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

Shakoor Hajat1, Ceri Whitmore2, Christophe Sarran3, Andy Haines1, Brian Golding3, Harriet Gordon-Brown2, Anthony Kessel4, Lora E Fleming2

1 Department of Social & Environmental Health Research, London School of Hygiene & Tropical Medicine, London, UK

2 European Centre for Environment and Human Health, University of Exeter Medical School, Truro, Cornwall, UK

3 Met Office, Exeter, UK

4 Public Health England, London, UK

Corresponding author:

Dr S Hajat

Department of Social & Environmental Health Research

London School of Hygiene & Tropical Medicine

15-17 Tavistock Place

London WC1H 9SH

UK

Email: [Shakoor.hajat@lshtm.ac.uk](mailto:Shakoor.hajat@lshtm.ac.uk)

Tel: +44 (0)20 7927 2512

Running title: Browser application for environmental health research

Acknowledgements: The research was funded in part by the UK Medical Research Council (MRC) and UK Natural Environment Research Council (NERC) for the MEDMI Project; the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Environmental Change and Health at the London School of Hygiene and Tropical Medicine in partnership with Public Health England (PHE), and in collaboration with the University of Exeter, University College London, and the Met Oﬃce; and the European Regional Development Fund Programme and European Social Fund Convergence Programme for Cornwall and the Isles of Scilly (University of Exeter Medical School).

Competing financial interests: None.

**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.

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.

**Keywords**

Big data, browser application, human health, environment, climate change, time-series regression

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

**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.[1] These risks will span a greater variety of mechanisms, non-linear relationships and spatio-temporal scales than epidemiologists are traditionally used to assessing.[2] 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.[3] 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.[4, 5]

Initiatives that link and explore big data in the environment and human health arenas are now being established.[6, 7] 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.[8]

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.[9, 10] 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.[11] 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.

**Material and methods**

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](file:///C:\Users\ppehshaj\AppData\Local\Temp\7\XPgrpwise\55D4BD7ELSHTM_POphpmailc10013268651126F11\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 (figure 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.[12] Although the production of spatial information via the browser is currently limited, this will be the focus of future development.

*Figure 1: MED-MI browser and server architecture*

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.[13]

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.[14] 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.

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, epidemiology, geographic information systems (GIS), meteorology, modelling, and database management). It enabled the computer scientist to understand the work flow of an epidemiologic analysis, and for the researchers to grasp the feasibility and limitations of developing such a tool to run within a browser.

The application was built through a partnership between the researchers and the computer scientist (author CW), using techniques adopted by the Agile approach to software development, such as iterative development and collaborative partner programming.[15, 16]

An iterative approach allowed for the researchers to view and check the interactions with the browser application, advising where changes needed to be made, as well as checking the analysis algorithms against more commonly used statistical tools such as STATA, at each cycle of the process. At the same time, this ‘fail-early approach’ meant that the computer scientist, who had limited statistical experience, did not take the browser application in the wrong direction.

A type of pair-programming was also used, whereby the researchers would advise the computer scientist whilst writing algorithms for the modelling code. This close interaction between the different team members gave crucial insights into the work involved as the project developed.

To allow for new models and datasets to be plugged into the browser application, the server code followed an ‘object-orientated approach’. For example, the time-series regression model expects any generic object of type data time-series, without being concerned with the specifics of the dataset; the same model could be run against other time-series variables without changing or adding additional model code. For instance, although the application has been validated using a mortality, air pollution and temperature dataset (see section below), the browser has also been used to assess associations between daily pollen exposure and emergency hospital admissions for asthma.

The client-side code used the Javascript framework, Angularjs, to give a dynamic application allowing for parts of the browser page to change independently as required.[17] This enhances the user experience, leading them through the analysis process and giving as much feedback as possible.

