

Small or medium-scale focused research project (STREP)

ICT SME-DCA Call 2013

FP7-ICT-2013-SME-DCA

**Data Publishing through the Cloud:
A Data- and Platform-as-a-Service Approach to Efficient
Open Data Publication and Consumption**

DaPaaS



Deliverable D5.4

Use cases collection

Date:	30 th October 2015
Author(s):	Jamie Fawcett (ODI), Jack Hardinges (ODI), Amanda Smith (ODI), Leonard Mack (ODI)
Dissemination level:	PU
WP:	WP5
Version:	1.0

Document metadata

Quality assurors and contributors

Quality assessor(s)	Bill Roberts (SWIRRL), Dumitru Roman (SINTEF)
Contributor(s)	DaPaaS Consortium, External experts

Version history

Version	Date	Description
0.1	28-09-2015	Initial draft sent to reviewers
0.2	01-10-2015	Review comments from DR
0.3	19-10-2015	Review comments from BR
0.4	25-10-2015	Review comments implemented
1.0	29-10-2015	Adjustments and finalization.

Executive Summary

This report explores the real world potential of the DataGraft platform and tools developed during the DaPaaS project. It does this by identifying a set of sector-specific challenges in which data, particularly open and linked data, can play a significant role in solving. The challenges considered are: Tackling urban air pollution, effectively managing water resources, improving cities' resilience to extreme weather events and bringing new drugs to a global market. For each of these data can play, and is playing, a key part in helping to solve the challenges. In particular, open and linked data strategies are demonstrating their value. The potential of the DataGraft platform is discussed in each of the cases in its ability to maximise this value and thus helping to tackle the challenges. These use cases are unique; the 'grand challenges' that we tackle have not been discussed in this level of detail, and with this perspective, before. We outline why these use cases were chosen in the introduction of this report.

TABLE OF CONTENTS

EXECUTIVE SUMMARY	3
1 INTRODUCTION	5
2 TACKLING URBAN AIR POLLUTION	6
2.1 WHAT IS THE CHALLENGE?	6
2.2 HOW CAN WE TACKLE THIS PROBLEM USING DATA?	6
2.2.1 Understanding and targeting air pollution	7
2.2.2 Informing citizens about air pollution	8
2.2.3 Changing behaviour to avoid and reduce air pollution	8
2.3 WHAT (OPEN) DATA DO WE NEED?	9
2.4 HOW CAN THE DATAGRAFT PLATFORM HELP?	13
2.5 CONCLUSIONS	15
3 EFFECTIVELY MANAGING WATER RESOURCES.....	16
3.1 WHAT'S THE CHALLENGE?	16
3.1.1 Taking an integrated systems approach.....	16
3.2 HOW CAN WE TACKLE THIS PROBLEM USING DATA?	18
3.2.1 Data-driven modelling for management and planning of water resources	18
3.2.2 Controlling and reducing demand for water	20
3.2.3 Understanding and tackling pollution	21
3.3 WHAT (OPEN) DATA DO WE NEED?	22
3.4 HOW CAN THE DATAGRAFT PLATFORM HELP?	25
3.5 CONCLUSIONS	26
4 IMPROVING CITIES' RESILIENCE TO EXTREME WEATHER EVENTS	27
4.1 WHAT IS THE CHALLENGE?	27
4.2 HOW CAN WE ADDRESS THE CHALLENGE USING DATA?	27
4.2.1 Better understanding the risks	27
4.2.2 Improving response, recovery and reconstruction	30
4.3 HOW CAN THE DATAGRAFT PLATFORM HELP?	32
4.3.1 Making it easier to publish data with limited resources	32
4.3.2 Enabling the publishing of interoperable linked data	33
5 BRINGING NEW DRUGS TO A GLOBAL MARKET.....	34
5.1 WHAT IS THE CHALLENGE?.....	34
5.2 HOW CAN WE ADDRESS THE CHALLENGE USING DATA?	35
5.2.1 Clinical trials data.....	35
5.2.2 Real world data	37
5.3 HOW CAN THE DATAGRAFT PLATFORM HELP?	40
5.3.1 Publishing clinical trial data effectively	40
5.3.2 Exploiting the value of real world data	41
6 CONCLUSIONS.....	42

1 Introduction

This report explores the real world potential of the DataGraft platform and tools developed during the DaPaaS project. It does this by identifying a set of sector-specific challenges in which data, particularly open and linked data, can play a significant role in solving. To evaluate the utility of DataGraft, it first interrogates the role data is playing with particular focus on open and linked data. Following this overview of current and planned approaches, each use case explores the issues faced in solving the challenges, with particular attention paid to the amount and quality of data publishing. They then propose how the DataGraft platform could help resolve some of these issues and thus improve the approach to each of the challenges.

An initial set of 14 challenges from a range of sectors were identified from urban, environmental and global challenge literature. Each of these challenges was then assessed on its suitability in demonstrating the value of DataGraft. This involved scoring each challenge in 5 categories on a scale of 1 to 5 based on initial research and existing knowledge of sectors. The categories considered were:

- Number of publishers
- Number of relevant datasets
- Existing barriers to publishing
- Importance of non-tabular data
- Importance of data mashups

In addition to the overall score of each challenge, the nature and scope of each in terms of environmental, economic and social impact (triple-bottom line) was factored in. Those which had wider influence and greater impact were considered more valuable in demonstrating the utility of DataGraft to relevant actors. On top of this, it was decided that the use cases should be clustered around a few key sectors in order to provide consistency to the overall report. As such the report focuses on challenges around health, environment and smart cities.

The challenges considered in this report are:

1. Tackling urban air pollution

Challenge: Urban air pollution is one of the biggest environmental, economic and public health challenges faced by cities in the 21st century.

Domain(s): Smart cities, health, environment

2. Effectively managing water resources

Challenge: Water is a vital resource for both nature and society which is under pressure from increasing demand, climate change and natural disaster.

Domain(s): Environment

3. Improving cities' resilience to extreme weather events

Challenge: Extreme weather events cause devastating human and economic damage around the world each year.

Domain(s): Smart cities, environment

4. Bringing new drugs to a global market

Challenge: Bringing new medicines to market is a long and expensive process which is stifling life-saving innovation.

Domain(s): Health

2 Tackling urban air pollution

2.1 What is the challenge?

Urban air quality is one of the biggest environmental, economic and public health challenges faced by cities in the 21st century. Outdoor air pollution is one of the top 10 risks to global health¹ and the prime environmental cause of death in the European Union (EU).² In Europe, poor air quality causes approximately 600,000 premature deaths a year³ with an economic cost of \$1.6 trillion⁴ according to the World Health Organisation (WHO). This does not take into account the environmental damage,⁵ developmental effects on children⁶ and other impacts poor air quality has on society.

Air pollution is not a problem limited to any one section of urban society. The WHO estimates that 83% of the urban population of Europe⁷ is exposed to above guideline levels of particulate matter⁸ (PM). This is especially concerning as the same report finds that despite setting the guidelines for exposure “*there is no evidence of a safe level of exposure*” to PM.⁹ Tackling urban air pollution is a huge challenge for urban society and one that demands an immediate and concerted effort to solve.

2.2 How can we tackle this problem using data?

Owing to years of scientific research we now have a relatively good understanding of the direct causes and effects of poor air quality. Using this information we can now produce complex models which estimate the levels and effects of air pollution on society as a whole. This research has also helped us develop a wide range of innovative technology and policy solutions. The WHO estimates that “*up to 80% of particulate air pollution in Europe can be reduced with currently available technologies*”.¹⁰ Such solutions must however be carefully targeted and properly implemented to have the desired effect.

To recognise this fully requires “*concerted action by public authorities, industry and individuals at national, regional and even international levels*”.¹¹ In order to establish this level of cooperation all actors must be understand the causes, effects and solutions to air pollution that relate to them directly. Only then will they be able to adjust their behaviour to recognise the danger of poor air quality and play a role in tackling it.

To tackle urban air pollution effectively we need to solve some of these issues. Here we examine three key areas where open, shared and closed data is being used to do just this.

¹ http://www.who.int/healthinfo/global_burden_disease/GlobalHealthRisks_report_full.pdf (last accessed on 2015-10-25)

² <http://ec.europa.eu/environment/life/publications/lifepublications/lifefocus/documents/airquality.pdf> (last accessed on 2015-10-25)

³ [http://www.euro.who.int/en/media-centre/sections/press-releases/2015/air-pollution-costs-european-economies-us\\$-1.6-trillion-a-year-in-diseases-and-deaths.-new-who-study-says](http://www.euro.who.int/en/media-centre/sections/press-releases/2015/air-pollution-costs-european-economies-us$-1.6-trillion-a-year-in-diseases-and-deaths.-new-who-study-says) (last accessed on 2015-10-25)

⁴ http://www.euro.who.int/_data/assets/pdf_file/0004/276772/Economic-cost-health-impact-air-pollution-en.pdf (last accessed on 2015-10-25)

⁵ <http://airuse.eu/wp-content/uploads/2012/12/Air-qualityEUROPE-2014.pdf> (last accessed on 2015-10-25)

⁶ http://www.euro.who.int/_data/assets/pdf_file/0010/74728/E86575.pdf (last accessed on 2015-10-25)

⁷ http://www.euro.who.int/_data/assets/pdf_file/0006/189051/Health-effects-of-particulate-matter-final-Eng.pdf (last accessed on 2015-10-25)

⁸ <http://www.eea.europa.eu/themes/air/air-quality/resources/glossary/particulate-matter> (last accessed on 2015-10-25)

⁹ http://www.euro.who.int/_data/assets/pdf_file/0006/189051/Health-effects-of-particulate-matter-final-Eng.pdf (last accessed on 2015-10-25)

¹⁰ http://www.euro.who.int/_data/assets/pdf_file/0006/189051/Health-effects-of-particulate-matter-final-Eng.pdf (last accessed on 2015-10-25)

¹¹ http://www.euro.who.int/_data/assets/pdf_file/0006/189051/Health-effects-of-particulate-matter-final-Eng.pdf (last accessed on 2015-10-25)

2.2.1 Understanding and targeting air pollution

Understanding the level and nature of air pollution in a specific area, city or region is key to tackling poor air quality. While national and international targets, regulation and legislation can be effective in tackling overall emissions levels they can fail to capture and tackle more localised issues.¹² The level of particular pollutants, nature of sources and the effects vary widely between localities. This means that different regions, cities and even streets will have different concerns which require more targeted solutions. Therefore, understanding air pollution in their area allows public authorities and others to choose the most appropriate solutions to tackle the issues relevant to them.

The majority of this information currently comes from complex environmental models. These models are derived from environmental research and applied to real-world context using a wide variety of observed data sources. One example is the ADMS Urban model,¹³ produced by Cambridge Environmental Research Consultants (CERC).¹⁴ The model allows authorities to visualise and interpret air pollution levels in urban areas. This can allow them to target specific technical and policy interventions, such as implementing low emissions zones¹⁵ or bicycle hire schemes.¹⁶ More than this, it can be used to test and compare the effects that these interventions could have on air quality.

