

# DEVELOPMENT OF AN ONLINE PORTAL AND METADATA CATALOGUE OF EARTH OBSERVATION DATA TYPES, SOURCES AND PRODUCTS FOR HUMAN HEALTH RESEARCH IN EXPOSOMICS

M. N. Kamel Boulos<sup>1\*</sup>, M. V. Iyyanki<sup>2</sup>, A. Dewan<sup>3</sup>, B. Bwambale<sup>4</sup>, K. Koh<sup>5</sup>

<sup>1</sup> School of Medicine, University of Lisbon, 1649-028 Lisbon, Portugal – mnkboulos@ieeee.org

<sup>2</sup> JN Technological University Hyderabad, Hyderabad, 500008, India - ivm@ieeee.org

<sup>3</sup> School of Earth & Planetary Sciences, Curtin University, Perth, WA 6845, Australia - a.dewan@curtin.edu.au

<sup>4</sup> Mountains of The Moon University, P. O. Box 837, Fort-Portal, Uganda - bwmbllbosco@gmail.com

<sup>5</sup> Dept. of Geography, Faculty of Social Sciences, The University of Hong Kong, Hong Kong 999077, China - peterkoh@hku.hk

**KEY WORDS:** Exposomics, Human exposome, Earth observation data, Remote sensing, Data sources, Metadata catalogue.

## 1. INTRODUCTION

Predisposition and development of various diseases, communicable, such as malaria and dengue fever, and non-communicable, such as diabetes, cardiovascular diseases, various types of cancer and mental health problems, involve a complex interplay between genetic factors (the genome) and environmental and lifestyle parameters that populations are exposed to (the exposome). When combined with other relevant data, remote sensing data from various sources/products and of different types can help us better map and investigate the latter (exposomic determinants of disease / population-level exposures). However, these Earth observation data sources and types remain a largely untapped resource for many researchers in the fields of population health and medicine, who are not familiar with the potential and value of these data and the unique insights that can be revealed by using them. Examples of such data obtained via remote sensing include data about air pollutants, land cover, green space, nocturnal outdoor light pollution and noise pollution, among others (Cui et al., 2022; Kamel Boulos and Le Blond, 2016; Kamel Boulos and Wilson, 2023; Sogno et al., 2020; Urbano et al., 2021).

This article introduces an ISPRS-funded project entitled ‘geospatial database for exposomics’. This project is funded under ISPRS Scientific Initiatives 2023 (SI2023 Awards), see [www.isprs.org/news/announcements/details.aspx?ID=314](http://www.isprs.org/news/announcements/details.aspx?ID=314), and involves proof concepting and disseminating a much-needed metadata catalogue of Earth observation data sources/products and types that are relevant to human health research as a free service to interested researchers worldwide. It is hoped future, more comprehensive versions of this service will enable more researchers and studies to discover and use remote sensing data about population-level exposures to disease determinants to reveal fresh insights that could improve our understanding of the aetiology, pathogenesis and spread of relevant diseases, and hence contribute to the development of better-optimised prevention and management programmes to tackle them.

## 2. METHODS

The catalogue is informed by extensive input from existing research literature on the subject (desk research) as well as online communications with, and relevant research publications collected from, a small panel of select ‘stakeholder representatives’ from the academia in three countries (China,

UK and USA) to guarantee some basic form of user engagement and involvement in the project from its very early stages, which is always highly desirable. Stakeholder representatives are providing us with valuable insights and intelligence about relevant remote sensing applications and available products in public health, in addition to important feedback on our catalogue during its development over two iterations in 2023 to help us improve it to best serve their needs and those of the wider research community they represent.

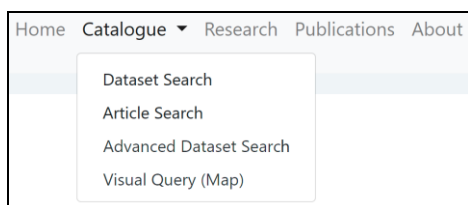
The research team compiled the catalogue based on all collected input in the form of an Excel workbook. Our Web developer then converted the Excel workbook into an online database and developed the project’s online portal in MySQL, an open-source relational database management system, Python, a high-level, general-purpose programming language, and Django, an open source, Python-based Web framework. The portal’s public user interface was designed with both desktop and mobile devices in mind. The portal also features an access-protected ‘Catalogue Management Interface’ where the catalogue curator can modify existing catalogue records or add new ones as necessary.

## 3. RESULTS

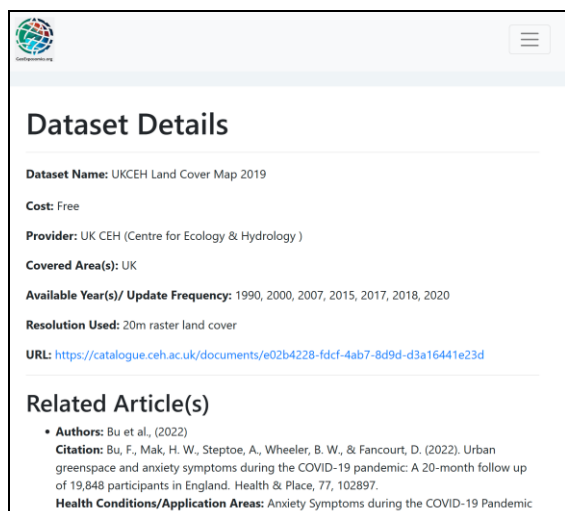
The catalogue currently has 90 metadata records of relevant Earth observation datasets (n = 40) and associated health-focused research publications (n = 50). Dataset and research publication records are stored in two separate but fully cross-referenced Excel worksheets/database tables. For each catalogued record, the following metadata elements were compiled: dataset name, provider, available year(s)/update frequency, cost (licence), covered geographical areas, resolution used, URL, related research publications, and the health conditions/application areas that featured in those publications.

