Science of the Total Environment 575 (2017) 79–86

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Development of a browser application to foster research on linking climate and health datasets: Challenges and opportunities

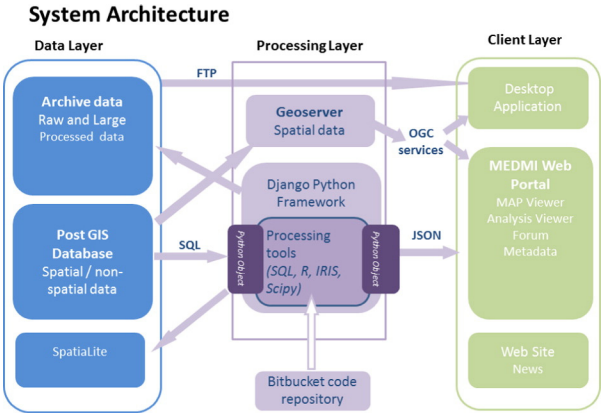
Shakoor Hajat [a,*], Ceri Whitmore [b], Christophe Sarran [c], Andy Haines [a], Brian Golding [c], Harriet Gordon-Brown [b], Anthony Kessel [d], Lora E Fleming [b]

[a] Department of Social & Environmental Health Research, London School of Hygiene & Tropical Medicine, London, UK
[b] European Centre for Environment and Human Health, University of Exeter Medical School, Truro, Cornwall, UK
[c] Met Office, Exeter, UK
[d] Public Health England, London, UK

HIGHLIGHTS

- Associations between public health and environmental factors are often complex.
- Big data analytics have the potential to provide new insights into such associations.
- This article documents the development of a browser application to facilitate assessments.
- Factors that shaped development are discussed, including licensing limits and data confidentiality.
- The issues raised can inform the development of similar tools by others working in this field.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 30 May 2016
Received in revised form 30 August 2016
Accepted 19 September 2016
Available online xxxx

Editor: D. Barcelo

Keywords:
Big data
Browser application
Human health
Environment
Climate change
Time-series regression

ABSTRACT

Background: Improved data linkages between diverse environment and health datasets have the potential to provide new insights into the health impacts of environmental exposures, including complex climate change processes. Initiatives that link and explore big data in the environment and health arenas are now being established.
Objectives: To encourage advances in this nascent field, this article documents the development of a web browser application to facilitate such future research, the challenges encountered to date, and how they were addressed.
Methods: A 'storyboard approach' was used to aid the initial design and development of the application. The application followed a 3-tier architecture: a spatial database server for storing and querying data, server-side code for processing and running models, and client-side browser code for user interaction and for displaying data and results. The browser was validated by reproducing previously published results from a regression analysis of time-series datasets of daily mortality, air pollution and temperature in London.
Results: Data visualisation and analysis options of the application are presented. The main factors that shaped the development of the browser were: accessibility, open-source software, flexibility, efficiency, user-friendliness, licensing restrictions and data confidentiality, visualisation limitations, cost-effectiveness, and sustainability.

* Corresponding author at: Department of Social & Environmental Health Research, London School of Hygiene & Tropical Medicine, 15-17 Tavistock Place, London WC1H 9SH, UK.
E-mail address: Shakoor.hajat@lshtm.ac.uk (S. Hajat).

http://dx.doi.org/10.1016/j.scitotenv.2016.09.162
0048-9697/© 2016 Elsevier B.V. All rights reserved.

Conclusions: Creating dedicated data and analysis resources, such as the one described here, will become an increasingly vital step in improving understanding of the complex interconnections between the environment and human health and wellbeing, whilst still ensuring appropriate confidentiality safeguards. The issues raised in this paper can inform the future development of similar tools by other researchers working in this field.
© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Although population health is closely linked to environmental factors, demonstrating associations can often be hampered by the lack of both common tools and databases available for research. Such limitations may become more apparent in future in the context of climate change, with many health risks of climate and other global environmental changes likely to be mediated by complex, often distal, pathways (Morris et al., n.d.). These risks will span a greater variety of mechanisms, non-linear relationships and spatio-temporal scales than epidemiologists are traditionally used to assessing (McMichael, 2013). Improved data linkages between environment, health and socio-economic datasets have the potential to overcome some of these new challenges of integrating complex information that is both spatially and temporally diverse (Reis et al., 2015). Such 'data mash-ups' can lead to new and innovative uses of environment and health data by a wide range of analysts, including those assessing complex climate change impacts (Lloyd et al., 2011; Semenza and Menne, 2009).