These models require a large amount of data, not just observed air quality data but also emissions sources and profiles, traffic density and flows, and weather patterns. The more data that is ingested into the model, the more accurate estimations and forecasts of air pollution can be made. Another such modelling tool for policy analysis is the GAINS model¹⁷ produced by IIASA.¹⁸ The model is used to inform a wide range of policy decisions in Europe.¹⁹ Integration of air quality data from monitoring stations across Europe improved the accuracy of the model²⁰ and therefore its effectiveness in informing policy.

Other data sources can also be used to improve the usefulness of models in targeting interventions. One novel example, from the US, is the use of GPS tracked inhalers by a company called Propeller Health.²¹ They use devices attached to inhalers to register where asthma sufferers use their inhalers and map this against known pollution sources. This can help asthma sufferers track their own usage but can also be used to target pollution hotspots likely to trigger health problems. Although, it is worth noting that there might be concerns around the sharing of potentially identifiable data with researchers and authorities,²² the benefits to public health and the participants themselves of targeted interventions must be considered.

In all these cases the advantages of more data are greatly apparent, be it data on emissions and conditions or air quality data itself. Data is a fundamental part of understanding air pollution and using this understanding to target technical and policy interventions.

¹² http://ec.europa.eu/environment/archives/cafe/activities/pdf/task_3_3.pdf (last accessed on 2015-10-25)

¹³ <http://www.cerc.co.uk/environmental-software/ADMS-Urban-model.html> (last accessed on 2015-10-25)

¹⁴ <http://www.cerc.co.uk/index.php> (last accessed on 2015-10-25)

¹⁵ http://www.londonair.org.uk/london/asp/LAQNSeminar/pdf/september2010/Berlin_LEZ_impacts_analysis.pdf (last accessed on 2015-10-25)

¹⁶ <http://www.bmj.com/content/343/bmj.d4521> (last accessed on 2015-10-25)

¹⁷ <http://www.iiasa.ac.at/web/home/research/modelsData/GAINS/GAINS.en.html> (last accessed on 2015-10-25)

¹⁸ <http://www.iiasa.ac.at/> (last accessed on 2015-10-25)

¹⁹ <http://www.iiasa.ac.at/web/home/research/researchPrograms/Europe.en.html> (last accessed on 2015-10-25)

²⁰ <http://www.iiasa.ac.at/web/home/about/news/150219-EU-air.html> (last accessed on 2015-10-25)

²¹ <http://propellerhealth.com/> (last accessed on 2015-10-25)

²² <http://www.popsoci.com/every-breath-you-take-theyll-be-tracking-you> (last accessed on 2015-10-25)

2.2.2 Informing citizens about air pollution

Top-down targeted solutions cannot alone solve this problem. Tackling air pollution demands concerted action and political will. This requires that citizens be engaged and informed when it comes to the negative effects of poor air quality. When citizens are aware of the direct effect air pollution is having on them, they are far more likely to pressure political actors for change.

A clear example of this from the US is the State of the Air campaign²³ run by the American Lung Association (ALA).²⁴ Citizens are provided simple high-level analysis of air quality in their county or city. They can use this information to lobby locally elected representatives and can join in the ALA's own national campaigns. This high level simple analysis is based on publicly available monitoring data provided by the US Environmental Protection Agency (EPA).²⁵

Empowering citizens to highlight the need for air pollution control measures can be extremely beneficial in generating action. However, ensuring this action is effective in tackling air pollution can be difficult. This can however be achieved by comparing actual air pollution against regulatory targets. One example of a UK project doing this is the London Air Quality Network (LAQN).²⁶ Using air quality sensor data the LAQN monitors the effectiveness of air pollution solutions in specific locations around London.²⁷

This project and others like it²⁸ can be used drive concerted action by putting useful information in the hands of citizens. To be useful, this information must be accessible and independently verifiable. Empowering civil society by providing them with the data gives them the opportunity to ensure this is the case.

2.2.3 Changing behaviour to avoid and reduce air pollution

Providing citizens with the information they need to drive action on air pollution and monitor its effects can help drive change. Empowering citizens to change their own relationship with air pollution can have an enormous impact on people's everyday lives. Access to real time information on pollutant levels gives citizens the power to avoid high levels of air pollution. When presented in an easily accessible and understandable fashion, it can be particularly helpful for high risk individuals such as those with asthma.

One project aimed at helping people recognise the real-time levels of air pollution is Plume Labs.²⁹ The service provides detailed live pollution reports for major cities around the world on the web³⁰ and as a smartphone application³¹. The simple, engaging reports provide city level pollution information in the context of historic, worldwide and recommended pollution levels. It also provides guidance on whether citizens should engage in certain activities, which expose them to air pollution risks such as outdoor exercise and eating outside.

Although these city wide services can be useful, they do not provide people with exact levels in their local areas. Air pollution levels can vary drastically in urban areas. For some who are especially vul-

²³ <http://www.stateoftheair.org/> (last accessed on 2015-10-25)

²⁴ <http://www.lung.org/> (last accessed on 2015-10-25)

²⁵ <http://www.epa.gov/> (last accessed on 2015-10-25)

²⁶ <http://www.londonair.org.uk/LondonAir/Default.aspx> (last accessed on 2015-10-25)

²⁷ <http://www.londonair.org.uk/london/asp/publicstats.asp?statyear=2014> (last accessed on 2015-10-25)

²⁸ <http://www.cleanerairforlondon.org.uk/londons-air/air-quality-data/data-and-maps/annual-objectives> (last accessed on 2015-10-25)

²⁹ <https://www.plumelabs.com/#air-report> (last accessed on 2015-10-25)

³⁰ <https://air.plumelabs.com/London> (last accessed on 2015-10-25)

³¹ <https://itunes.apple.com/us/app/plume-air-report-beat-pollution/id950289243?mt=8> (last accessed on 2015-10-25)

nerable knowing local levels is very important in planning their activities, to mitigate health risks. Services such as Skopje air quality³² can present this localised information in a simple and informative fashion. Providing the current readings, the past 9 hours and the 'safe' levels for specific areas within the city makes it very accessible and user friendly. Putting this granular information on a map as BreezoMeter aims to,³³ can provide even more useful information. These can be particularly useful when linked directly to user locations through smartphone GPS location.³⁴

Highly granular monitoring systems can be made even more useful when combined with forecast modelling. This is the approach taken by the airText system for Greater London produced by CERC.³⁵ The service aims to make the information as accessible as possible by enabling users to sign up for alerts via text, voicemail, twitter or email.³⁶ This is particularly useful for those who are especially vulnerable to air pollution as it allows them to easily and passively remain informed about daily exposure risks.

These systems give information to citizens which they can use to actively avoid air pollution. In daily practice, this means citizens must make the decision on how best to avoid the negative effects themselves. Several recent projects, such as Air Sensa,³⁷ are planning to help citizens to make these decisions more easily. One way they are doing this is by producing smart route planning services which account for air pollution levels.³⁸ Such a service allows citizens to opt to avoid areas of high pollution when walking or cycling without having to check the levels themselves. Another recent EC-funded project, CITI-SENSE, aims to empower citizens by involving them in the collection, dissemination and use of air quality data.³⁹ This holistic approach, underpinned by strong technical solutions, can help meet all the problems raised.

These projects not only put the appropriate information into the hands of society but give citizens the power to act on it. The proliferation of these kinds of tools requires a great deal of access to data. Access allows innovative citizens and businesses to develop useful tools. Citizens are then able to help mitigate the effects of air pollution has on them and likewise the contribution they make to air pollution.

2.3 What (open) data do we need?

It is clear that tackling urban air pollution requires data-driven solutions. From targeting policy responses to changing individual behaviour, data underpins effective action to tackle poor air quality and its effects. The key requirement is air pollution measurement data from monitoring stations and sensors. But there is a wide range of other data sources and types which can be used to build air pollution models and understand the effects of poor air quality. The availability of both types of data is vital to the success of any solutions tackling air pollution. As is its usability, in terms of data quality and authority.

There is a lot of existing data that can be used to combat air pollution but is not fully utilised because it is not available to all those who want to use it, in the way they need it, and with reliable access. To explore how we could improve the situation we will examine what relevant data is made available as

³² <http://gorjan.rocks/clients/airquality/index.html#draw?location=Centar&type=PM10> (last accessed on 2015-10-25)

³³ <http://breezometer.com/> (last accessed on 2015-10-25)

³⁴ <https://itunes.apple.com/us/app/breezometer-air-quality/id989623380?mt=8> (last accessed on 2015-10-25)

³⁵ <http://www.airtext.info/> (last accessed on 2015-10-25)

³⁶ <https://www.airtext.info/signup> (last accessed on 2015-10-25)

³⁷ <http://www.airsensa.org/> (last accessed on 2015-10-25)

³⁸ <http://diversify-project.eu/papers/nallur15.pdf> (last accessed on 2015-10-25)

³⁹ <http://www.citi-sense.eu/> (last accessed on 2015-10-25)

open data, data which anyone can access, use and share. We will then explore how opening more, high quality, data might help to tackle the problem of air pollution.

There is a great deal of important open data available that can be used to generate models, help develop different solutions and understand the impacts of poor air quality. Table 1, below, provides a small sub-section of useful data types and examples of where they have been released openly. It is worth noting however that not all useful datasets have been released as open data which means there is still plenty of progress to be made.

Table 1 - Relevant supplementary data

Data	Typical owner	Description	Example owner	Example open dataset	Data format(s) for examples
Emissions profiles	Industry, national government	Statutory reporting of pollutants from regulated industry processes	UK Department for Environment, Food & Rural Affairs (DEFRA) ⁴⁰	UK National Atmospheric Emissions Inventory (NAEI) ⁴¹	XLSX
Land use	National government, supranational government, GIS businesses	Locations and areas of heavy industry, airports, power stations, green space and other land uses relevant to pollution emissions	European Environment Agency (EEA) ⁴²	Urban Atlas ⁴³	ESRI shapefile, PDF, MS Word (for metadata descriptions)
Traffic density	National government, mobile telecoms companies	Live or historic data on the usage of specific roads and junctions	UK Department for Transport ⁴⁴	GB Road Traffic Counts ⁴⁵	CSV
Weather	National government	Live weather observation data as well as historic and forecast weather	Met Office ⁴⁶	UK hourly site-specific observations ⁴⁷	XML, JSON
Weather	Citizen sensors	Live sound pressure and wind speed data from CITI-SENSE sensors	CITI-SENSE project ⁴⁸	Sound pressure and wind speed ⁴⁹	CSV

⁴⁰ <https://www.gov.uk/government/organisations/department-for-environment-food-rural-affairs> (last accessed on 2015-10-25)

⁴¹ <http://naei.defra.gov.uk/data/> (last accessed on 2015-10-25)

⁴² <http://www.eea.europa.eu/> (last accessed on 2015-10-25)

⁴³ <http://www.eea.europa.eu/data-and-maps/data/urban-atlas> (last accessed on 2015-10-25)

⁴⁴ <https://www.gov.uk/government/organisations/department-for-transport> (last accessed on 2015-10-25)

⁴⁵ <http://data.gov.uk/dataset/gb-road-traffic-counts> (last accessed on 2015-10-25)

⁴⁶ <http://www.metoffice.gov.uk/> (last accessed on 2015-10-25)

⁴⁷ <http://www.metoffice.gov.uk/datapoint/product/uk-hourly-site-specific-observations> (last accessed on 2015-10-25)

⁴⁸ <http://www.citi-sense.eu/> (last accessed on 2015-10-25)

⁴⁹ <http://co.citi-sense.eu/ProductsServices/Data.aspx> (last accessed on 2015-10-25)

Instead we will focus mainly on the availability and quality of air pollution monitoring data. Air quality monitoring data is provided as open data from a variety of sources. Chief amongst those are regional, national and supranational government bodies. For instance, the European Environment Agency (EEA)⁵⁰ provides historic time series data from monitoring stations through the Air Quality e-Reporting⁵¹ (formerly AirBase⁵²). This can be used to validate models by providing baseline conditions. These are based on the national monitoring systems of member countries, in which the relevant agencies often also publish openly. In the UK for instance, DEFRA⁵³ provides a wide range⁵⁴ of air quality data⁵⁵ including historic time series pollution concentrations monitoring going back to 1972.