**Theory/calculation**

Browser validation

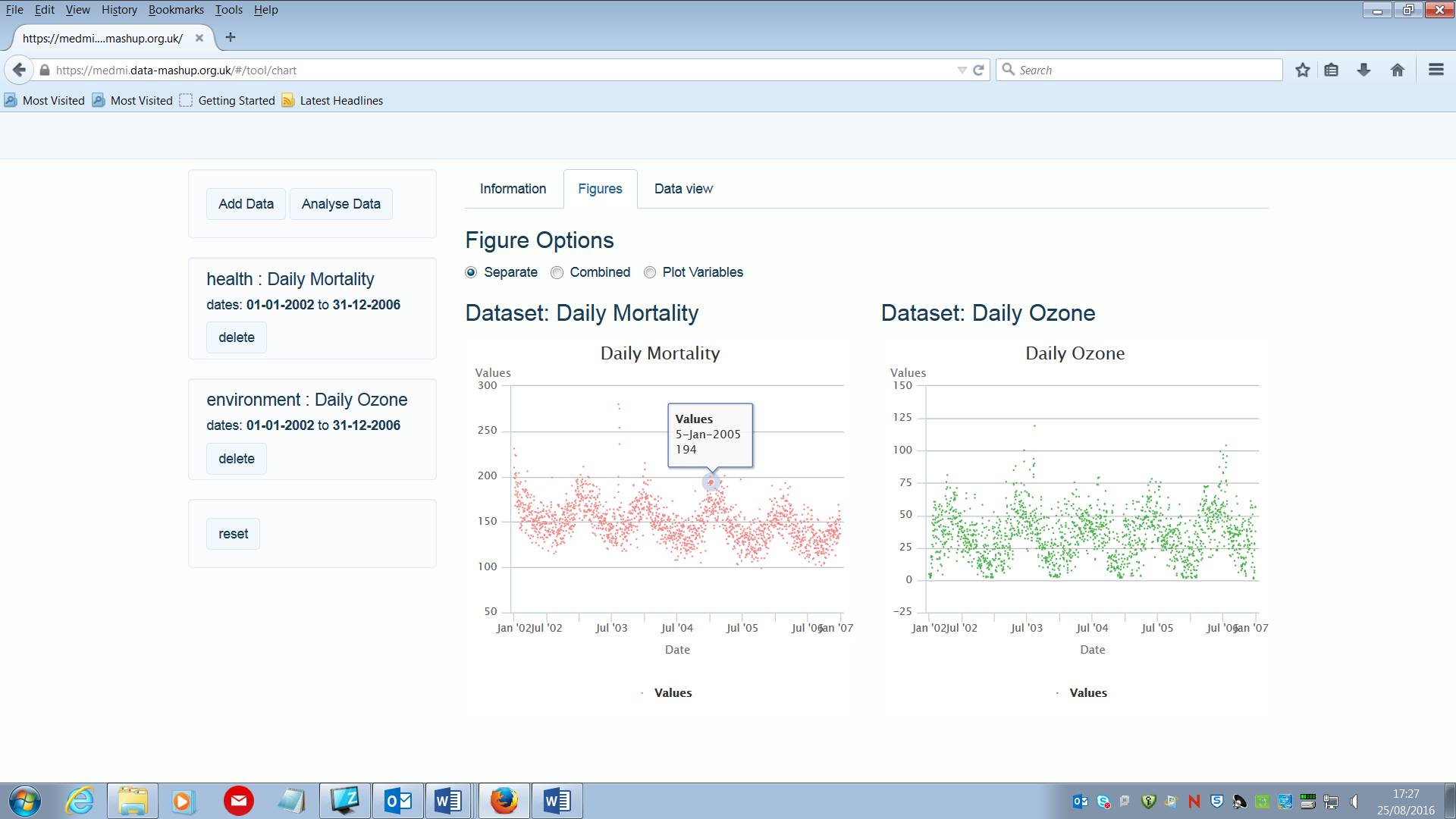
The browser was validated by reproducing previously published results from time-series datasets of daily mortality, air pollution and ambient temperature in London, UK.[18] All data used are in the public domain and downloadable from <http://ije.oxfordjournals.org/content/42/4/1187/> suppl/DC1. In this time-series regression analysis, the outcome consists of the daily number of all-cause deaths occurring in London over a 5-year period (January 1st 2002 to December 31st 2006). The environmental exposure of interest is daily concentrations of the air pollutant tropospheric ozone in London over the same time-period. The time-series regression study allows for the assessment of acute effects only. So, the analysis is appropriate to determine if daily fluctuations in ozone levels are associated with the daily number of deaths, but an alternative study design would be needed to assess possible chronic effects of air pollution exposure, such as a cohort study.[19] We also considered ambient temperature as an exposure to demonstrate additional features of the browser, such as the efficient identification of possible threshold effects.[20] In building up a time-series regression model, other features of the data also require special consideration. These are not discussed here, but are detailed in the original publication.[18]

**Results**

Data visualisation

A key feature of any browser application developed for big data analyses is the data visualisation component. When multiple exposure and outcome datasets can be brought together from disparate sources and linked, visual assessments in either the spatial or temporal domain, or both, may be used initially to indicate potential associations which can then be more rigorously explored in more formal data analyses. Such ‘hypothesis generation tools’ are useful to reveal novel and hitherto undetected pathways through which the environment may impact on human health. In our simple example, both the exposure (daily ozone concentrations) and outcome series (daily mortality) exhibit clear temporal patterns which can be viewed in the browser either in data spreadsheet form or viewed graphically (figure 2).

