

International Journal of Population Data Science

Journal Website: www.ijpds.org



Bibliometric Analysis to Scan and Scrape New Datasets: It's all about that BASS

Tingay, Karen^{1*}, Perkins, Brian¹, and Anastasiou, Athanasios¹

¹Swansea University

Objectives

The main objective of this poster is to present a pilot project in determining emerging population health themes and identifying key research-enabling datasets ahead of time.

At present, large-scale databanks, such as the Secure Anonymised Information Linkage (SAIL) Databank at Swansea University Medical School, already manage large quantities of health and administrative linked datasets. While these datasets are valuable for research purposes, complementary datasets may be required by collaborating researchers to answer detailed population health research questions. Dataset acquisition can take several years, which is a serious delay to a project with time-limited funding.

The ability to pre-emptively acquire datasets so that these are ready for use before a researcher requests them would obviously be beneficial. However, a recent study conducted by the Farr Cipher team at Swansea University identified over 800 health and administrative datasets in Wales alone.

With limited resources such as available funding and time, which of these datasets is worth its effort in acquiring?

Approach

Bibliometrics has long been a means of measuring the impact of papers on the wider academic community. Lately, the focus of analyses has been extended to include the topics, authorship and citations of the publications. Existing bibliometric data mining techniques suggest that it is possible to identify emerging topic trends and through this assist in prioritising dataset identification and acquisition.

The project explored mining available literature through bibliometric analysis in order to predict emerging trends and through these identify potentially relevant and valuable datasets for acquisition on behalf of the Dementias Platform UK (DPUK). Literature searches were conducted for papers published on the topic of "dementia" over the last 20 years. Additional keywords

and topics were extracted to identify emerging areas of research and clinical interest. These were then compared against an existing list of over 800 Welsh datasets currently not held in SAIL.

Results

Results focus on:

- Using bibliometric methods in the context of DPUK cohort publications
- Identifying emerging trends in the field of dementia research.
- Identifying and prioritising datasets which might be useful for the SAIL Databank to acquire

*Corresponding Author:

Email Address: k.s.tingay@swansea.ac.uk (K. Tingay)