These monitoring networks however usually consist of large permanent facilities which are expensive to install and maintain. This means there are relatively few stations covering large areas.⁵⁶ They also rarely publish real-time data feeds, with many government bodies opting to provide only the information themselves.⁵⁷ However, these are not the only authority owned air quality sensors. Local and regional governments, especially in cities, often have their own small scale monitoring networks. Releasing data from these networks, especially in real-time, can give citizens and companies the opportunity to create useful tools for avoiding and tackling air quality issues.

However, only a few of these networks provide access to real time data feeds. One example is Bristol City Council which provides near real-time pollutant data from three sensors in the city.⁵⁸ In general, government sensor networks are very useful for collecting data in specific areas but they still leave a lot of gaps in actual air quality data. Given that air pollution levels can vary dramatically from one junction to the next, we must still rely heavily on modelling to get anywhere near street-level pollution levels. This lack of granular data can greatly hinder the development of more complex behavioural change solutions, such as route planning.

Following the low levels of coverage and lack of real-time data available from existing government networks, there has been a proliferation of sensing projects. Projects like sensing cities⁵⁹ and CITI-SENSE⁶⁰ are being driven by supranational and national government bodies. Outside of government there are also attempts to improve the availability of air quality data. For instance the LAQN which is run by King's College supplies real-time open sensing data.⁶¹ An ambitious new project, Air Sensa is a private non-profit venture to install 40'000 sensors in 20 urban areas across the UK.⁶² The aim is to generate street level real-time, open data in order to kick-start the creation of new tools.

Simultaneously, there has also been a dramatic rise in low cost sensing⁶³ and citizen-sensor projects.⁶⁴ In these cases, individuals set up sensors to monitor their own environment and often publish or display the data on a central platform. Examples include open-source raspberry pi based AirPi⁶⁵, air

⁵⁰ <http://www.eea.europa.eu/> (last accessed on 2015-10-25)

⁵¹ <http://www.eea.europa.eu/data-and-maps/data/eqereporting#tab-european-data> (last accessed on 2015-10-25)

⁵² <http://www.eea.europa.eu/data-and-maps/data/airbase-the-european-air-quality-database-8#tab-european-data> (last accessed on 2015-10-25)

⁵³ <https://www.gov.uk/government/organisations/department-for-environment-food-rural-affairs> (last accessed on 2015-10-25)

⁵⁴ http://uk-air.defra.gov.uk/assets/documents/reports/cat14/1307241318_Guide_to_UK_Air_Pollution_Information_Resources.pdf (last accessed on 2015-10-25)

⁵⁵ http://uk-air.defra.gov.uk/data/data_selector (last accessed on 2015-10-25)

⁵⁶ <http://uk-air.defra.gov.uk/interactive-map> (last accessed on 2015-10-25)

⁵⁷ <http://www.scottishairquality.co.uk/> (last accessed on 2015-10-25)

⁵⁸ <https://opendata.bristol.gov.uk/Environment/Latest-Air-Quality/hnkb-7z35> (last accessed on 2015-10-25)

⁵⁹ <http://www.sensingcities.org/> (last accessed on 2015-10-25)

⁶⁰ <http://citi-sense.nilu.no/> (last accessed on 2015-10-25)

⁶¹ <http://www.londonair.org.uk/LondonAir/API/> (last accessed on 2015-10-25)

⁶² <http://www.airsensa.org/> (last accessed on 2015-10-25)

⁶³ <http://www.sciencedirect.com/science/article/pii/S0160412014003547> (last accessed on 2015-10-25)

⁶⁴ <http://www.greencitystreets.com/crowdsourcing/crowdsourced-air-quality-data/> (last accessed on 2015-10-25)

⁶⁵ <http://airpi.es/index.php> (last accessed on 2015-10-25)

sensors attached to hire bikes⁶⁶ and a project aimed at empowering young people to tackle air quality in Kosovo.⁶⁷ These systems all depend on the development of Internet of Things (IoT) platforms which can handle vast amounts of data. One such platform is OpenSensors.io,⁶⁸ currently used by the Air Quality Egg network.⁶⁹ It which hosts projects that publish open data free of charge.

Generally, there remain some concerns over the quality of data gathered by lower-cost sensors however as the technology develops and scales these are likely to be somewhat abated. Currently these concerns can be partially mitigated by having low-cost sensors and traditional monitoring stations co-located to confirm accuracy. Tracing the data provenance might also help, for example enabling services to weight results by the type of sensor.

Table 2, below, lays out all the types of air quality monitoring data from the various sensor networks and owners.

Table 2 - Air monitoring data by type of data owner

Data owner	Description	Example open (or potentially open) dataset	Example owner	Data format(s) for examples
Supranational organisations	The European Commission collects air quality data submitted by member countries for reporting and policy purposes	Air Quality e-Reporting ⁷⁰ (formerly AirBase ⁷¹)	European Environment Agency (EEA) ⁷²	CSV
National government	National governments, through environment agency bodies collect air quality data for reporting and policy purposes	Many ⁷³	UK Department for Environment, Food & Rural Affairs (DEFRA) ⁷⁴	CSV
Local government	Local government organisations collect air quality data from their local area	Latest Air Quality ⁷⁵	Bristol City Council ⁷⁶	CSV, JSON, PDF, RDF, RSS, XLS, XLSX, XML
Private initiative	Initiatives have been set up to capture more data in order to tackle air quality issues	Potential - Full dataset download and real-time access through API	Air Sensa ⁷⁷	Unknown

⁶⁶ <http://www.techworld.com/blog/machination/mapping-air-quality-with-hire-bike-sensors-3618839/> (last accessed on 2015-10-25)

⁶⁷ <http://www.mobilisationlab.org/kosovo-youth-set-agenda-and-gather-data-for-change/#.VaFFoxNViko> (last accessed on 2015-10-25)

⁶⁸ <https://opensensors.io/> (last accessed on 2015-10-25)

⁶⁹ <http://airqualityegg.com/> (last accessed on 2015-10-25)

⁷⁰ <http://www.eea.europa.eu/data-and-maps/data/aireporting#tab-european-data> (last accessed on 2015-10-25)

⁷¹ <http://www.eea.europa.eu/data-and-maps/data/airbase-the-european-air-quality-database-8#tab-european-data> (last accessed on 2015-10-25)

⁷² <http://www.eea.europa.eu/> (last accessed on 2015-10-25)

⁷³ http://uk-air.defra.gov.uk/data/data_selector (last accessed on 2015-10-25)

⁷⁴ <https://www.gov.uk/government/organisations/department-for-environment-food-rural-affairs> (last accessed on 2015-10-25)

⁷⁵ <https://opendata.bristol.gov.uk/Environment/Latest-Air-Quality/hnkb-7z35> (last accessed on 2015-10-25)

⁷⁶ <http://www.bristol.gov.uk/> (last accessed on 2015-10-25)

⁷⁷ <http://www.airsensa.org/> (last accessed on 2015-10-25)

Citizen generated	Citizens are able to buy low-cost sensors in order to monitor the air quality where they live	Potential - Full dataset download and real-time access through API	Air Quality Egg ⁷⁸	Unknown
-------------------	---	--	-------------------------------	---------

There is a clear move by many different actors aimed at generating much more air quality data. There are two key factors in using this increasing amount of data to successfully tackle urban air pollution. Firstly, ensuring there is good coverage within and between these networks.⁷⁹ Secondly, making sure the data produced is interoperable through use of shared identifiers and common standards as well as providing consistent documentation on accuracy, timeliness, methods of collection and a whole range of other aspects. With such a large demand for this data, ensuring it is readily available as open data will amplify the solutions created and significantly reduce any duplication of effort.

2.4 How can the DataGraft platform help?

Opening data can have many benefits in helping tackle poor air quality. But these depend greatly on the quality, usefulness and usability of the data itself. These aspects must be considered from the outset when publishing data, otherwise the potential benefits might never be realised. Additionally, the scalability of data consumption is another core challenge. To what extent the benefits and synergies of multiple, colocated air sensor systems can be reaped will also depend on whether these systems, or at least the data they produce, are interoperable. Otherwise, a multitude of sensor systems might exist in silos, with none of their potential synergies being exploitable. In sum, ensuring that data is of maximum quality, usability, and interoperable across systems can require considerable efforts and technical skills from publishers. Both the publication and consumption of open data can suffer substantially when publishers don't have or cannot invest the time, skill and resources to meet these criteria.

The DaPaaS project⁸⁰ aims to tackle this concern through DataGraft,⁸¹ a platform and tools which significantly reduce the burden on publishing quality, useful and usable open data. Using DataGraft to publish data openly could facilitate the use of data in tackling air pollution in a number of ways.

Firstly, DataGraft provides users with a centralised, low-cost hosting platform for open data. It can be used to host up to 1GB of data for free and can therefore be useful to publishers with small amounts of data or for organisations to experiment with publishing without committing significant resources. For private users with air sensors or weather stations, this could provide a means to share their data without requiring any financial investment. For larger organisations seeking to publish large quantities of monitoring or supporting data but not wanting to invest in setting up a bespoke portal, the platform can be used not only to demonstrate the viability but act as a long term low-cost alternative. This increases the prospect of more actors publishing their data openly. Releasing more air quality monitoring data, as well as additional and contextual data, is likely to lead to more and improved solutions to better air quality.