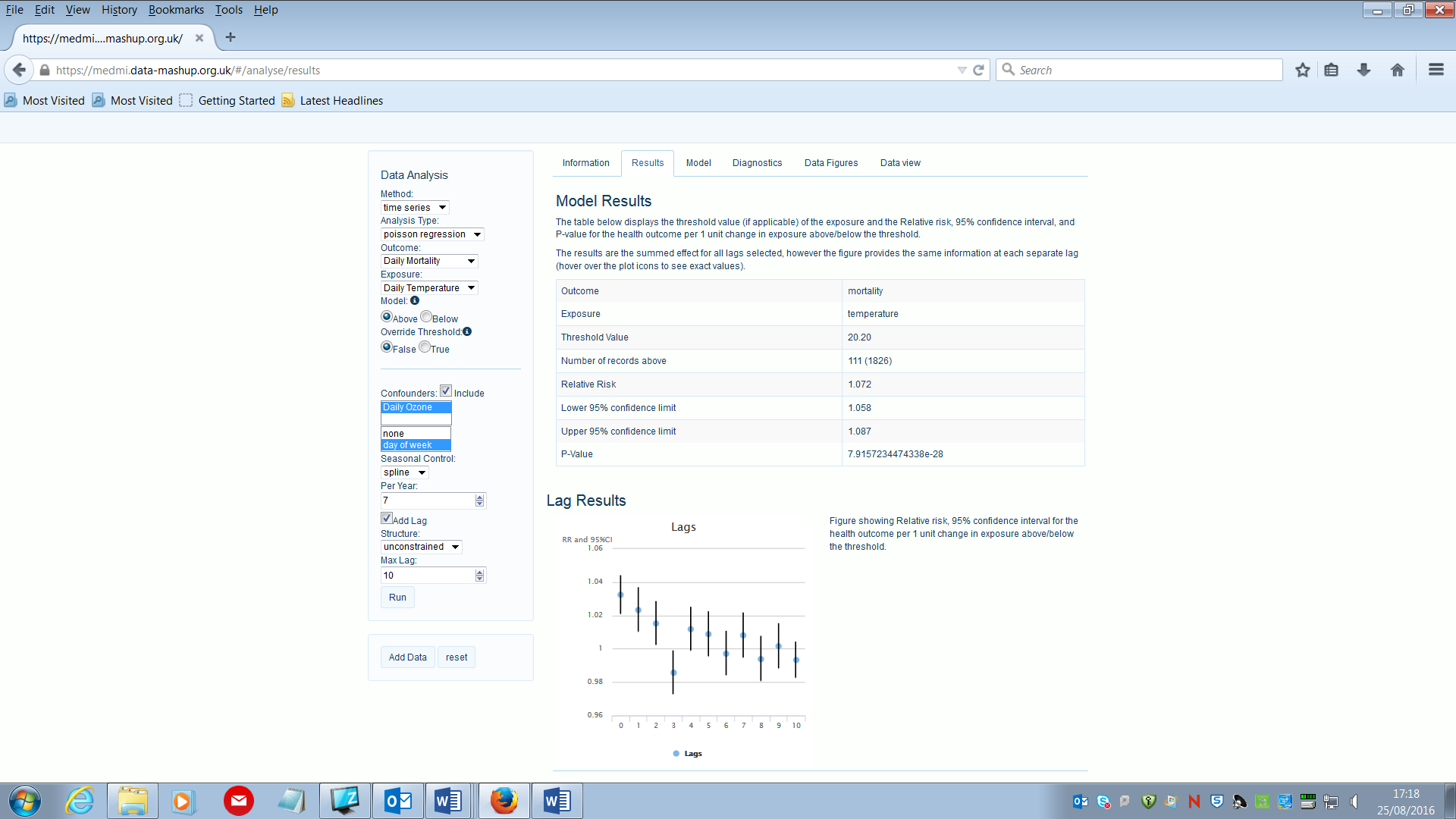
*Figure 2: Data visualisation on browser*



