**README document for Improving Peer Online Forums (iPOF) Datasets**

**Forum Raw Data**

**Prof. Fiona Lobban, Spectrum Centre, Department for Mental Health Research, Lancaster University.** **F.lobban@lancaster.ac.uk**

**Date of data collection: 1/3/2022-31/12/2024**

This project is funded by the National Institute for Health and Social Care Research (NIHR), 134035. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

The further information file contains:

* The “iPOF Forum Codes- bird names” spreadsheet contains the mapping between the anonymised bird names and their corresponding real forum names.
* IPOF Triangulation file which allows data from the survey, forum and interviews to be triangulated.
* Summary of Forum Databases file which gives information on where the data came from and when and the format.

Within the ORIGINALS folder, there are two subfolders “2023” and “2024” representing the two forum data collection time points. Within each year subfolder, there are further subfolders containing the data downloaded for each forum.

Depending on the collection method, these are stored as “tar.gz”, “xls” (Excel spreadsheet), or “csv” (comma separated values). CSV format is a plain text files that stores tabular data, and can be opened with a number of software applications including spreadsheet tools. Tarfiles can be expanded using 7-Zip or WinZip in Windows, or via the ‘tar’ command on Linux and MacOS command lines.

**This dataset should be kept closed and not shared. It is on PURE for storage only.**

The forum data was collected with the permission of the forum providers or moderators via bespoke Python corpus scraping scripts developed by Matthew Coole and John Vidler of the UCREL NLP group (<https://ucrel.lancs.ac.uk/>), or shared with us directly from the forum providers using database export scripts of their own.

Methodological information is available from the study protocol and other project publications:

Glossop, Z., Čermáková, A. L., Marshall, P., Rayson, P., Robinson, H., Semino, E., & Lobban, F. (2025). Who uses mental health support forums, and why? Triangulating findings from surveys, interviews, and forum posts. medRxiv. <https://doi.org/10.1101/2025.05.11.25327409>

Marshall, P., Booth, M., Coole, M., Fothergill, L., Glossop, Z., Haines, J., Harding, A., Johnston, R., Jones, S., Lodge, C., Machin, K., Meacock, R., Nielson, K., Puddephatt, J.-A., Rakic, T., Rayson, P., Robinson, H., Rycroft-Malone, J., Shryane, N., ... Lobban, F. (2024). Understanding the Impacts of Online Mental Health Peer Support Forums: Realist Synthesis. JMIR Mental Health, 11, Article e55750. <https://doi.org/10.2196/55750>

Coole, M., Rayson, P., Glossop, Z., Lobban, F., Marshall, P., & Vidler, J. (2024). Exploring the Suitability of Transformer Models to Analyse Mental Health Peer Support Forum Data for a Realist Evaluation. In D. Demner-Fushman, S. Ananiadou, P. Thompson, & B. Ondov (Eds.), Proceedings of the First Workshop on Patient-Oriented Language Processing (CL4Health) @ LREC-COLING 2024 (pp. 184-188). ELRA and ICCL. <https://aclanthology.org/2024.cl4health-1.22>

Coole, M., Marshall, P., Rayson, P., & Lobban, F. (2023). iPOF: Improving Peer Online Forums Update. Poster presented at Healthcare text analytics conference 2023, Manchester, United Kingdom.

Lobban, F., Coole, M., Donaldson, E., Glossop, Z., Haines, J., Johnston, R., ... & Wise, S. (2023). Improving Peer Online Forums (iPOF): protocol for a realist evaluation of peer online mental health forums to inform practice and policy. BMJ open, 13(7), e075142.

Lobban, F., Rayson, P., & Coole, M. (2022). iPOF: Improving Peer Online Forums. Poster presented at HealTAC 2022: 5th Healthcare Text Analytics Conference, United Kingdom. https://youtu.be/ryA\_obyJnfU

Ethical approval: Solihull Research Ethics Committee. 20 June 2022 (IRAS314029).