With all or at least a significant amount of monitoring data being made open and provided using a set of shared identifiers and under a set of common standards, it will become easier to identify where the monitoring gaps are. It can also help to avoid duplication of efforts by effectively unifying various sen-

⁷⁸ <http://airqualityegg.com/> (last accessed on 2015-10-25)

⁷⁹ <http://dsg.nutn.edu.tw/msrg/home%20page/member/96patrick/pdf/Worst%20and%20Best-Case%20Coverage%20in%20Sensor%20Networks.pdf> (last accessed on 2015-10-25)

⁸⁰ <http://project.dapaas.eu/> (last accessed on 2015-10-25)

⁸¹ <http://datagraft.net/> (last accessed on 2015-10-25)

sor networks into a single federated system that is connected through a common standards. Although this does not replace the need to expand certain networks to include parts covered by others, because sensors vary in terms of cost and accuracy, it can at least help with these decisions. What is more, using larger more centralised hosting solutions makes it far easier for developers and interested parties to find useful data from a wide variety of publishers. This can greatly lessen the time it takes them to implement scalable solutions and services. By placing large amounts of relevant data in fewer places it is also likely to attract new solutions which do not require an in depth knowledge of all the data owners and providers to implement.

Secondly, DataGraft provides a set of tools which help publishers easily clean, transform and link the data. Together this toolbox has a large number of benefits which eventually could help to facilitate the development of scalable, reliable, high-quality services - which can also tackle air pollution more effectively. For publishers, this is based on the fact that uploading data to DataGraft is designed to be user-friendly for non linked data specialists. This user-oriented design helps to reduce the time and resource burden for preparing high quality, usable open data. For users, it centres on the utility, quality and accessibility of the data output in a number of ways. To begin with, DataGraft allows users to carry out commercial scale cleansing of tabular data using predefined operations with the option to specify more complex operations. This means that users can immediately access data hosted on the platform which has already been cleaned. For air pollution data this might involve removing nonsensical monitoring readings or providing interoperable data formatting, e.g. through consistently formatted timings.

A key aspect for publishers is that the operations carried out on one dataset can be saved and then repeated on another dataset. This is particularly useful when publishing data from the same source on a regular basis which is likely to have consistent quality issues. This is most important in the case of real-time data which can be pushed into the platform automatically on a regular basis by publishers. The utility of most of the air pollution data and supplementary data is particularly high when it is updated in either real-time or on a regular basis. After the cleansing, data can be transformed from common tabular formats such as excel spreadsheets to conform with common data standards. This greatly improves the usability of the data by increasing the number of tools users are able to process the data with. For instance with transformation to tabular csv (comma separated values) data.⁸²

Even more importantly, DataGraft allows publishers to transform tabular data to the RDF⁸³ linked data standard. This involves making the data far more useful for use by developers. By giving commonly referenced objects URIs (Unique Resource Identifiers), machines can be made to understand the relationship between different data about the same objects. Using such solutions, analytics services can easily draw on a significant number of different data sources and types. By adhering to the common standards, developers can access data published by a variety of publishers using the same SPARQL⁸⁴ queries. This makes it easier for programmers to help develop solutions using a wide variety of datasets. Without having to dig into the specifics of each individual dataset, developers will still be able integrate a number of variables and applicable data resources. For air pollution which often requires a particularly complex understanding of a range of factors this is particularly useful.

One of the key features of the DataGraft platform is that all these features and functions can be accessed through both the user-friendly web interface and programmatically through a set of well documented APIs. This means that publishers and users can interact with the platform, and therefore the data, in the manner which they find most comfortable and therefore most useful.

⁸² <http://data.okfn.org/doc/csv> (last accessed on 2015-10-25)

⁸³ <http://www.w3.org/RDF/> (last accessed on 2015-10-25)

⁸⁴ <http://www.w3.org/TR/sparql11-overview/> (last accessed on 2015-10-25)

Using high quality linked open data provided on the DataGraft platform could greatly impact the effort to tackle urban air pollution.

2.5 Conclusions

Tackling urban air pollution requires a wide variety of data-driven solutions implemented by a wide variety of actors at various levels. To best encourage evidence-based concerted action we need a lot of data to be made available. To ensure that this data is provided to those who are able and willing to tackle these problems demands the opening of this data to all. Efforts to do this are proliferating but progress is still slow. DEFRA carried out an assessment of data requirements,⁸⁵ finding that:

“Fewer than half of the UK’s air quality dataset users have access to all of the air quality monitoring, modelling and emissions data that they require, and the majority do not have access to the associated information which is needed to provide context and relationships between the causes and impacts of air pollution”.

Recommending that they integrate the data they provide to try to tackle these issues. Efforts have also been undertaken to provide linked air quality data to enhance the solutions proposed. Air Quality+ (AQ+) in Sheffield, England⁸⁶ and a project in Skopje, Macedonia⁸⁷ are two good examples where progress is being made but they remain the exceptions not the rule. OpenSensors.io,⁸⁸ a centralised platform for publishing real-time sensor data openly, has greatly increased the amount of citizen generated air quality data available. However, it still requires a great deal of technical skill to get the data and clean up it up before use.

The DaPaaS project’s DataGraft brings together a wide range of functions for making the publishing of high-quality, usable open data relevant to tackling air pollution easy. By providing a centralised platform for a variety of publishers to use, it could be used to accelerate the opening of data for tackling air pollution. This in turn could accelerate the coordinated concerted data-driven action required to tackle the problems posed by urban air pollution.

⁸⁵ http://uk-air.defra.gov.uk/assets/documents/reports/cat09/1102161123_Data_Integration_Report_v1-2.pdf (last accessed on 2015-10-25)

⁸⁶ <http://betterwithdata.co/2014/linked-open-data-for-air-quality/> (last accessed on 2015-10-25)

⁸⁷ http://www.researchgate.net/publication/275647451_Publishing_Skopje_Air_Quality_Data_as_Linked_Data (last accessed on 2015-10-25)

⁸⁸ <https://opensensors.io/> (last accessed on 2015-10-25)

3 Effectively managing water resources

3.1 What's the challenge?

Water is a vital natural resource on which all life on earth depends. It's the lifeblood which sustains the vast array of natural ecosystems which also has to meet the many domestic, agricultural and industrial demands of humanity.⁸⁹ Balancing these various demands presents a serious challenge to those responsible for managing water resources. Managing water resources effectively requires the tackling of a wide range of interrelated problems.

The natural systems which drive water distribution are inherently complex combinations of chemical, meteorological, geological and many other processes.⁹⁰ These systems are facing unprecedented change from the emerging effects of human-driven climate change.⁹¹ As well as increasing pressures from a demanding and growing society, which brings with it a wide variety of related social, political and economic challenges.⁹² The result is an incredibly complicated human and environmental system governed by a wide range of diverse factors and interrelationships.

So many things depend on this complex system; flooding and droughts, biodiversity and agricultural yield, drinking water supply and wastewater removal. A careful balancing act is required to keep all these things functioning. The danger being that too much focus on one might result indirectly negatively affect all the others. For instance, water pollution from agricultural, domestic and industrial water use causes damage to natural ecosystems.⁹³ But it also causes increased financial cost of treatment of water for drinking, estimated to cost up to £1.3 billion per year in England and Wales alone.⁹⁴

Overuse or 'over-abstraction', removing excessive amounts of water from natural sources for use by society, is a key threat to the balance of natural systems.⁹⁵ It is just one of the many societal and environmental factors affecting the number and severity of droughts. Droughts can have a huge impact on the environment, economy and society.⁹⁶ Even in Europe's temperate climate water scarcity is estimated to have affected at least a tenth of the population in 2007 and cost €100bn over the last thirty years.⁹⁷ Human-driven climate change and widespread construction of non-porous structures, buildings and car-parks, has also increased the risk of damage from flooding.⁹⁸ Flooding can cause significant environmental and societal damage. Europe suffered 213 major floods between 1998 and 2009, causing 1126 deaths and at least €52 billion in insured economic losses.⁹⁹

3.1.1 Taking an integrated systems approach

Effectively managing water resources requires us to balance the numerous interrelated supply and demand pressures on water systems. The best way to do this is to use an integrated systems ap-

⁸⁹ <http://www.eea.europa.eu/publications/water-resources-across-europe/> (last accessed on 2015-10-25)

⁹⁰ http://pubs.usgs.gov/circ/circ1139/htdocs/natural_processes_of_ground.htm (last accessed on 2015-10-25)

⁹¹ https://ec.europa.eu/research/environment/pdf/kina22422ens_web_water_and_cc.pdf (last accessed on 2015-10-25)

⁹² <http://www.fao.org/docrep/003/t0800e/t0800e0b.htm> (last accessed on 2015-10-25)

⁹³ <http://www.theguardian.com/environment/2011/jun/09/fish-thames-sewage> (last accessed on 2015-10-25)

⁹⁴ <http://www.nao.org.uk/wp-content/uploads/2010/07/1011188es.pdf> (last accessed on 2015-10-25)

⁹⁵ https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/297309/LIT_4892_20f775.pdf (last accessed on 2015-10-25)

⁹⁶ http://cip.cornell.edu/DPubS/Repository/1.0/Disseminate?view=body&id=pdf_1&handle=dns.gfs/1279121772 (last accessed on 2015-10-25)

⁹⁷ http://ec.europa.eu/environment/water/quantity/scarcity_en.htm (last accessed on 2015-10-25)

⁹⁸ https://www.ipcc.ch/publications_and_data/ar4/wg2/en/ch3s3-5-2.html (last accessed on 2015-10-25)

⁹⁹ http://ec.europa.eu/environment/water/flood_risk/index.htm (last accessed on 2015-10-25)

proach,¹⁰⁰ which considers the multitude of factors involved and their relationships.¹⁰¹ This is especially difficult because complex water systems can often transcend political and administrative boundaries. In recognition of this, countries often attempt to implement “integrated river basin management”¹⁰² strategies. The requirements of the European Union (EU) Water Framework Directive are a good example of where such an approach has been adopted at a transnational level.¹⁰³

These practices require a great degree of cooperation and coordination to implement. Even when encompassed within national boundaries there can be a wide variety of stakeholders with a vested interest in water resources. Table 1 below gives a breakdown of UK stakeholder types and examples.