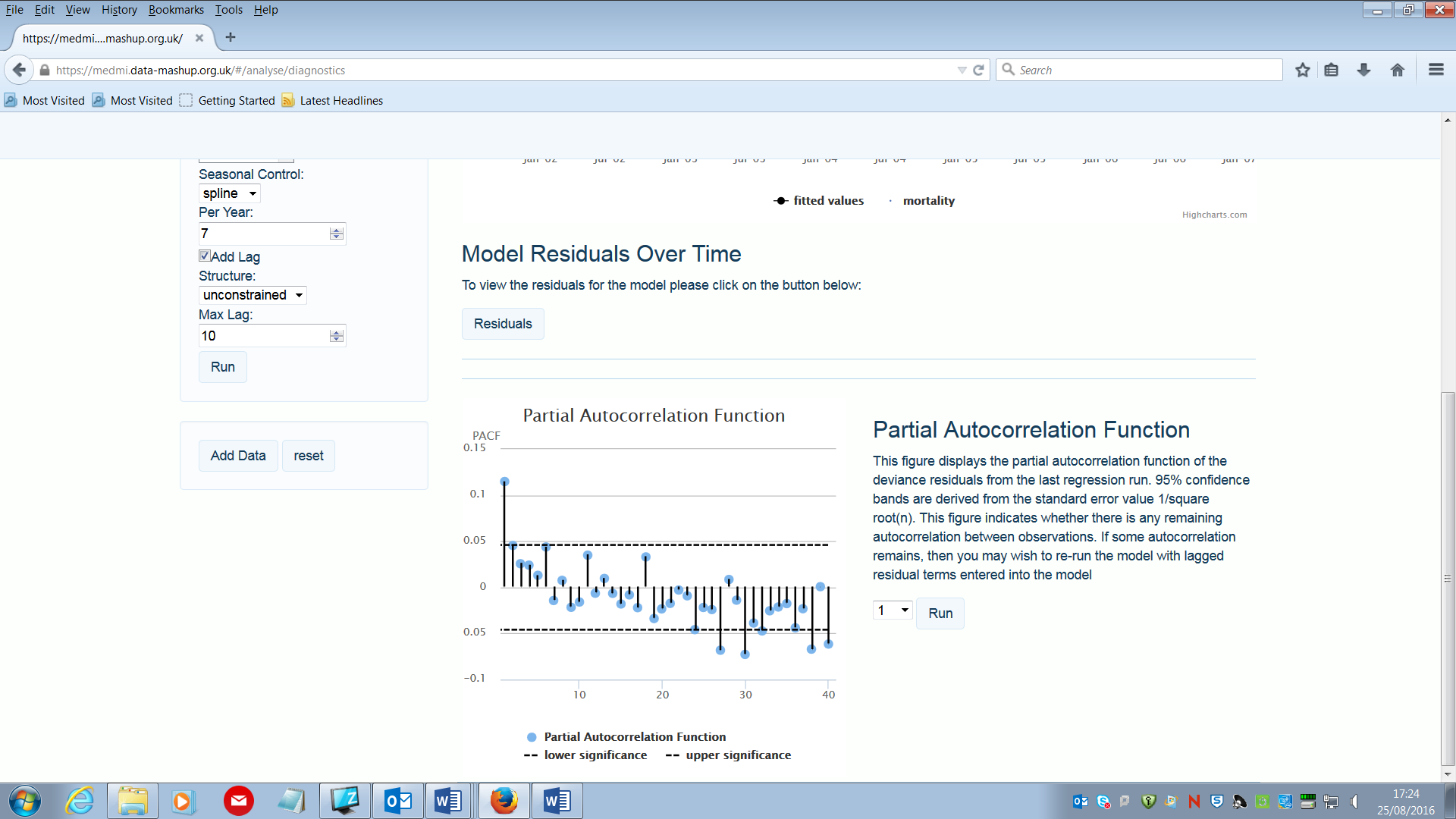
Data analysis

Analytical tools in the application have been designed to be sufficiently flexible to allow both experienced and less-experienced analysts to undertake useful analysis to explore linkages between human health and the environment. An important consideration in the development was the extent to which the interface and analytic options should remain generic enough to allow for future study designs to be incorporated into the browser without the need for complete redevelopment. Also, depending on the study design chosen, further specifications should become available for the experienced analyst. For example, for the time-series regression study, the experienced user may wish to specify the type and degree of seasonal control, the functional form of the relationship and identification of possible thresholds, lagged effects of exposure, etc. For the less experienced analyst, default settings are provided. The output produced can be viewed in both detailed form and viewed as a summary measure (figure 3). Regression model diagnostics are also provided, with explanatory text. For example, figure 4 shows the Partial Autocorrelation Function of the regression model residuals to describe the amount of remaining serial correlation.

