LabPipe: A Modular Informatics Toolkit for Sample Collection and Data Integration

Zhao B[1], Wilde M[2], Cordell C[2], Salman D[5], Bryant L[2], Ruszkiewicz D[5], Ibrahim W[1], Singapuri A[1], Parmar A[4], Coats T[3], Gaillard E[1], Beardsmore C[1], Suzuki T[3], Ng L[3], Blower J[6], Greening N[1], Thomas P[5], Brightling C[1], Siddiqui S[1] and Free R[1]

[1]Department of Infection, Immunity and Inflammation, [2]Department of Chemistry, [3]Department of Cardiovascular Sciences, [4]Research and Enterprise Partnerships, University of Leicester, UK; [5]Centre for Analytical Science, Loughborough University, UK; [6]Leicester Fertility Centre, UHL Trust, UK

Contacts: Dr Bo Zhao bo.zhao@Leicester.ac.uk, Dr Robert Free rcf8@Leicester.ac.uk

Background

The EMBER (East Midlands Breathomics Pathology Node) project is a multi-disciplinary project which aims to discover how breathomics can be used to diagnose cardiorespiratory diseases. Data generation and collection is a key part of this project. In EMBER, data is generated using various analytical chemistry techniques and tools across different remote sites with each technique and site having distinct operational and data handling protocols.

Previous research shows that a sustainable and trustable data pipeline that directly serves endpoints from data collection to statistical analysis is critical to networked medical researches[1]. Therefore, while single solutions exist for handling aspects of data management, we aimed to create an integrated solution to handle the combined integration of direct chemometric data collection, sample tracking and clinical data from remote sites.

Methods

We created custom informatics solutions to handle this effectively, i.e. LabPipe, which is an informatics toolkit which supports sample/data collection from multiple hospital sites for different analytical chemistry equipment. The toolkit also includes data integration; user notifications and data warehouse linkage.

- A centrally hosted data warehouse is designed to facilitate the process, storage and query of fragmented data from multiple locations, instruments, visits and handlers with different requirements, for the purpose of data reuse and managed access[2][3].
- A data collection tool is also designed to be deployed with each instrument to help organise data and related information under standardised structure prior to data warehouse entry[4].
- A web service is created as a bridge to link remote data collection tool and data warehouse together.
- To extend usage scenarios and to provide additional functions, plugins are available for the web service.

Structure

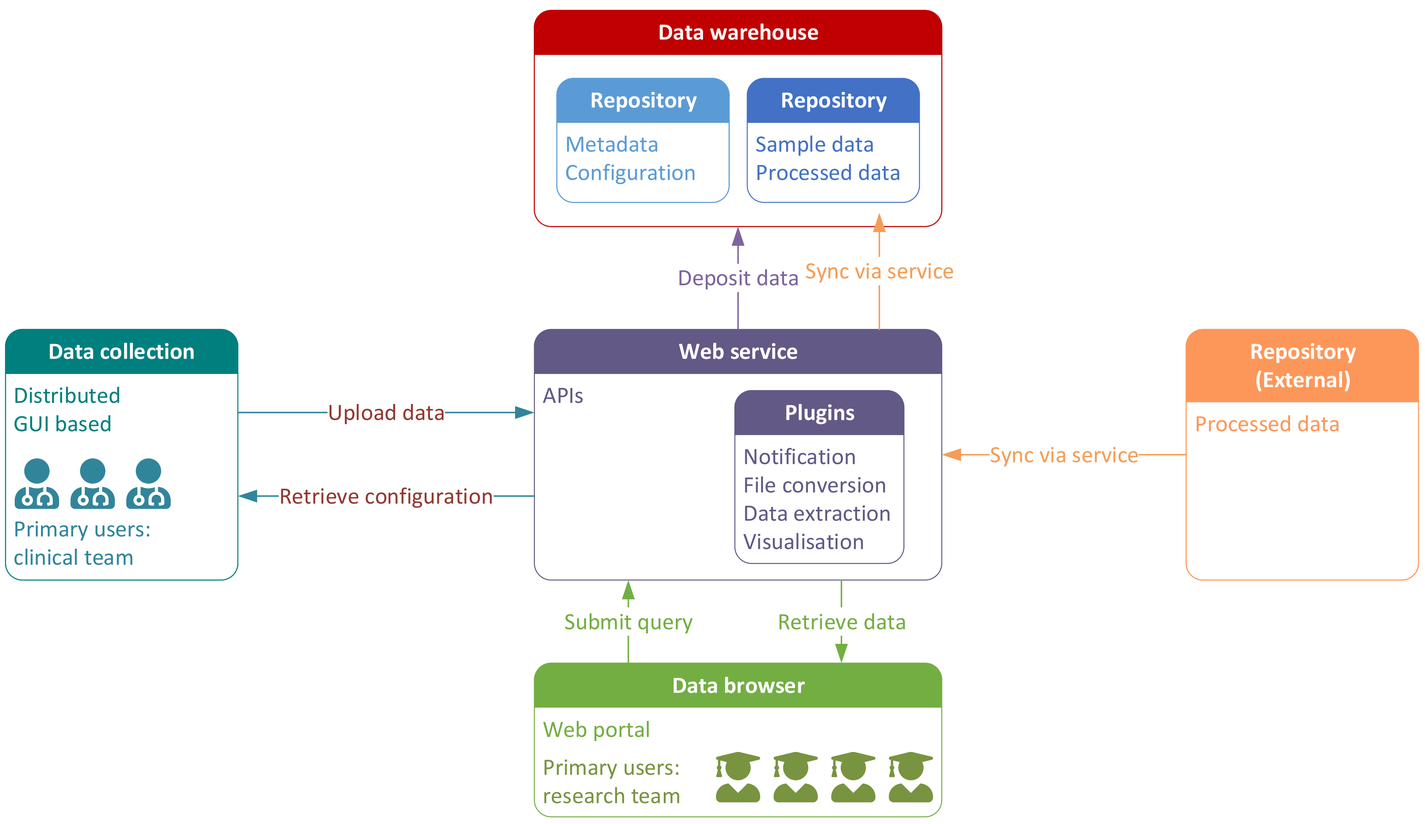

Figure 1. LabPipe modular structure

The diagram (Fig. 1) above illustrates the prime structure of LabPipe.