Table 1 - Key UK stakeholder types with some examples

Stakeholder type	Role(s)	Examples
Government departments, agencies and regulators	To protect environment, regulate water supply, ensure careful planning of infrastructure and demand	<ul style="list-style-type: none"> • Environment Agency¹⁰⁴ • Defra¹⁰⁵ • Ofwat¹⁰⁶ • Drinking Water Inspectorate¹⁰⁷ • DCLG¹⁰⁸
Water industry	Provide water and wastewater services to consumers and industry	<ul style="list-style-type: none"> • 32 water companies¹⁰⁹ • Professional and industry representative bodies <ul style="list-style-type: none"> ◦ Water UK¹¹⁰ ◦ Water Industry Forum¹¹¹ ◦ Institute of Water¹¹² ◦ CIWEM¹¹³
Non-profit organisations	Carry out research and lobby government to maintain water quality and supply	<ul style="list-style-type: none"> • Environmental groups (Greenpeace UK¹¹⁴, Surfers Against Sewage¹¹⁵) • Research institutions (CEH¹¹⁶, CLUWRR¹¹⁷)
Water-dependent industry	Use water in the production of goods	<ul style="list-style-type: none"> • Variety of users¹¹⁸ <ul style="list-style-type: none"> ◦ Food manufacture

¹⁰⁰ <https://www.unesco-ihe.org/academic-departments/integrated-water-systems-governance> (last accessed on 2015-10-25)

¹⁰¹ <http://unesdoc.unesco.org/images/0013/001363/136355e.pdf> (last accessed on 2015-10-25)

¹⁰² http://www-wds.worldbank.org/external/default/WDSPContentServer/WDSP/IB/2007/10/17/000020953_20071017093644/Rendered/PDF/411500Intro0to1mgmt0NOTE1101PUBLIC1.pdf (last accessed on 2015-10-25)

¹⁰³ http://ec.europa.eu/environment/water/water-framework/info/intro_en.htm (last accessed on 2015-10-25)

¹⁰⁴ <https://www.gov.uk/government/organisations/environment-agency> (last accessed on 2015-10-25)

¹⁰⁵ <https://www.gov.uk/government/organisations/departments-for-environment-food-rural-affairs> (last accessed on 2015-10-25)

¹⁰⁶ <https://www.ofwat.gov.uk/> (last accessed on 2015-10-25)

¹⁰⁷ <http://dwi.defra.gov.uk/about/annual-report/index.htm> (last accessed on 2015-10-25)

¹⁰⁸ <https://www.gov.uk/government/organisations/departments-for-communities-and-local-government> (last accessed on 2015-10-25)

¹⁰⁹ <http://www.ofwat.gov.uk/industryoverview/today/watercompanies/> (last accessed on 2015-10-25)

¹¹⁰ <http://www.water.org.uk/about-water-uk> (last accessed on 2015-10-25)

¹¹¹ <http://www.waterindustryforum.com/> (last accessed on 2015-10-25)

¹¹² <https://www.instituteofwater.org.uk/> (last accessed on 2015-10-25)

¹¹³ <http://www.ciwem.org/> (last accessed on 2015-10-25)

¹¹⁴ <http://www.greenpeace.org.uk/> (last accessed on 2015-10-25)

¹¹⁵ <http://www.sas.org.uk/> (last accessed on 2015-10-25)

¹¹⁶ <http://www.ceh.ac.uk/> (last accessed on 2015-10-25)

¹¹⁷ <http://research.ncl.ac.uk/cluwrr/> (last accessed on 2015-10-25)

¹¹⁸ <http://data.gov.uk/dataset/water-use-by-industry-in-england-and-wales-2006-07> (last accessed on 2015-10-25)

		<ul style="list-style-type: none"> ○ Energy production (hydroelectric¹¹⁹, fracking¹²⁰)
Agricultural users	Use water to grow crops and sustain animals across the 17m hectares of agricultural land ¹²¹	<ul style="list-style-type: none"> ● Crop and animal farmers ● Vineyards (Denbies¹²²) ● Farmers' associations (NFU¹²³)
Households	Use water for drinking, washing and bathing. As well as recreational activities such as boating and fishing.	<ul style="list-style-type: none"> ● 26.5m households (99%¹²⁴ of 26.7m households¹²⁵) ● Consumer groups (Consumer Council for Water¹²⁶) ● Recreational users

With such a diverse range and large number of stakeholders, including in many cases stakeholders with no conventional relationship, implementing integrated water basin management practices is an incredibly difficult task. Without careful management and consideration, the effect on a community can be huge.¹²⁷ It requires the establishment of “an open, participatory decision-making process [of] close coordination among the many institutions that manage water resources and a strong focus on the stakeholders impacted”.¹²⁸

3.2 How can we tackle this problem using data?

Building a collaborative, integrated approach is key to tackling the challenge of managing water resources effectively. For such an approach to work in practice, it must be built on collective and shared understanding of the natural and human systems involved. Such an understanding must be built not only on solid scientific foundations but also on shared, real-world data sources. Ensuring that real-world data sources are available to the many organisations involved in managing water resources and also those who are using them is a necessary step for building a participatory integrated approach.

3.2.1 Data-driven modelling for management and planning of water resources

Understanding natural water systems and their relationship to the needs and nature of society is key to solving the challenges of managing water resources. This understanding is fundamentally rooted in the acquisition, sharing and use of relevant data.¹²⁹ By analysing the data directly or feeding it into complex computer models, water resource managers and planners can make more informed decisions.

¹¹⁹ <https://www.gov.uk/guidance/harnessing-hydroelectric-power> (last accessed on 2015-10-25)

¹²⁰ https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/277211/Water.pdf (last accessed on 2015-10-25)

¹²¹ <http://data.gov.uk/dataset/agriculture-in-the-united-kingdom/resource/9825419f-1d9b-430f-b5ba-486762260c61> (last accessed on 2015-10-25)

¹²² <http://www.denbies.co.uk/> (last accessed on 2015-10-25)

¹²³ <http://www.nfu.org/> (last accessed on 2015-10-25)

¹²⁴ <http://www.niassembly.gov.uk/globalassets/Documents/RaISe/Publications/2010/Regional-Development/11310.pdf> (last accessed on 2015-10-25)

¹²⁵ <http://www.ons.gov.uk/ons/rel/family-demography/families-and-households/2014/families-and-households-in-the-uk-2014.html> (last accessed on 2015-10-25)

¹²⁶ <http://www.ccwater.org.uk/> (last accessed on 2015-10-25)

¹²⁷ <https://www.opendemocracy.net/transformation/javan-briggs/we-have-right-to-have-that-basic-thing%E2%80%94%E2%80%99s-water> (last accessed on 2015-10-25)

¹²⁸ <https://www.nae.edu/Publications/Bridge/V38N2/WaterResourceManagementModels.aspx> (last accessed on 2015-10-25)

¹²⁹ <https://www.unesco-ihp.org/node/5665> (last accessed on 2015-10-25)

Many water resources managers opt to use modelling software which available from a wide range of specialist companies, such as WHS¹³⁰, DHI Group¹³¹, ESI¹³² and Oxford Scientific Software¹³³. However, some of these use proprietary data formats presenting interoperability issues. Projects, such as ENVISION¹³⁴, have attempted to tackle these difficulties. Government bodies, such as U.S. Geological Survey (USGS)¹³⁵, also share software¹³⁶ and techniques¹³⁷ in the public domain, under open source licenses, to help stakeholders take a holistic and open approach. One exemplar is Colorado's Decision Support Systems (CDSS) providing tools, models and data about Colorado's river basins.¹³⁸

Using measurement data and environmental modelling techniques, agencies such as the UK Environment Agency¹³⁹ can make decisions on water abstraction based on environmental and societal needs.¹⁴⁰ This process involves using data to determine an "environmental flow indicator (EFI)" which assesses whether river flows are sufficient to support ecological demands.¹⁴¹ They use this to assess the sustainability and efficacy of abstraction permit requests, taking into account other abstractions as well as natural features such as rainfall. When they grant licenses, they often stipulate the use abstraction meters to monitor the actual amount of abstraction.¹⁴²

These data-driven models also provide the relevant authorities with tools to plan future infrastructure and capacity. Whether this involves the balancing of different needs of stakeholders, supporting ecosystem renewal or mitigating the ecological and human risk of extreme events such as droughts.¹⁴³ They can also be used to include, map, plan for and control future demand. This can include involving water suppliers and government environmental organisations in housing and commercial planning decisions based on available water resources.¹⁴⁴

Taking this approach is not limited to the agencies responsible for environmental protection. The same data-driven models can also provide those concerned with the supply of water to users, both commercial and domestic. A number of projects, funded by the Natural Environment Research Council (NERC)¹⁴⁵ are using environmental data to build products for water companies.¹⁴⁶ For example, Remote Sensing Applications Consultants¹⁴⁷ are building a more detailed map of agricultural land usage to determine water usage. And Maxeler Technologies¹⁴⁸ are providing shorter-timescale groundwater models to help manage resources during drought and flood stresses.¹⁴⁹ Similar work has been done in the US by USGS and local water authorities.¹⁵⁰

¹³⁰ <http://www.hydrosolutions.co.uk/software-3.asp> (last accessed on 2015-10-25)

¹³¹ <http://www.dhigroup.com/upload/publications/brochures/waterresourcesmodelling.pdf> (last accessed on 2015-10-25)

¹³² <http://esinternational.com/water/> (last accessed on 2015-10-25)

¹³³ <http://www.oxscisoft.com/> (last accessed on 2015-10-25)

¹³⁴ <http://www.envision-project.eu/> (last accessed on 2015-10-25)

¹³⁵ <http://www.usgs.gov/> (last accessed on 2015-10-25)

¹³⁶ <http://water.usgs.gov/software/lists/alphabetical> (last accessed on 2015-10-25)

¹³⁷ <http://water.usgs.gov/techniques.html> (last accessed on 2015-10-25)

¹³⁸ <http://cdss.state.co.us/Pages/CDSSHome.aspx> (last accessed on 2015-10-25)

¹³⁹ <https://www.gov.uk/government/organisations/environment-agency> (last accessed on 2015-10-25)

¹⁴⁰ https://consult.defra.gov.uk/water/abstraction-reform/supporting_documents/abstractreformconsultmanage20131217.pdf (last accessed on 2015-10-25)

¹⁴¹ <https://www.gov.uk/government/collections/water-abstraction-licensing-strategies-cams-process> (last accessed on 2015-10-25)

¹⁴² https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/297309/LIT_4892_20f775.pdf (last accessed on 2015-10-25)

¹⁴³ <https://www.nae.edu/Publications/Bridge/V38N2/WaterResourceManagementModels.aspx> (last accessed on 2015-10-25)

¹⁴⁴ https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/292913/gean0107blln-e-e.pdf (last accessed on 2015-10-25)

¹⁴⁵ <http://www.nerc.ac.uk/> (last accessed on 2015-10-25)

¹⁴⁶ <http://www.nerc.ac.uk/innovation/activities/environmentaldata/businessproblems/env-data-projects/> (last accessed on 2015-10-25)

¹⁴⁷ <http://www.rsac.co.uk/> (last accessed on 2015-10-25)

¹⁴⁸ <https://www.maxeler.com/> (last accessed on 2015-10-25)

¹⁴⁹ <http://gtr.rcuk.ac.uk/project/B736CD6D-DF5B-42E7-B9A0-FEA2EA50734E> (last accessed on 2015-10-25)

¹⁵⁰ <http://pubs.usgs.gov/of/2014/1108/pdf/ofr2014-1108.pdf> (last accessed on 2015-10-25)

3.2.2 Controlling and reducing demand for water

Effective planning and management of supply is not enough to deal with all the problems facing the supply of water. Current systems were often set up to meet only competing human demands and thus dealt primarily focused on expanding supply.¹⁵¹ Taking a holistic approach to managing water resources requires including users in managing their own demand. One key way of doing this is through managing agricultural demand which is responsible for 24% of European abstraction, rising to more than 80% in some rural regions.¹⁵² Providing the agricultural industry with detailed water usage models can greatly improve its water usage. For instance, the wine producer Gallo used satellite data and modelling on a field by field basis to determine the amount of water they should use, resulting in up to 30% reduction in water used¹⁵³ and associated economic benefit.

