie Locard-Paulet^{1,6}, Yasset Perez-Riverol^{1,7}, Veit Schwämmle^{1,8}, Julian Uszkoreit^{1,9,10}, and Tim Van Den Bossche^{*1,11,12}*

- (1) European Bioinformatics Community for Mass Spectrometry (info@eubic-ms.org)
- (2) University of California San Diego, La Jolla, CA, USA
- (3) University of Antwerp, Antwerpen, Belgium
- (4) IPBS, University of Toulouse, CNRS, UPS, Toulouse, France
- (5) University of Applied Sciences Upper Austria, Bioinformatics Research Group, Hagenberg, Austria
- (6) Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Denmark
- (7) European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK
- (8) Department of Biochemistry and Molecular Biology, University of Southern Denmark, Odense, Denmark
- (9) Ruhr University Bochum, Center for Protein Diagnostics (PRODI), Medical Proteome Analysis, Bochum, Germany
- (10) Ruhr University Bochum, Medical Faculty, Medizinisches Proteom-Center, Bochum, Germany
- (11) VIB-UGent Center for Medical Biotechnology, VIB, Ghent, Belgium
- (12) Department of Biomolecular Medicine, Ghent University, Ghent, Belgium

* Corresponding author (tim.vandenbossche@ugent.be)

Abstract

The European Bioinformatics Community for Mass Spectrometry (EuBIC-MS; eubic-ms.org) was founded in 2014 to unite European computational mass spectrometry researchers and proteomics bioinformaticians working in academia and industry. Since then, we have organised two Winter Schools and two Developers' Meetings, which have contributed to strengthening the European mass spectrometry network and which have fostered international collaboration in this field, even beyond Europe. Besides these conferences, EuBIC-MS members also maintain educational resources (proteomics-academy.org) and organise workshops at national and international conferences on proteomics and mass spectrometry. Moreover, we are currently actively developing a community-driven standard dedicated to mass spectrometry data annotation (SDRF-Proteomics) that will facilitate data reuse and collaboration. Besides these activities originating from within EuBIC-MS, we are also involved in several other community initiatives such as the Human Proteome Organization's Proteomics Standards Initiative. In this manuscript, we highlight what EuBIC-MS is, what we do, and what we have already achieved. We would also like to extend a warm invitation to new researchers at all career stages to join the EuBIC-MS community on our Slack channel (eubic.slack.com).

Brief communication

The European Bioinformatics Community for Mass Spectrometry (EuBIC-MS; <https://eubic-ms.org/>) is a community for proteomics bioinformatics and computational mass spectrometry (MS). EuBIC-MS aims to bring together European MS bioinformatics researchers, including students and early-career researchers, as well as long-standing experts from both academia and industry. Through the setup of community-driven initiatives, EuBIC-MS promotes education in bioinformatics and computational mass spectrometry, highlights job and funding opportunities, fosters international collaborations, contributes to the publication of specialised studies and provides training of basic and advanced software tools.

EuBIC-MS is an official initiative of the European Proteomics Association (EuPA), and together they have created and maintained the Proteomics Academy website (<https://www.proteomics-academy.org/>), an online resource that contains educational material, including videos from educational talks, tutorials of popular software tools, keynote talks from our conferences, and bioinformatic tool descriptions. This platform also includes a Q&A message board to ask questions and discuss proteomics experiments, and an online job fair where vacant job offers can be posted and researchers can upload their CV. This job fair was created with the help of the Young Proteomics Investigators Club (YPIC), another EuPA Initiative. Apart from the Proteomics Academy, EuBIC-MS members also actively take part in international workshops and training events such as the “Integrative Biological Interpretation using Proteomics” course at the Gulbenkian Training Programme in Bioinformatics, Lisbon, Portugal (<http://gtpb.igc.gulbenkian.pt/bicourses/2019/IBIP19/index.html>). We also (co-)organise workshops at the bioinformatics hubs of the major international conferences on proteomics and mass spectrometry, including the Annual Congress of the European Proteomics Association (EuPA), the Human Proteome Organization (HUPO) World Congress, and the annual conference of the American Society for Mass Spectrometry (ASMS). These bioinformatics hubs form a central gathering place for conference attendees who are interested in computational mass spectrometry, and stimulate informal debates and lively discussions on topics of interest to the field. During these activities, EuBIC-MS strongly supports open science by promoting open data

policies, providing freely available training and teaching material, and collaborating to open source software development.