The project’s portal offers a searchable version of the catalogue, see [www.geoexposomics.org](http://www.geoexposomics.org) (Figures 1 and 2). It allows users to search and filter dataset records by health conditions/application areas, cost, covered area(s), provider, or a combination of one or more of these elements. Users can also filter the associated research article records by health conditions/application areas and/or dataset name. The evaluation feedback received from the same panel of stakeholder representatives about the online portal was positive and provided valuable suggestions that were later implemented in a second iteration of the interface (Figures 3 and 4).

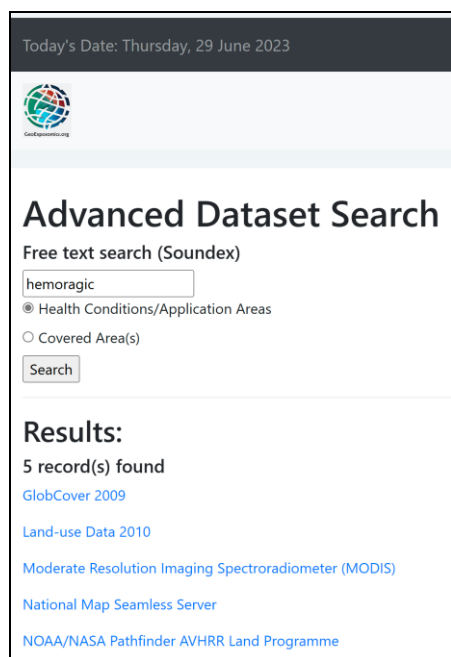
\* Corresponding author



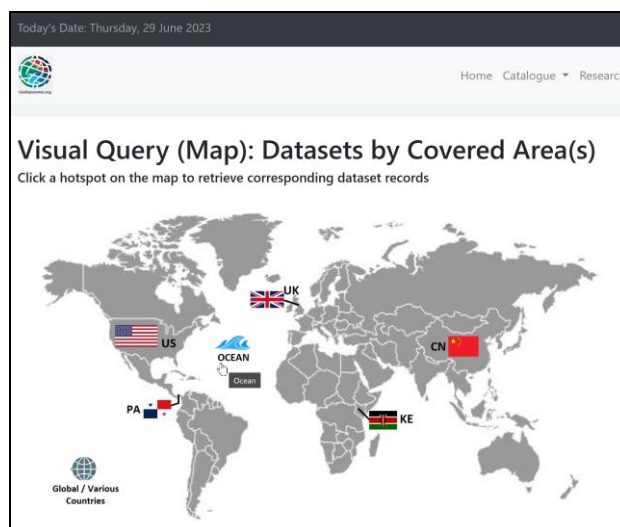
**Figure 1.** An online searchable version of the catalogue can be accessed by clicking the ‘Catalogue’ menu on the project portal’s navigation bar. Various search modes and options are currently available. Future directions are briefly discussed at [www.geoexposomics.org/catalogue/research](http://www.geoexposomics.org/catalogue/research).



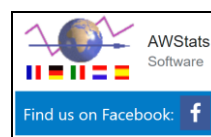
**Figure 2.** A typical metadata record in the online catalogue.



**Figure 3.** Following user evaluation feedback, we implemented a free text search option of datasets (by health applications or covered areas) using a custom Soundex algorithm. As an example of how this algorithm works, searching health conditions for ‘haemorrhagic fever’ (British spelling), ‘hemorrhagic fever’ (American spelling), or ‘hemoragic fever’ (typo/misspelling) will all retrieve the correct/same results.



**Figure 4.** User evaluation feedback included a suggestion to provide a handy “online map to help locating datasets”. In response, we introduced an option allowing users to click location ‘hotspots’ on a map of the world with country flags and other relevant icons to retrieve corresponding dataset records.



**Figure 5.** Dissemination: Between 15–30 June 2023, the portal received >200 unique visitors from more than 25 countries according to anonymous server log statistics generated by AWStats tool using IP geolocation. We also have a companion Facebook page where relevant news items and updates are regularly shared with a growing community, see [www.facebook.com/geoexposomics](https://www.facebook.com/geoexposomics).

## REFERENCES

- Cui, Y., Eccles, K.M., Kwok, R.K., Joubert, B.R., Messier, K.P., Balshaw, D.M., 2022: Integrating Multiscale Geospatial Environmental Data into Large Population Health Studies: Challenges and Opportunities. *Toxics* 10(7), 403. doi.org/10.3390/toxics10070403
- Kamel Boulos, M.N., Le Blond, J., 2016: On the road to personalised and precision geomedicine: medical geology and a renewed call for interdisciplinarity. *International Journal of Health Geographics* 15(5). doi.org/10.1186/s12942-016-0033-0
- Kamel Boulos, M.N., Wilson, J.P., 2023: Geospatial techniques for monitoring and mitigating climate change and its effects on human health. *International Journal of Health Geographics* 22(2). doi.org/10.1186/s12942-023-00324-9
- Sogno, P., Traidl-Hoffmann, C., Kuenzer, C., 2020: Earth Observation Data Supporting Non-Communicable Disease Research: A Review. *Remote Sensing* 12(16):2541. doi.org/10.3390/rs12162541
- Urbano, T., Vinceti, M., Wise, L.A., Filippini, T., 2021: Light at night and risk of breast cancer: a systematic review and dose–response meta-analysis. *International Journal of Health Geographics* 20(44). doi.org/10.1186/s12942-021-00297-7