Giving users the context and data about their use can help them change their own behaviour. This can be considered one of the major drivers in the move towards smart water metering,¹⁵⁴ lead in the UK by providers such as Thames Water¹⁵⁵ and Anglian Water.¹⁵⁶ Providing consumers with real-time information on their water usage can help them to identify leakages and control usage, and therefore cost.¹⁵⁷ It is very useful in drought conditions where it can also help water companies identify violators of imposed restrictions.¹⁵⁸ Providing this data to the consumer may not however be enough. Providing the context for personal water usage and providing means to actively change behaviour is also important. This was the focus of the Leeds Data Dive¹⁵⁹ open-innovation event jointly held by Yorkshire Water¹⁶⁰ and ODI Leeds.¹⁶¹ Teams created applications which allowed users to change their behaviour through gamification and community engagement strategies that went beyond simple reporting.¹⁶²

Beyond just helping to manage demand, giving users data on water flow can provide a number of extra benefits. For example, the GaugeMap¹⁶³ tool provided by Shoothill¹⁶⁴ can be used by a wide variety of recreational users to check river levels and flows which are important for anglers, boaters and rowers.¹⁶⁵ Similar recreational uses also include whitewater rafting in areas where this is possible.¹⁶⁶ More importantly, providing citizens with information on water can also be useful in times of extreme weather events such as flooding. Another example from Shoothill is FloodAlerts which provides citizens with real time warnings about flooding.¹⁶⁷

¹⁵¹ https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/228861/8230.pdf (last accessed on 2015-10-25)

¹⁵² <http://www.eea.europa.eu/publications/water-resources-across-europe/> (last accessed on 2015-10-25)

¹⁵³ <http://westgov.org/drought-forum/case-studies/324-agriculture/816-case-study> (last accessed on 2015-10-25)

¹⁵⁴ <http://www.metering.com/water-meters-one-in-four-will-be-smart-by-2020-as-europe-market-hots-up/> (last accessed on 2015-10-25)

¹⁵⁵ <https://www.thameswater.co.uk/your-account/17386.htm> (last accessed on 2015-10-25)

¹⁵⁶ <http://www.anglianwater.co.uk/business/business-services/smart-metering.aspx> (last accessed on 2015-10-25)

¹⁵⁷ <http://www.waterworld.com/articles/wwi/print/volume-26/issue-5/regulars/creative-finance/smart-water-metering-networks-an-intelligent-investment.html> (last accessed on 2015-10-25)

¹⁵⁸ <http://www.wired.com/2015/06/smart-water-meters-let-cities-spot-drought-defiers/> (last accessed on 2015-10-25)

¹⁵⁹ <http://www.eventbrite.co.uk/e/water-challenge-data-dive-odi-leeds-yorkshire-water-300pm-600pm-ish-on-friday-930-am-445pm-ish-on-tickets-17103817980> (last accessed on 2015-10-25)

¹⁶⁰ <https://www.yorkshirewater.com/> (last accessed on 2015-10-25)

¹⁶¹ <http://actuatedfutures.com/> (last accessed on 2015-10-25)

¹⁶² <http://leeds.theodi.org/2015/07/17/nine-teams-six-themes/> (last accessed on 2015-10-25)

¹⁶³ <http://www.gaugemap.co.uk/#!Map> (last accessed on 2015-10-25)

¹⁶⁴ <http://www.shoothill.com/> (last accessed on 2015-10-25)

¹⁶⁵ <http://www.shoothill.com/using-gaugemap-just-floodaware/> (last accessed on 2015-10-25)

¹⁶⁶ http://www.usgs.gov/blogs/features/usgs_top_story/measuring-the-flow-uses-of-streamflow-information/ (last accessed on 2015-10-25)

¹⁶⁷ <http://www.shoothill.com/floodmap/> (last accessed on 2015-10-25)

3.2.3 Understanding and tackling pollution

Most models are primarily built around managing the many factors flow and allocation of water resources. However, this does not take into account all the factors involved for successful integrated systems management. Key to this understanding is bringing in the causes and effects of water pollution. Especially given that only a quarter of UK water bodies meet EU standards for healthy functioning ecosystems.^{168,169} Society plays a huge role in causing water pollution. This comes in the form of both point source pollution¹⁷⁰ and diffuse source pollution.¹⁷¹ Point source pollution in most cases originates from heavy industry and wastewater treatment plants. One particular problem is from combined sewage overflows during periods of heavy rainfall.¹⁷² Another rising concern is the effect of hydraulic fracking on water pollution.¹⁷³ Diffuse pollution is most commonly associated with both agricultural¹⁷⁴ and urban¹⁷⁵ runoff, the process by which polluting materials are driven from fields and roads into the water supply.

Key to tackling both types of pollution is effective monitoring and regulation. First, there is clearly a need to monitor the health of water bodies in order to determine the level of the problem. Government agencies responsible for the environment use a range of techniques to determine the level of pollution in water bodies,¹⁷⁶ mainly based on sampling specific sites.¹⁷⁷ They use this to gauge the effects pollution is having on the natural ecosystems that depend on those bodies of water. In addition to understanding the extent of the problem it can be used to help estimate levels of diffuse pollution. The extent of point source pollution can be monitored through wastewater sampling¹⁷⁸ and inflow real time sensors.¹⁷⁹ This data can be used to some extent to determine policy and enforce regulation.

However, monitoring and attempting to tackle pollution on its own is only of limited effect. There are many factors in the wider water system that can affect and be affected by pollution. For instance, over-abstraction from natural sources compounds water pollution issues because less water means more concentrated, and therefore more damaging, pollutants.¹⁸⁰ Likewise diffuse pollution from agricultural runoff is often down to poor farming practices which might in turn be affecting farming output.¹⁸¹ Water pollution itself has an economic effect where polluted sources require more treatment to be made 'safe' for drinking water.¹⁸² It can also have a severe impact on recreational users such as surfers¹⁸³ and anglers,¹⁸⁴ which has a knock-on economic impact.

To maximise the effectiveness of water monitoring data it must be contextualised within the wider water system, taking a truly integrated systems approach. It must also be readily shared to help recrea-

¹⁶⁸ https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/228861/8230.pdf (last accessed on 2015-10-25)

¹⁶⁹ http://www.doeni.gov.uk/niea/overview_stds_and_classification.pdf (last accessed on 2015-10-25)

¹⁷⁰ <http://www.eea.europa.eu/themes/water/water-pollution/point-sources/point-sources> (last accessed on 2015-10-25)

¹⁷¹ <http://www.eea.europa.eu/themes/water/water-pollution/diffuse-sources> (last accessed on 2015-10-25)

¹⁷² <http://www.theguardian.com/environment/2011/jun/09/fish-thames-sewage> (last accessed on 2015-10-25)

¹⁷³ <http://theyee.ca/News/2015/06/08/Water-Pollution-Fracking/> (last accessed on 2015-10-25)

¹⁷⁴ http://water.epa.gov/polwaste/nps/agriculture_facts.cfm (last accessed on 2015-10-25)

¹⁷⁵ http://water.epa.gov/polwaste/nps/urban_facts.cfm (last accessed on 2015-10-25)

¹⁷⁶ http://www.who.int/water_sanitation_health/resourcesquality/wqmonitor/en/ (last accessed on 2015-10-25)

¹⁷⁷ <http://www.eea.europa.eu/themes/water/status-and-monitoring/monitoring-of-waters/introduction-and-overview-of-monitoring-activities> (last accessed on 2015-10-25)

¹⁷⁸ https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/69592/pb13811-waste-water-2012.pdf (last accessed on 2015-10-25)

¹⁷⁹ http://web.sbe.hw.ac.uk/staffprofiles/bdgsa/11th_International_Conference_on_Urban_Drainage_CD/ICUD08/pdfs/560.pdf (last accessed on 2015-10-25)

¹⁸⁰ https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/228861/8230.pdf (last accessed on 2015-10-25)

¹⁸¹ <http://water.epa.gov/polwaste/nps/agriculture.cfm> (last accessed on 2015-10-25)

¹⁸² <http://www.nao.org.uk/wp-content/uploads/2010/07/1011188es.pdf> (last accessed on 2015-10-25)

¹⁸³ <http://www.bbc.co.uk/news/uk-england-cornwall-33681269> (last accessed on 2015-10-25)

¹⁸⁴ <http://www.anglingtrust.net/page.asp?section=551> (last accessed on 2015-10-25)

tional users, conservationists and other stakeholders plan their activities accordingly as well as lobby government and regulators when good conditions are not achieved.

3.3 What (open) data do we need?

Integrated systems approaches must be built on a robust data infrastructure. This is due to the necessity of extensive and diverse data, particularly when attempting to understand and manage the entire complex water systems. The main challenges emerge from the diverse range of sources which can provide data both directly related to water systems and data which provides the wider context of water resources management. To match or exceed the wide range of sources are a wide range of data users, often the same organisations that publish but also a wide variety of end users (public, private and third sector organisations as well as citizens).

The most efficient means of sharing this data with all those who require it is by making it open to, and usable by, all. This approach is especially useful in fostering an open and inclusive decision making, which is necessary for good integrated systems management. Given the extent and diversity of data, this requires not only clear open licensing but efforts to make the data more interoperable. This is primarily achieved through use of shared identifiers and common standards as well as providing consistent documentation on accuracy, timeliness, methods of collection and a whole range of other aspects.

Recognition of the need for open data in tackling water resources management problems in a concerted way is already prevalent in certain circles. Before examining the UK and EU water data infrastructure it is worth noting the trend in the US. This began with a sharing agreement¹⁸⁵ between United States Army Corps of Engineers (USACE)¹⁸⁶, National Oceanic and Atmospheric Administration (NOAA)¹⁸⁷, and USGS. Recognising the benefits of this joined up approach to data infrastructure but also the benefits of wider sharing¹⁸⁸ lead to the establishment of the Open Water Data Initiative.¹⁸⁹ This initiative created a wide range of opportunities which would not have occurred without it, such as the National Flood Interoperability Experiment¹⁹⁰ and the establishment and work of the Open Water Foundation.¹⁹¹

In the UK, the Environment Agency is leading the way on publishing data concerned with water systems. This is perhaps unsurprising given their mandate to coordinate and ensure sustainable water resources. There are however, a wide range of government agencies and departments which are engaged in publishing relevant data. A brief outline of the types of data being published by these publishers about water systems is presented in Table 2.

Table 2 - Overview of environmental data types provided by various governmental bodies

Data type	Description	Publishers	Example dataset	Data format(s) for examples
Water resource	Geo-located natural and man made physical	Ordnance Survey ¹⁹²	Open Rivers ¹⁹³	ESRI shape files, CSV, GML.