*Figure 3: Summary results of time-series regression analysis on browser*



*Figure 4: Partial Autocorrelation Function on browser*



In the above figures, temperature has been used as the main exposure of interest in order to illustrate certain model specifications. For validation purposes, using the browser application we also estimated the adjusted effect of a 10 µg/m3 rise in ozone, resulting in a 0.3% increase in mortality (95% CI -0.1, 0.6), which is the same as the effect size reported in the original publication.[18]

Challenges and priority considerations

In working through the above process of setting up the platform, seeking out databases, and the user requirements and browser development, a number of challenges were encountered which help shaped the eventual form of the browser. Table 1 list these since these are likely to be priority considerations for others interested in developing similar browsers and platforms suitable for data mash-ups.

*Table 1: Challenges and working solutions for MED-MI browser and platform*

|  |  |
| --- | --- |
| Challenges: | Solutions: |
| 1. *The need to improve accessibility of environmental health research and data*  A fundamental justification for developing a web-browser application was to reach as wide an audience as possible. | Even though the type of basic analysis described above can be conducted using statistical computing packages such as STATA or R, the use of a browser circumvents the need to have specialist software installed, and its accessibility via a simple web-link allows analysts to share and repeat results quickly and easily without the need to make available datasets or computer code. |
| 2. *The need to make software open-source*  Allows programming code to be available to all and analytical options can be extended by future users with programming experience. | We used Python for the server aspect of the application since it is a multipurpose language that has web framework libraries, as well as statistical and spatial libraries useful for environmental health research, although other languages could also be used. |
| 3. *The need for flexibility*  Data records should be loosely coupled to allow as much flexibility as possible. | Traditionally in a time-series study, the variables would largely be stored on a single dataset. The browser utilises each variable previously split into an individual ‘dataset’ to maximise flexibility. |
| 4. *Efficiency*  When running a query or analysis that requires extracting big data from a server, users will not expect to wait more than a few seconds, perhaps minutes, before seeing results from their query. | Knowledge of the structure of the datasets can be exploited to develop data-seeking code that returns the required data much quicker than standard Python enumeration methods. Python generator functions can also be developed to output data as soon as it is ready and before the function finishes extracting or processing all of the requested data. |
| 5. User-friendliness  The interface should be easy to follow by novice users. Experienced researchers should also find the features and analysis options sufficiently extensive for their purposes. | Both the input and output functions have been designed to be used without the need for highly developed technical skills. Brief text descriptions are provided of each tool, its usage, outputs and limitations. However, it is envisaged that following further development, additional explanatory information will be provided to aid less-experienced users. Output is displayed using a combination of charts and tables, and explanatory text is provided alongside all results and diagnostics, making the browser also useful as a teaching tool. |
| *6. Licensing restrictions and data confidentiality*  Although it would be desirable to allow users to conduct unspecified research without prescription from the application, there are always likely to be limitations imposed by data providers/owners on the extent to which any potentially sensitive data can be accessed and used. A platform needs to be designed with this in mind. | Development of a browser application allows for substantial interaction of the environment and health datasets by a range of potential users, whilst protecting the data separately on the server. Furthermore, different degrees of access can be granted to different users. |
| *7. Visualisation and interpretation limitations*  Users should be cautioned that exploration of data is primarily a hypothesis-generating exercise which should be more formally assessed using the analytic tools. | The MED-MI website has a section devoted to caveats for interpreting data. |
| *8. Cost-effectiveness*  The browser application should be maintainable with minimal upkeep. | The use of open-source software that is standard in the research community allows programming code to be accessible and extendable by other users. |
| *9. Sustainability*  Future-proofing the application to accommodate multiple databases and multiple uses. | Create with sufficiently flexible features to allow for the further building of the platform. As the system is loosely coupled, new applications can be added with minimal amendment to the system layers. As more data are added, the use of big-data storage solutions (e.g. cloud) could be considered. |

**Discussion**

Human health, climate and other environmental factors are intricately interconnected. Climate change and other environmental changes are having, and will continue to have, a profound effect on these relationships.[21] To support rapid, novel, and rigorous research in this area, there is an increasing need to link and analyse complex meteorological, other environmental, and epidemiological data. The MED-MI platform, including the provision of a web browser, was developed to provide a data and analysis resource to address this need for researchers and other users, whilst ensuring appropriate confidentiality safeguards.