Additionally, EuBIC-MS organises a yearly conference dedicated to computational mass spectrometry and proteomics bioinformatics. This forms an important community outreach effort to bring together bioinformatics researchers, junior scientists, and industry partners from all over Europe. The EuBIC-MS Winter School, held every two years, is the major European conference dedicated to computational mass spectrometry and proteomics bioinformatics^{1,2}. This conference features five educational days dedicated to workshops and training sessions of computational mass spectrometry tools and workflows, as well as keynote lectures by internationally renowned speakers and practical workshops covering aspects of identification, quantification, result interpretation, and integration of mass spectrometry data. Alternating with the Winter School, the EuBIC-MS Developers' Meeting is organised every second year^{3,4}. The aim of the Developers' Meeting is to bring together bioinformaticians, computer scientists and developers in computational mass spectrometry to discuss and collaborate in an open and constructive spirit. These meetings act as incubators for new ideas and provide a structure for new open source projects to arise. (e.g. MegaGO⁵, Universal Spectrum Explorer (USE)⁶, IsoProt⁷, and an update of ThermoRawFileParser⁸). Although EuBIC-MS is a young initiative (only two Developers' Meetings took place to date), the Developers' Meetings have already nurtured new developments in computational proteomics and have provided the environment for community-driven implementation sprints. In addition to these collaborative developments, each EuBIC-MS Winter School and Developers' Meeting has resulted in the publication of a Proceeding¹⁻⁴.

Next to the collaborative projects conducted during the Developers' Meetings, we also stimulate scientific exchanges between laboratories and foster financial support for visits and travels between collaborating research groups in Europe. We also collaborate with major bioinformatics and proteomics consortia and infrastructures (**Figure 1**). Herein, several EuBIC-MS members have contributed to new ELIXIR implementation studies¹⁶. While these projects are federated and funded by the ELIXIR European infrastructure, the strong relationship between ELIXIR and EuBIC-MS gives young researchers the opportunity to widely strengthen their network. A current ongoing implementation study focuses on the setup and benchmarking of proteomics workflows specifically designed for execution in a cloud environment

(<https://elixir-europe.org/about-us/commissioned-services/proteomics-pipelines>). This will facilitate the reanalysis of datasets from public data repositories, such as PRIDE¹⁷, which is currently an important challenge for the proteomics community. Reuse of proteomics data also fundamentally relies on its metadata, such as the sample type, instrument parameters, and the experimental design of the underlying study. Recently, EuBIC-MS has started a collaborative project to tackle this challenge with the HUPO Proteomics Standards Initiative (HUPO-PSI)¹⁸, an initiative that defines community standards for data representation in proteomics and interactomics to facilitate data comparison, exchange and verification, e.g. the well-established mzML format¹⁹, and the more recent PSI Extended Fasta Format (PEFF)²⁰ and the Universal Spectrum Identifier (USI)²¹. Here, we have proposed a new proteomics metadata standard: the Sample and Data Relationship Format for Proteomics file format (SDRF-Proteomics)²². This standard is cross-compatible with the MAGE-TAB format²³, which is widely used in the transcriptomics community. As of December 2020, 146 ProteomeXchange²⁴ projects have been annotated, and this number is expected to increase quickly in 2021, after the official publication of the SDRF-Proteomics file format.

Since its creation five years ago, EuBIC-MS has contributed to the establishment of a new computational mass spectrometry community that provides research opportunities for its members. This has led to unprecedented interactions between bioinformaticians and biostatisticians working in the MS and proteomics fields and to the emergence of several community-driven projects. As ambassadors of open science and open software development we consider open research to be of major importance, as this improves the quality and sustainability of bioinformatics solutions in the long term. In the future, EuBIC-MS plans to strengthen its relationship with other EuPA initiatives and community-driven initiatives. Moreover, we plan to reach out to other computational mass spectrometry fields, such as computational lipidomics and metabolomics, and support developments, standards, and training that matter to the entire community.

EuBIC-MS is always on the lookout for enthusiastic bioinformaticians and computer scientists at all career stages who would like to become members of our open community, and therefore also gain access to an extensive network of prominent bioinformaticians and computational mass spectrometrists. EuBIC-MS welcomes everyone interested in computational MS and proteomics

bioinformatics from all over the world. There is no official registration or expectation of active engagement. If you want to help out with organising meetings, maintaining EuBIC-MS web resources, providing support, join our discussions, or just see what EuBIC-MS members are up to, please join our Slack channel (eubic.slack.com).