¹⁸⁵ <http://www.usgs.gov/newsroom/article.asp?ID=2797&from=rss#.Vdhb1VNVhBd> (last accessed on 2015-10-25)

¹⁸⁶ <http://www.usace.army.mil/> (last accessed on 2015-10-25)

¹⁸⁷ <http://www.noaa.gov/> (last accessed on 2015-10-25)

¹⁸⁸ http://www.usgs.gov/blogs/features/usgs_top_story/using-every-drop-of-information-the-open-water-data-initiative/ (last accessed on 2015-10-25)

¹⁸⁹ <http://acwi.gov/spatial/owdi/> (last accessed on 2015-10-25)

¹⁹⁰ <http://www.cae.utexas.edu/prof/maidment/giswr2014/Synopsis/GISWRSynopsis12.pdf> (last accessed on 2015-10-25)

¹⁹¹ <http://openwaterfoundation.org/> (last accessed on 2015-10-25)

¹⁹² <https://www.ordnancesurvey.co.uk/> (last accessed on 2015-10-25)

infrastructure	infrastructure such as rivers, dams and reservoirs			
Water gauge	Direct data feeds from river gauges capturing flow and depth of water bodies	Environment Agency	Real-time and near-real-time river level data ¹⁹⁴	JSON, CSV (archive data)
Water quality	Chemical and biological sampling and monitoring data from specific site inspections	Environment Agency	UK Water Quality Sampling Harmonised Monitoring Scheme Detailed Data ¹⁹⁵	Shape file, XML, HTML.
Flooding	Geo-located risks and warnings in real time	Environment Agency	Risk of Flooding from Rivers and Sea ¹⁹⁶	Shape file, XML, HTML.
Wastewater	Flow, quality and resource locations of wastewater system	European Environment Agency ¹⁹⁷	Wastewater treatment locations and compliance (EU wide) ¹⁹⁸	Microsoft access, CSV.
Land use	Granular land use data mainly agricultural (satellite data is very useful)	ESA ¹⁹⁹ (USGS)	SENTINEL ²⁰⁰ (LANDSAT ²⁰¹)	<i>Unknown</i>
Topography	Geospatial mapping of terrain elevation profiles	Environment Agency	LIDAR (various) ²⁰²	ESRI ASCII Raster, georeferenced JPEG
Weather data	Real-time and forecast weather data particularly precipitation and temperature	Met Office ²⁰³	UK hourly site-specific observations ²⁰⁴	XML, JSON

There is also a wide range of data which provides context for integrated systems approaches. For example, Ofwat releases the statutory financial reporting of water companies,²⁰⁵ although this is not clearly licensed and is only provided as PDF. Other forms of contextual data originate outside of government. The work of the Carbon Disclosure Project (CDP)²⁰⁶ in collecting survey responses about the

¹⁹³ <https://www.ordnancesurvey.co.uk/business-and-government/products/os-open-rivers.html> (last accessed on 2015-10-25)

¹⁹⁴ <http://data.gov.uk/dataset/real-time-and-near-real-time-river-level-data> (last accessed on 2015-10-25)

¹⁹⁵ <http://data.gov.uk/dataset/uk-water-quality-sampling-harmonised-monitoring-scheme-detailed-data1> (last accessed on 2015-10-25)

¹⁹⁶ <http://data.gov.uk/dataset/risk-of-flooding-from-rivers-and-sea> (last accessed on 2015-10-25)

¹⁹⁷ <http://www.eea.europa.eu/> (last accessed on 2015-10-25)

¹⁹⁸ <http://www.eea.europa.eu/data-and-maps/data/waterbase-uwatd-urban-waste-water-treatment-directive-4> (last accessed on 2015-10-25)

¹⁹⁹ <http://www.esa.int/> (last accessed on 2015-10-25)

²⁰⁰ <https://scihub.esa.int/> (last accessed on 2015-10-25)

²⁰¹ <http://landsat.usgs.gov/> (last accessed on 2015-10-25)

²⁰² <https://data.gov.uk/data/search?q=LIDAR&publisher=environment-agency&page=1> (last accessed on 2015-10-25)

²⁰³ <http://www.metoffice.gov.uk/> (last accessed on 2015-10-25)

²⁰⁴ <http://www.metoffice.gov.uk/datapoint/product/uk-hourly-site-specific-observations> (last accessed on 2015-10-25)

²⁰⁵ <https://www.ofwat.gov.uk/regulating/junereturn/jrhistoricdata/> (last accessed on 2015-10-25)

²⁰⁶ <https://www.cdp.net/en-US/Pages/HomePage.aspx> (last accessed on 2015-10-25)

water needs and usage of commercial organisations is an interesting example of this.²⁰⁷ We must also recognise the work of intergovernmental organisations, such as the Food and Agricultural Organisation (FAO),²⁰⁸ in collating high level statistics²⁰⁹ and also more recent efforts to provide satellite data.²¹⁰ These statistics can provide political context for the attitudes of certain governments to water resources management.

A large number of the key datasets required to build a decent water systems data infrastructure are now readily available as open data. This is a very promising start but by no means comprehensive. There is for instance recognition of the need for more real-time data, particularly in the case of wastewater flow and quality.²¹¹ Another useful addition would be the provision of comprehensive data on current abstraction licences and actual withdrawals. The Environment Agency already provides aggregated statistics²¹² and presents actual licence information on a map²¹³ but does not appear to publish the full records in a queryable manner that would allow for external modelling.²¹⁴ In addition, some government and academic organisations such as the Centre for Ecology and Hydrology (CEH)²¹⁵ have a wide range of useful data which are not made open, or at least not clearly open.²¹⁶ Making this data open and clearly so, by attaching clear licences to the data is important for further developing truly integrated approaches to water resource management.

More data is also required around the demand, potential demand and actual usage of public water supply. In particular there is a need for water companies to share more data. Yorkshire Water²¹⁷ is a good example of a utility which recognises this need by publishing data²¹⁸ (such as anonymised consumer water usage data²¹⁹) on Leeds Data Mill.²²⁰ This data can be useful in building up a picture of the amount of and variation in domestic usage. This can then be contextualised with a wide variety of other data concerning residential areas like demographics and house prices. Likewise open data on future housebuilding and planning must be made available to allow for effective future planning of water infrastructure. Initiatives such as Open Planning²²¹ which aim to create more inclusive and open decision making planning would be useful in creating a strong water resource data infrastructure as well as the wider benefits they pursue. To drive down demand to reduce overall water usage water utilities must also share real-time data with consumers about their own usage, using smart metering. Allowing customers to track their own usage in real time empowers them to change their behaviour. If they have access to other data then they can contextualise this usage and it may drive them further in adopting new practices.

There are efforts underway, both top-down and bottom-up approaches, aiming to tackle the current data shortfalls. Firstly, Defra has committed to opening up at least 8000 datasets²²² many of which will be useful in water resources management, including more satellite and biodiversity data. A number of

²⁰⁷ <https://data.cdp.net/Water/Water-2014-Open-dataset/5fe7-nx93> (last accessed on 2015-10-25)

²⁰⁸ <http://www.fao.org/home/en/> (last accessed on 2015-10-25)

²⁰⁹ <http://www.fao.org/nr/water/aquastat/wastewater/index.stm> (last accessed on 2015-10-25)

²¹⁰ <http://www.fao.org/news/story/en/item/326000/icode/> (last accessed on 2015-10-25)

²¹¹ <http://www.sciencedirect.com/science/article/pii/S0378377413002163> (last accessed on 2015-10-25)

²¹² <https://www.gov.uk/government/statistical-data-sets/env15-water-abstraction-tables> (last accessed on 2015-10-25)

²¹³ http://maps.environment-agency.gov.uk/wiyby/wiybyController?x=357683&y=355134&scale=1&layerGroups=default&ep=map&textonly=off&lang=en&topic=water_abstractions

(last accessed on 2015-10-25)

²¹⁴ <http://data.gov.uk/data-request/current-water-abstraction-licences> (last accessed on 2015-10-25)

²¹⁵ <http://www.ceh.ac.uk/> (last accessed on 2015-10-25)

²¹⁶ <https://catalogue.ceh.ac.uk/documents#page=1> (last accessed on 2015-10-25)

²¹⁷ <https://www.yorkshirewater.com/> (last accessed on 2015-10-25)

²¹⁸ http://leedsdatamill.org/dataset?q=&sort=score+desc%2C+dp_modified+desc&organization=yorkshire-water (last accessed on 2015-10-25)

²¹⁹ <http://leedsdatamill.org/dataset/customer-meter-data> (last accessed on 2015-10-25)

²²⁰ <http://leedsdatamill.org/> (last accessed on 2015-10-25)

²²¹ <http://thecreativeexchange.org/projects/open-planning> (last accessed on 2015-10-25)

²²² <https://www.gov.uk/government/news/environment-secretary-unveils-vision-for-open-data-to-transform-food-and-farming> (last accessed on 2015-10-25)

civil society initiatives are working to improve the amount of open water data through open source monitoring hardware and crowdsourcing.²²³ Examples include the Open Water project,²²⁴ a program to monitor flooding in Venice²²⁵ and NASA inspired water quality monitoring application.²²⁶

But the amount of data required is only one aspect, of equal importance is that the data is reliable, up-to-date and accessible as well as of appropriate spatial and temporal scale.²²⁷ In this regard, there is a genuine mixture of quality when it comes to water systems data infrastructure. While some datasets miss out on a few of these key areas there are also so examples of exceptional publishing. One of which is the Bathing Water Quality data²²⁸ published by the Environment Agency as linked open data,²²⁹ with the help of linked data specialists Epimorphics.²³⁰ Providing the data at such a high standard greatly increases its usefulness and value for integrated infrastructure management.

3.4 How can the DataGraft platform help?

The best approach to comprehensive and sustainable water resources management is through a holistic integrated systems approach. We have explored how such a collaborative approach fundamentally relies on a consistent, extensive and efficient data infrastructure. To ensure that the data infrastructure is all of these it must be built on the principle of 'open by default'. Everything that can be made open should be, subject only to concerns over privacy. The fundamental problem with meeting this need is meeting the requirements of the technical infrastructure to host it. However, the DataGraft platform has been designed to specifically meet many of these needs. We will explore how using it might help to tackle the problems with managing water resources effectively and sustainably.

DataGraft provides a cloud-based platform for transforming, publishing and hosting open data. It can be used by both government and non-government organisations and actors to host data and provide reliable access to the data. This allows the many producers of data on water resources to publish it as open data in the same place, including private citizens engaged in crowdsourcing initiatives. This differentiates it from the existing governmental²³¹ and inter-governmental²³² attempts to unify public sector water resource data. This more unified approach allows for even greater collection and wider comparison of data relating to water resources. This can be useful for identifying critical spatial or temporal gaps which might be filled by alternative sources. A single point of reliable data access makes it a lot easier for developers and modellers to identify data relevant to their projects. This is especially useful for modelling which can often be improved with more data. The platform itself is reliable, persistent and scalable which means users are more likely to invest time and effort into using it, as they know it will consistently provide the data they require.