It is stressed that the aim of the browser was to make a complex model available to researchers through a user-friendly interface. We wished to explore the ease with which other model developers could contribute to a web system of this nature and allow for their models to be shown in a similar user-friendly way. If new models are written to use python data objects as designed here and provide results as JSON, then they could be incorporated within the web interface. Similarly for new data, if they are added with the same data access calls then they could be used within the model without a code rewrite.

In the early stages of development, even with the simple example highlighted in this paper, there have been many challenges, both anticipated and unanticipated, to make the data and browser as useful as possible. These broadly relate to issues of data formats and structures, and data visualisation and analytics – illustrations of which are discussed below.

*Data formats and structures*

A key consideration in development was that any available datasets could be linked, in whichever combination the user selects. This posed significant challenges since the large variation in the possible data formats and structures of the diverse environment and health datasets meant that different inputs would be needed depending on the type of data selected by the user. For example, the platform includes 5km resolution gridded climate data from Met Office surface observations, in raster image format.[22] However, health datasets are generally available within administrative boundaries such as postcode, Lower Super Output Area, and Government region. In the first instance linking the two was achieved by offline processing, using spatial queries to extract information from the image data and attach to the administrative boundary. Once in the same vector time series format, data could be loaded onto the applications database to be used as an additional input. In the future however, this analysis between raster and vector formats would ideally be achieved as a dynamic process through the application, allowing for the user to select their administrative boundary of choice to analyse against the raster format. This could also allow for real-time processing. To achieve this, different data stores would need to be used, moving away from the classic relation database and considering a combination of data stores more commonly used for big data solutions. The idea behind the application is that these necessary changes would only need to occur at the data store layer, and by developing the right connectors, the modelling and presentation code would continue to work.

In addition, as the available data grow, the approach of storing data locally will no longer be feasible and making use of data services could be the adopted approach. This would allow for real-time analysis and avoid the problems and inconsistencies of duplicating data. This approach would, however, rely on the development of data services being provided by data owners, as well as standards being developed for the format of the data being delivered. When analysing data from a number of sources, it will also be a challenge to decide and develop where the model is run – is the model passed to the data location or vice versa?

*Data visualisation and analytics*

Latest web technologies provide a feature-rich browser application, resulting in as much functionality and user interaction as possible in data visualisation and analytics. For example, the combination of Javascript and JSON in the browser allows options such as hovering over data points in a chart to read exact values. This would not be possible if the chart had been created on the server and sent to the browser as an image. As much of the data processing as possible was moved to the browser in order to visualise the data in multiple ways, such as with charts and tables, without having to re-query the original data. This approach is more efficient for resource-heavy queries since the number of requests to the server are reduced. Nevertheless, speed is also a challenge to create original time-series from raw data on the database server. For instance, with the introduction of the gridded climate data, the extraction of maximum temperature measurements at locations in London for a specified time range from the over 9 billion observations and statistics available on the server requires special consideration.

The team wanted the browser application to be as flexible and as widely useable as possible, allowing for analyses to be carried out on other environment and health datasets that may become available in future. The requirement for the application to be extendable was to allow for future visualisation and analysis options and data to be added without having to rewrite aspects of the browser application. It was also recognised that this approach would benefit the end user performing their research. By having access to multiple models and datasets through a user-friendly application, the user can run a variety of analyses quickly and efficiently. Allowing both environment and health variables to be disassociated from individual datasets promotes a wider range of hypotheses to be potentially explored. The data visualisation features of the browser may reveal temporal and/or spatial associations between previously unlinked environmental exposures and health outcomes. Such flexibility will become increasingly important in future as the number of environment and health databases made available via the server, and the inherent complexities of the analyses, grows. The key consideration is that any available datasets could be linked, in whichever combination the user selects. However, it is vitally important that there is transparency in the purposes for which the browser tools should be used, as well as any limitations. For example, the data visualiser may indicate patterns between variables, however the user should be cautioned that these associations may not necessarily be causal and that they should then be iteratively and more robustly assessed with further exploration using the appropriate analytical tools available.[23]

*Other considerations*

As more datasets become available to be linked on the server, there is potential for increased tension between the desire by analysts to use ever finer spatial information and the need to preserve individual privacy.[24] Such concerns are commonplace in environmental health research,[25] and most providers of health data have confidentiality-preserving restrictions which may include processing of data to mask precise information on location or to provide aggregate information only. Safeguards on the platform are also necessary to prevent potential re-identification of individuals once linked to environmental information that may be more precisely geocoded.