Acknowledgements

We would like to thank the European Proteomics Association (EuPA) for the support of this initiative, and all EuBIC-MS members for their engagement and continuous effort. We would like to acknowledge the Research Foundation – Flanders (SB grant 1S90918N to TVDB; postdoctoral grant 12W0418N to WB). MLP is supported financially by the Novo Nordisk Foundation (Grant agreement NNF14CC0001).

References

1. Willems S, Bouyssié D, David M, et al. Proceedings of the EuBIC Winter School 2017. *J Proteomics*. 2017;161:78-80.
2. Kopczynski D, Bittremieux W, Bouyssié D, et al. Proceedings of the EuBIC Winter School 2019. *EuPA Open Proteomics*. 2019;22-23:4-7.
3. Willems S, Bouyssié D, Deforce D, et al. Proceedings of the EuBIC developer's meeting 2018. *J Proteomics*. 2018;187:25-27.
4. Ashwood C, Bittremieux W, Deutsch EW, et al. Proceedings of the EuBIC-MS 2020 Developers' Meeting. *EuPA Open Proteomics*. 2020;24:1-6.
5. Verschaffelt P, Van Den Bossche T, Gabriel W, et al. MegaGO: a fast yet powerful approach to assess functional similarity across meta-omics data sets. *bioRxiv*. 2020; (biorxiv;2020.11.16.384834v1). doi:10.1101/2020.11.16.384834
6. Schmidt T, Samaras P, Dorfer V, et al. Universal Spectrum Explorer: A standalone (web-)application for cross-resource spectrum comparison. *bioRxiv*. Published online September 9, 2020. doi:10.1101/2020.09.08.287557
7. Griss J, Vinterhalter G, Schwämmle V. IsoProt: A Complete and Reproducible Workflow To Analyze iTRAQ/TMT Experiments. *J Proteome Res*. 2019;18(4):1751-1759.
8. Hulstaert N, Shofstahl J, Sachsenberg T, et al. ThermoRawFileParser: Modular, Scalable, and Cross-Platform RAW File Conversion. *J Proteome Res*. 2020;19(1):537-542.
9. Walzer M, Pernas LE, Nasso S, et al. qcML: an exchange format for quality control metrics from mass spectrometry experiments. *Mol Cell Proteomics*. 2014;13(8):1905-1913.

10. Wang X, Chambers MC, Vega-Montoto LJ, Bunk DM, Stein SE, Tabb DL. QC metrics from CPTAC raw LC-MS/MS data interpreted through multivariate statistics. *Anal Chem*. 2014;86(5):2497-2509.
11. Röst HL, Sachsenberg T, Aiche S, et al. OpenMS: a flexible open-source software platform for mass spectrometry data analysis. *Nat Methods*. 2016;13(9):741-748.
12. The M, Käll L. MaRaCluster: A Fragment Rarity Metric for Clustering Fragment Spectra in Shotgun Proteomics. *J Proteome Res*. 2016;15(3):713-720.
13. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498-2504.
14. Verschaffelt P, Van Den Bossche T, Gabriel W, et al. MegaGO: a fast yet powerful approach to assess functional similarity across meta-omics data sets. *Cold Spring Harbor Laboratory*. Published online November 17, 2020:2020.11.16.384834. doi:10.1101/2020.11.16.384834
15. Deutsch EW, Perez-Riverol Y, Carver J, et al. Universal Spectrum Identifier for mass spectra. *Cold Spring Harbor Laboratory*. Published online December 9, 2020:2020.12.07.415539. doi:10.1101/2020.12.07.415539
16. Vizcaíno JA, Walzer M, Jiménez RC, et al. A community proposal to integrate proteomics activities in ELIXIR. *F1000Res*. 2017;6:875.
17. Perez-Riverol Y, Csordas A, Bai J, et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res*. 2019;47(D1):D442-D450.
18. Deutsch EW, Orchard S, Binz P-A, et al. Proteomics Standards Initiative: Fifteen Years of Progress and Future Work. *J Proteome Res*. 2017;16(12):4288-4298.
19. Martens L, Chambers M, Sturm M, et al. mzML--a community standard for mass spectrometry data. *Mol Cell Proteomics*. 2011;10(1):R110.000133.
20. Binz P-A, Shofstahl J, Vizcaíno JA, et al. Proteomics Standards Initiative Extended FASTA Format. *J Proteome Res*. 2019;18(6):2686-2692.
21. Deutsch EW, Perez-Riverol Y, Carver J, et al. Universal Spectrum Identifier for mass spectra. *bioRxiv*. 2020;(biorxiv;2020.12.07.415539v1):160018. doi:10.1101/2020.12.07.415539
22. Perez-Riverol Y, European Bioinformatics Community for Mass Spectrometry. Toward a Sample Metadata Standard in Public Proteomics Repositories. *J Proteome Res*. 2020;19(10):3906-3909.
23. Rayner TF, Rocca-Serra P, Spellman PT, et al. A simple spreadsheet-based, MIAME-supportive format for microarray data: MAGE-TAB. *BMC Bioinformatics*. 2006;7(1):365.