Initiatives that link and explore big data in the environment and human health arenas are now being established (U.S. Federal Government, 2014; The World Bank Group, 2014). A recent article highlighted the potential for big data to inform decision support on climate change and health and introduced the MED-MI (Medical and Environmental Data – a Mashup Infrastructure) partnership, which has been set-up with the primary aim to explore the creation of a central data and analysis source as an internet-based platform to provide a vital new common resource for public health research in the UK and elsewhere (Fleming et al., 2014).

Integral to initiatives such as MED-MI is the facilitation, with appropriate safeguards, of access by analysts to multiple, linked, health and environment databases so that customised analyses can be undertaken which will provide characterisation and quantification of a range of health and wellbeing effects of climate, weather and other environmental exposures. In the case of MED-MI, a web browser application has been developed to aid this process. This refers to a program that is created in a browser-supported programming language and relies on a web browser to render the application. One key aspect of the MED-MI browser application is that it has been specifically designed to allow any interested parties to explore hypotheses using the available environment and health data and to conduct appropriate statistical and other analyses, including visualisation, without the need for detailed knowledge of the underlying epidemiological methods employed or the technical skills and software usually required.

In order to encourage advances in this nascent field, this article documents the initial development of the browser application, the challenges encountered, and how they have been addressed to date. A 'storyboarding approach' was adopted to aid development of the web application. This approach is in common usage in software design and refers to a graphic organiser that provides the developer with a high-level view of the process (Little, 2013; Microsoft Dynamics CRM Team Blog, 2006). The approach can serve as a co-creation interface between the software developers/computer scientists and other researchers. Although in this article we describe the development of the browser designed specifically for MED-MI, we anticipate that most of the issues raised are sufficiently generic to help inform the future development of similar browser applications. The functionality of our application is demonstrated with a study design that will be familiar to many environmental epidemiologists, namely a time-series regression analysis (Katsouyanni et al., 1996). The article concludes by discussing the confidentiality issues raised by the potential sharing of sensitive data, and the main factors that we believe should inform the future development of similar tools by other researchers working in the environmental health field.

2. Material and methods

2.1. Browser location and architecture

The web application is part of the MED-MI platform located on the MED-MI server, which is hosted by the University of Exeter and is also the repository of the datasets used by the application. The web address for the platform is www.data-mashup.org.uk. The platform is hosted on its own dedicated server. It is via the platform that the browser allows for access to user-selected subsets of the data.

The application was developed with a proposed 3-tier architecture: a spatial database server for storing and querying data, server-side code for processing and running models against the data, and then client-side browser code for user interaction and for displaying data and results (fig. 1). A key challenge for development was the difference in research cultures, languages and analytical approaches traditionally employed between the environment and health communities, although tools such as Geographic Information Systems (GIS) can straddle both. The standardisation of spatial data services by the Open Geospatial Consortium (OGC) (http://www.opengeospatial.org) has enabled interoperability between systems for the global geospatial community. And the need for OGC standards to be adopted by the health community are being increasingly recognised (Samarasundera et al., 2014). Although the production of spatial information via the browser is currently limited, this will be the focus of future development.

Each layer of the application has a clearly defined role, as well as being loosely coupled to the other layers of the architecture. This approach is important as the technologies involved in storing and processing different types of data change rapidly, and the aim was to not be tied to a specific technology. Being loosely coupled meant that the data-storage layer could be changed, with only minimal amendments to the processing layer and no amendments to the client code. The different layers also enable greater security of the underlying datasets which may contain confidential information. Access to the data is provided by python software modules (Python Software Foundation, 2001).

For the relationship between the server and the client/user, the loosely coupled goal was achieved by using the JavaScript Object Notation (JSON) web standard for communication (ECMA-404, 1999). This use of the software design principle 'separation of concern' meant that the client does not need to understand how the server runs the model, but only that a result will be received as JSON in a defined format. This allows for the models, or even the language running the model, to change without affecting the client code, thereby increasing the potential flexibility and sustainability of the browser.

2.2. Browser development

As previously noted, a 'storyboard approach' was used to aid in the initial discussions and the design of the browser application. This approach was particularly useful as the members of the project working on the application came from different fields (e.g. computer science,

Download English Version:

https://daneshyari.com/en/article/6319396

Download Persian Version:

https://daneshyari.com/article/6319396

Daneshyari.com