Initially, we envisage the data that are on the platform will be made available to approved users (i.e. researchers), with subsequent expansion of the platform and website portal to make at least some aspects accessible to a wider audience. Currently, researchers request approval by completing a brief questionnaire on the MED-MI website that, among other things, asks about the intended purpose of the data requested. In making such a platform available to interested parties, there naturally have to be restrictions and confidentiality constraints consistent with the data being accessed (e.g. individual confidentiality and commercial protections). For example, confidential data (i.e. individual-identifying) could be made available only via query/batched analyses to approved users; anonymised and/or perturbed data would be downloadable to individual PCs or via the browser for analysis. There would likely be layers of user access depending on their requirements to analyse the data (e.g. researchers) as opposed to querying, visualising, and mapping the data (e.g. the press, NGOs, policy makers, and individuals). In this initial stage, the security is achieved by having two versions of the application pointing to different data layers. One of these is for the public data, and one with restricted access for more sensitive data. When extending this kind of application in the future, further security will need to be considered. For example a user may have the right to use a certain dataset, but only with other specific datasets. If MED-MI is to give access to multiple datasets, then both the user and the data will need to be considered when providing licensed access. Our experience suggests that to achieve such a platform in the future, with full functionality of analysing varied datasets, security is a major challenge. Crucially such projects must involve teams with expertise from both technical and research fields, as well as a willingness of organisations to share their data.

In order to make the data and research tools as accessible and intuitive as possible for the additional users described above, our initial efforts have focussed on the development of web browser applications. As described in this paper, currently only one type of analysis is possible on the browser, and one that is only amenable to assess more direct relationships, such as those between daily ambient temperature and mortality, or between daily pollen concentrations and emergency asthma admissions.[26] We have also developed a separate interactive visualisation browser to explore reportable infectious disease cases (located to the diagnosing laboratory) linked with a range of environmental variables in England. In addition, we are developing another browser that will allow researchers to create bespoke linked datasets using dropdown menus from the MED-MI data.

In time, it is anticipated that, as more datasets and study designs are incorporated, the browser will enable the more complex current and future climate change processes to be more robustly characterised.[2] Such information will ultimately lead to a number of useful applications, including: better identification of ‘hotspots’ of vulnerable populations for targeted protection; the provision of relevant information to healthcare practitioners and public health planners to help improve services for locations and at-risk populations; the initiation and evaluation of intervention measures; and data access and dissemination as part of outreach and engagement activities with the research community, policymakers and civil society.[8] There is also the potential to introduce real-time data into the browser for public health surveillance of environmental exposures,[27] as well as the possibility of using Met Office weather forecasts to anticipate impacts on health and health-care utilisation.[28] This opens up the possibility of incorporating information from social media, with different sources potentially offering differing perspectives on an emerging environmental health problem. For example, analysis of a drinking water ban issued in Toledo, Ohio in 2014 observed that Twitter reflected a real-time public response to the issued notice with awareness that harmful algal blooms (HAB) were the underlying environmental issue, and concurrent activity on Google Trends illustrated that the Toledo incident raised general awareness of the health hazards associated with the bloom.[29]

**Conclusions**

In conclusion, creating dedicated data and analysis resources such as the MED-MI platform described here will become increasingly important in advancing environmental, epidemiologic, clinical, and commercial collaborative applications in the field of environment and human health. A vital next step will be for environmental, medical and public health researchers, as well as other potential stakeholders, to explore the tool and provide practical feedback on its usability and usefulness. Once demonstrated, dissemination of such shared resources in the UK and beyond will ultimately lead to improved understanding of the interconnections between the environment and human health and wellbeing.