24. Deutsch EW, Csordas A, Sun Z, et al. The ProteomeXchange consortium in 2017: supporting the cultural change in proteomics public data deposition. *Nucleic Acids Res.* 2017;45(D1):D1100-D1106.

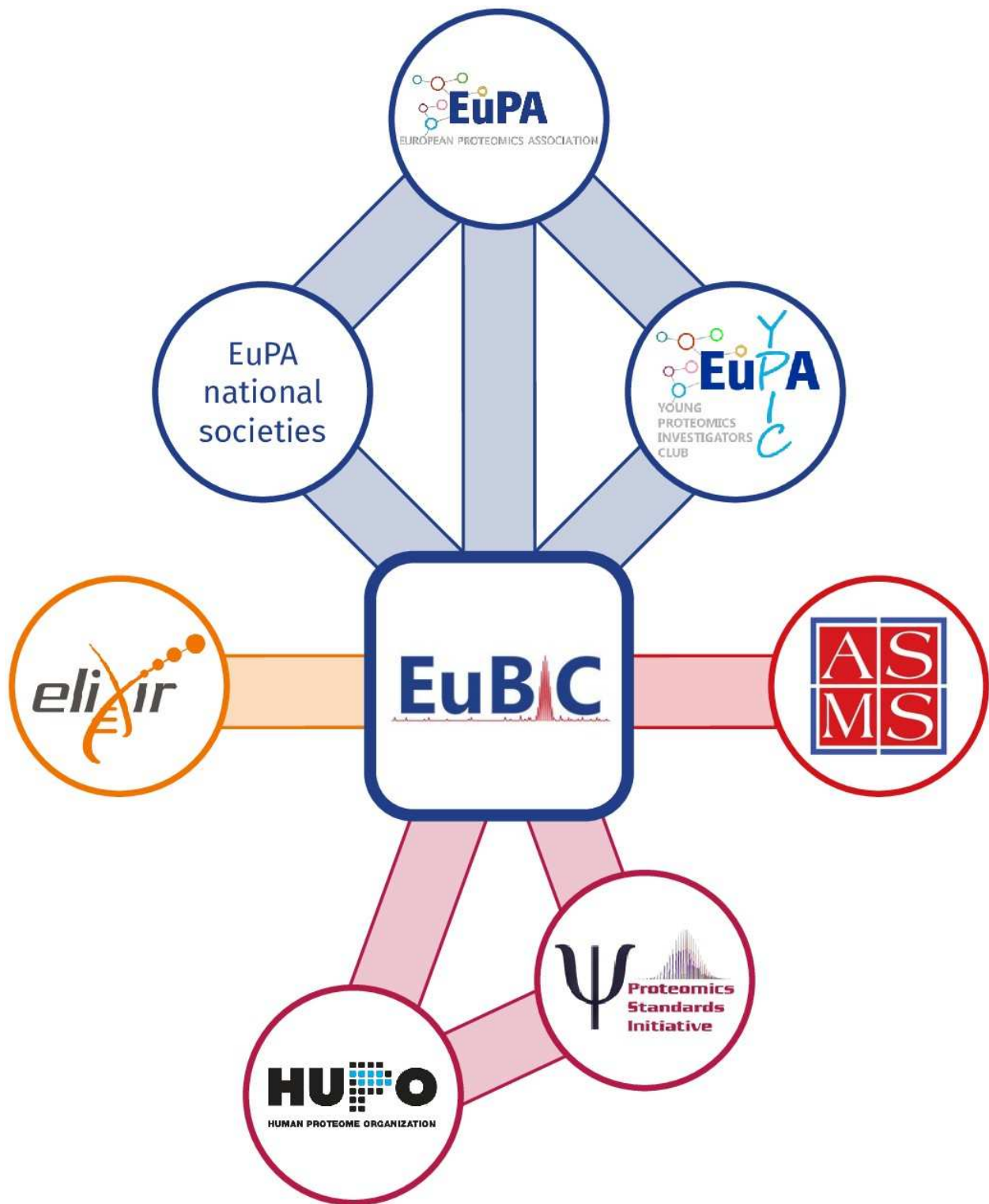


Figure 1. EuBIC-MS and its collaborators.