- **Data warehouse:** consists of a metadata repository and a sample data repository.
  - The metadata repository is used to store relevant information of each sample process, study and instrument specific configurations.
  - The sample data repository is used to store raw sample data and post-collection processed data of each analysis stage.
- **External repository:** used by collaborators for storage of processed data and synced to the data warehouse sample data repository.
- **Data collection:** designed to be a user interface based tool to assist sample collection process on labelling, structuring metadata, and prepare the record for transfer. Configurations are loaded from the metadata repository for each supported scenario.
- **Data browser:** is a web-based portal, which can be used to query, extract cohorts and linked data.
- **Web service:** used as a bridge to link all modules together and supports functionalities of other modules.
  - **Plugin:** can be used to provide additional functionalities to the web service, e.g. sample data pre-processing, user notification at different stages, file format conversion, data extraction and visualisation.

Results

LabPipe handles analytical chemistry data integration, including real-time notifications to researchers and users, as well as bridges the data warehouse. The system has been continuously running for over a year with active maintenance to improve performance and to meet new feature requests.

LabPipe has also reduced the person effort required to maintain data consistency, integrate data and perform quality control checks. A survey was carried out among the users with average 9 months use per user/scenario. Feedbacks indicate the toolkit is stable, reliable and efficient. Across all instruments deployed with LabPipe, over 90% of the scenarios have reported the tool as useful, with 55% reported very useful. Compared with paper-based record, 80% of the scenarios have reported very useful, with average 68% time saved.

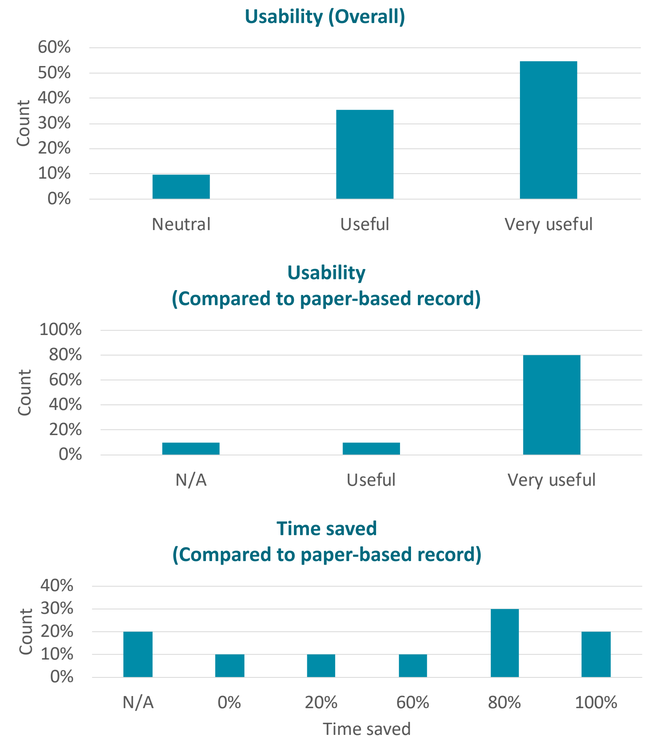

Figure 2. LabPipe user survey key results

Methods

LabPipe has helped to streamline the flow of data from the clinic to the researchers including notifying users of potential problems with data collection. The modular approach taken to the design of LabPipe has allowed it to be used across multiple studies and enabled different analytical chemistry data formats to be handled effectively.

Acknowledgements

This work is part of the EMBER project. The authors gratefully acknowledge the helps from our lovely clinical research team : Sheila Jones, Bharti Patel, Awal Asia, Liesl Carr, Rachael Phillips, Teresa McNally, Clare Foxon, Graham Fowkes, Sarah Parker, Joanne Finch, Vijay Mistry, Sarah Edwards and Dr. Ahmed Yousuf.

References

1. Bauer, C., Umbach, N., Baum, B., Buckow, K., Franke, T., Grütz, R., Gusky, L., Nussbeck, S., Quade, M., Rey, S., Rottmann, T., Rienhoff, O., Sax, U., 2016. Architecture of a Biomedical Informatics Research Data Management Pipeline. St Heal T 228, 262–6.
2. Holmes, J., Elliott, T., Brown, J., Raebel, M., Davidson, A., Nelson, A., Chung, A., Chance, P., Steiner, J., 2014. Clinical research data warehouse governance for distributed research networks in the USA: a systematic review of the literature. J Am Med Inform Assn 21, 730–736.
3. Khnaisser, C., Lavoie, L., Diab, H., Ethier, J.-F., 2015. Data Warehouse Design Methods Review: Trends, Challenges and Future Directions for the Healthcare Domain. springer, pp. 76–87.
4. Cho, M., Taylor, H., McCormick, J., Anderson, N., Barnard, D., Boyle, M., Capron, A., Dorfman, E., Havard, K., Reider, C., Sadler, J., Schwartz, P., Sharp, R., Danis, M., Wilfond, B., 2015. Building a Central Repository for Research Ethics Consultation Data: A Proposal for a Standard Data Collection Tool. Clin Transl Sci 8, 376–387.