**References**

1. Morris GP, Reis S, Beck S, Fleming LE, Adger WN, Benton TG, et al. Climate change and health in the UK: scoping and communicating longer term “distal” dimensions. *Env Health*; submitted.
2. McMichael AJ. Impediments to comprehensive research on climate change and health. *Int J Environ Res Public Health* 2013;10:6096-105
3. Reis S, Morris G, Fleming LE, Beck S, Taylor T, White M, et al. Integrating health and environmental impact analysis. *Public Health* 2013;pii: S0033-3506(13)00242-4
4. Lloyd SJ, Kovats RS, Chalabi Z. [Climate change, crop yields, and undernutrition: development of a model to quantify the impact of climate scenarios on child undernutrition](http://www.ncbi.nlm.nih.gov/pubmed/21844000). *Environ Health Perspect* 2011;119:1817-23
5. Semenza JC, Menne B. Climate change and infectious diseases in Europe. *Lancet Infect Dis* 2009;9:365-75.
6. U.S. Federal Government, 2014: U.S. Climate Resilience Toolkit. [Online] <http://toolkit.climate.gov>. Last accessed [30/05/16]
7. The World Bank Group. Climate Change Knowledge Portal. <http://sdwebx.worldbank.org/climateportal>. Last accessed [30/05/16]
8. Fleming LE, Haines A, Golding B, Kessel A, Cichowska A, Sabel CE, et al. Data mash-ups: potential contribution to decision support on climate change and health. *Int J Environ Res Public Health* 2014;11:1725-46
9. Little A, 2013. <https://uxmag.com/articles/storyboarding-in-the-software-design-process>. Last accessed [30/05/16]
10. Microsoft Dynamics CRM Team Blog, 2006. <http://blogs.msdn.com/b/crm/archive/2006/11/02/storyboarding-in-software-development-with-particular-reference-to-microsoft-crm.aspx>. Last accessed [30/05/16]
11. Katsouyanni K, Schwartz J, Spix C, Touloumi G, Zmirou D, Zanobetti A, et al. [Short term effects of air pollution on health: a European approach using epidemiologic time series data: the APHEA protocol](http://www.ncbi.nlm.nih.gov/pubmed/8758218). *J Epidemiol Community Health* 1996;50 Suppl1:s12-8
12. Samarasundera E, Hansell A, Leibovici D, Horwell CJ, Anand S, Oppenheimer C. Geological hazards: From early warning systems to public health toolkits. *Health & Place* 2014;30:116-9
13. Python Software Foundation. Python Language Reference, version 2.7. Available at <http://www.python.org>
14. [ECMA-404 The JSON Data Interchange Standard](http://www.ecma-international.org/publications/files/ECMA-ST/ECMA-404.pdf). Available at <http://www.json.org/>.
15. Waters K, 2007. [http://www.allaboutagile.com/what-is-agile-10-key-principles](http://www.allaboutagile.com/what-is-agile-10-key-principles/). Last accessed [30/05/16]
16. Dingsoyr T, Nerur S, Balijepally V, Brede Moe N. A decade of agile methodologies: towards explaining agile software development. *J Systems Software* 2012;85:1213-21.
17. <https://angularjs.org>. Last accessed [30/05/16]
18. Bhaskaran K, Gasparrini A, Hajat S, Smeeth L, Armstrong B. Time series regression studies in environmental epidemiology. *Int J Epidemiol* 2013;42:1187-95
19. Dockery DW, Pope CA 3rd, Xu X, Spengler JD, Ware JH, Fay ME, et al. An association between air pollution and mortality in six US cities. *N Engl J Med* 1993;329:1753-9
20. Muggeo VM. Estimating regression models with unknown break-points. *Stat Med* 2003:22:3055-71
21. Haines A, Kovats RS, Campbell-Lendrum D, Corvalan C. Climate change and human health: impacts, vulnerability, and mitigation. *Lancet* 2006;367:2101-9
22. <http://www.metoffice.gov.uk/climatechange/science/monitoring/ukcp09>. Last accessed [30/05/16]
23. Faghmous JH, Kumar V. A big data guide to understanding climate change: the case for theory-guided data science. *Big Data* 2014;2:155-63
24. Kamel Boulos MN, Curtis AJ, Abdelmalik P. Musings on privacy issues in health research involving disaggregate geographic data about individuals. *International Journal of Health Geographics* 2009:8:46
25. Smolders R, Casteleyn L, Joas R, Schoeters G. Human biomonitoring and the INSPIRE directive: spatial data as link for environment and health research. *J Toxicol Environ Health B Crit Rev* 2008;11:646-59
26. Dales RE, Cakmak S, Judek S, Dann T, Coates F, Brook JR, et al. Influence of outdoor aeroallergens on hospitalization for asthma in Canada. *J Allergy Clin Immunol* 2004;113:303-6
27. Elliot AJ, Bone A, Morbey R, Hughes HE, Harcourt S, Smith S, et al. Using real-time syndromic surveillance to assess the health impact of the 2013 heatwave in England. *Environ Res* 2014;135:31-6
28. Marno P, Chalder M, Laing-Morton T, Levy M, Sachon P, Halpin D. Can a health forecasting service offer COPD patients a novel way to manage their condition? *J Health Serv Res Policy* 2010;15:150-5
29. Cha Y, Stow CA. Mining web-based data to assess public response to environmental events. *Environmental Pollution* 2015;198:97-9