While having a centralised open platform to host the vast and varied data requirements of managing water resources is useful it does not ensure that the data itself is consistent and useable. However, the DataGraft platform includes commercial scale customisable data cleaning tools. Using these can help improve data quality when publishing rather than leaving it to each individual user. Increasing the

²²³ <http://www.forbes.com/sites/federicoguerrini/2014/05/12/grassroot-water-monitoring-and-open-data-a-fruitful-combination/> (last accessed on 2015-10-25)

²²⁴ <http://publiclab.org/wiki/open-water> (last accessed on 2015-10-25)

²²⁵ <http://www.zdnet.com/article/open-data-vs-the-flood-how-high-tech-is-helping-venice-deal-with-high-tides/> (last accessed on 2015-10-25)

²²⁶ <http://gcn.com/articles/2015/04/17/mwater-data-app.aspx> (last accessed on 2015-10-25)

²²⁷ <http://www.eea.europa.eu/publications/water-resources-across-europe/> (last accessed on 2015-10-25)

²²⁸ <http://data.gov.uk/dataset/bathing-water> (last accessed on 2015-10-25)

²²⁹ <http://www.epimorphics.com/web/projects/bathing-water-quality> (last accessed on 2015-10-25)

²³⁰ <http://www.epimorphics.com/> (last accessed on 2015-10-25)

²³¹ <http://acwi.gov/spatial/owdi/> (last accessed on 2015-10-25)

²³² <http://water.europa.eu/> (last accessed on 2015-10-25)

quality of individual datasets can make them far more useful to potential users especially when they can expect a minimum level of quality from all datasets on the platform.

This however does not necessarily help when integrating data from different publishers or about different phenomena, which is key in modelling exercises. It also does not make the data more accessible or manipulable for users. To aid in this respect DataGraft allows publishers to transform cleaned datasets from tabular (csv²³³) to queryable linked data based on common web standards (RDF²³⁴). This provides a whole host of benefits when it come to reuse of data on the platform. For one, the use of URIs allows different datasets to be more readily integrated with one another. Using common web standards allows developers to use the same SPARQL²³⁵ queries to interrogate similar datasets, reducing the time taken to integrate a wide variety of data. This feeds directly into the needs of integrated approaches to water resources modelling. The benefits of providing open data as easy to use linked open data are already witnessed in the wide variety of uses associated with the Environment Agency's Bathing Water Quality data.

In short, a strong water data infrastructure underpinned by the technical capabilities of the DataGraft platform could greatly improve efforts to manage water resources in an integrated systems manner.

3.5 Conclusions

Integrated water systems management approaches are mandated by EU directive. Taking this approach has been shown to be beneficial for the health of ecosystems and society. These however require a great deal of effort to implement and coordinate. To ensure their success and reduce the amount of effort required they must be underpinned by a strong data infrastructure. This infrastructure is most effective when it is made up of readily available and usable linked open data. The wider structures around decision making should reflect this inclusive and open attitude by involving all stakeholders, be they government bodies, private companies or citizens.

²³³ <http://data.okfn.org/doc/csv> (last accessed on 2015-10-25)

²³⁴ <http://www.w3.org/RDF/> (last accessed on 2015-10-25)

²³⁵ <http://www.w3.org/TR/sparql11-overview/> (last accessed on 2015-10-25)

4 Improving cities' resilience to extreme weather events

4.1 What is the challenge?

Extreme weather events - bouts of unusual, severe or unseasonal conditions - cause devastating human and economic damage around the world each year.

According to Rachel Kyte, World Bank Vice-President for Sustainable Development, "over the last 30 years, the world has lost more than 2.5 million people and almost \$4 trillion to natural disasters. Economic losses are rising – from \$50 billion each year in the 1980s, to just under \$200 billion each year in the last decade. And about three quarters of those losses are a result of extreme weather."²³⁶

Widespread scientific opinion suggests that climate change is increasing the frequency and intensity of these extreme weather events. The rising concentration of humans in cities and urban environments will also result in higher exposure and vulnerability to extreme weather than ever before, resulting in a critical challenge to humanity.

There are a number of strategies that can be adopted in order to improve our cities' resilience to these extreme weather events and the economic, societal and environmental disasters they cause. In 2009 the UN General Assembly established the United Nations Office for Disaster Risk Reduction (UNISDR)²³⁷ to serve as focal point for international disaster strategy.

The UNISDR's Sendai Framework 2015-2030²³⁸, which was adopted by UN Member States in March 2015, sets out four strategic priorities:

- understanding disaster risk
- strengthening disaster risk governance/management
- investing in disaster risk reduction for resilience
- enhancing disaster preparedness for effective response

Open data is beginning to impact on these strategies. In this report, we look at the way it can be used to both develop a better understanding of risk and improve response, recovery and reconstruction - ultimately increasing our cities' collective resilience to extreme weather events.

4.2 How can we address the challenge using data?

4.2.1 Better understanding the risks

At the foundation of improving cities' resilience to extreme weather events is the better understanding of the risks they may encounter.

The UNISDR's outgoing Hyogo Framework for Action (HFA)²³⁹ states that "unless cities have a clear understanding of the risks they face, planning for meaningful disaster risk reduction may be ineffective. Risk analysis and assessments are essential prerequisites for informed decision making."

²³⁶ <http://www.worldbank.org/en/news/press-release/2013/11/18/damages-extreme-weather-mount-climate-warms> (last accessed on 2015-10-20)

²³⁷ <http://www.unisdr.org/> (last accessed on 2015-10-20)

²³⁸ http://www.unisdr.org/files/43291_sendaiframeworkfordrren.pdf (last accessed on 2015-10-20)

²³⁹ http://www.unisdr.org/files/1037_hyogoframeworkforactionenglish.pdf (last accessed on 2015-10-20)

An open approach - what open data do we need?

A risk information and monitoring system may exist at a national, local or even individual level. Regardless of scope, an effective system needs to ingest a variety of input data from a number of different sources. Table 1 shows some types of data that may be used by a risk information and monitoring system to understand the risks resulting from extreme weather events.

Table 1: types of data used in risk information and monitoring systems

Data type	Description of data
Geospatial data	Data including topographic and physical maps of natural and man-made features such as mountains, rivers and forests, and transport, building and critical asset infrastructure.
Hydrological data	Data on the state of water, such as rivers, lakes and oceans, which may include real-time river and sea levels and flow data, flood zone locations, real-time and historical flood data, and water quality and temperature data.
Land ownership data	Data including the location, dimensions, boundaries and ownership of land parcels.
Land use data	Data regarding land usage and changes in usage, such as different building use and crop or vegetation cover.
Soil data	Data on the state of soil, including soil maps, expected soil conditions and nutrients, and emissions of land pollutants and contaminated land.
Statistical data	Various population data, which may include a variety of demographic statistics.
Weather / meteorological data	Real-time and historic observational, and forecast data, which may include weather states, temperatures, rainfall, radiation, moisture, humidity, evaporation, and climate maps.

The UN's HFA tasked cities to “consider creating an information and monitoring system that includes input data from and is accessible to all actors, including civil society, the production sector (for example, agriculture, mining, commerce and tourism) and the scientific and technical community.” In doing so, it provided an early indication of the need for a considered data infrastructure²⁴⁰ and the availability of open data.

Founded by the Global Facility for Disaster Reduction and Recovery (GFDRR)²⁴¹ in 2011, the Open Data for Resilience Initiative (OpenDRI)²⁴² attempts to harness open data to reduce vulnerability to natural hazards and the impacts of climate change. It partners with governments, international organizations, technical institutions, and civil society groups to support the development of open systems to create, share, and use disaster risk and climate change information. Over 100 million people in 50 countries have gained improved access to risk information through GFDRR-supported geospatial data sharing platforms since 2010.

²⁴⁰ <http://theodi.org/who-owns-our-data-infrastructure> (last accessed on 2015-10-16)

²⁴¹ <https://www.gfdr.org/> (last accessed on 2015-10-16)

²⁴² <https://www.gfdr.org/opendri> (last accessed on 2015-10-16)

In Malawi, for example, OpenDRI has helped to launch an open data platform - the MASDAP GeoNode²⁴³. It was created to ensure that historical and current data remains accessible, giving the public and other stakeholders access to key information about their disaster risk. This data related to the built environment has been used²⁴⁴ with InaSAFE software²⁴⁵ to calculate ex-ante disaster impact scenarios across the country.

An increasing amount of data pertinent to risk information and monitoring, such as geospatial data, has been made available by governments as open data over the past five years. It is often found on national data portals such as data.gov²⁴⁶ (US) and data.gov.uk²⁴⁷ (UK).

However, useful, and sometimes critical, data used in a risk information and monitoring system may not be collected, maintained or published by the city or government itself. A wide range of stakeholders may need to be involved in providing the data behind an effective system.

As Arup, the multinational services firm specialising in the built environment, have recently said, “everyone - business, government, civic society, academia and NGOs - has a role to play in building resilience, and everyone stands to benefit”²⁴⁸.

Building risk information and monitoring systems

Resurgence²⁴⁹, a UK startup incubated by the ODI, specialises in the use of open data to help assess the fluctuating nature of risk in cities. It helps to develop tools that harness open data along with other technologies, such as GIS mapping and Internet of Things (IoT), to build better urban resilience.

Alongside other experts in the field of urban resilience, Resurgence is working on the Data Toolkit for 100 Resilient Cities, to be published in 2016. The Toolkit seeks to enable cities to identify, manage and innovate from the data that is key to their city challenges and the delivery of their resilience strategies. It will help them to understand where the data they need is most likely to be held or whether it needs to be gathered from scratch, and provide guidance on how to negotiate access, use and re-use of data held by other parties.

The HFA also suggested that the outputs from a city's risk information system, such as dashboards or targeted reports, should be made available as widely as possible in order to maximise the collective awareness and understanding of risk. Through their Global Dialogue on Emerging Technologies for Urban Resilience²⁵⁰, the Red Cross²⁵¹ also argues the need for an open approach to risk information. It states that an effective resilience-strengthening technology solution “is accessible. It is open, inclusive and increasingly affordable for consumers. [It is also] governed by trustworthy leaders, systems and policies. It has access to relevant data and responsibly manages the data it generates.”²⁵²

Arup maintains a significant voice in the building of resilient cities. It has begun to develop resilience framework, invested in a number of related projects around the world and leads the 100 Resilient Cities Network. According to them, “there is a recognised increase in extreme weather and natural catastrophes, which are caused by population growth, demographics, denser development, increased

²⁴³ <http://www.masdap.mw> (last accessed on 2015-10-16)

²⁴⁴ https://www.gfdrr.org/sites/default/files/2_Innovation_lab_Factsheet_OpenDRI_rev.pdf (last accessed on 2015-10-16)

²⁴⁵ <https://github.com/AIFDR/inasafe> (last accessed on 2015-10-16)

²⁴⁶ <http://www.data.gov/> (last accessed on 2015-10-16)

²⁴⁷ <http://data.gov.uk/> (last accessed on 2015-10-16)

²⁴⁸ <http://digital.edition-on.net/wardour/ArupDesignBook2015> (last accessed on 2015-10-16)

²⁴⁹ <http://www.resurgence.io/> (last accessed on 2015-10-16)

²⁵⁰ <http://www.tech4resilience.org/> (last accessed on 2015-10-16)

²⁵¹ <http://www.redcross.org.uk/> (last accessed on 2015-10-16)

²⁵² <http://www.tech4resilience.org/the-basics.html> (last accessed on 2015-10-16)

complexity in infrastructure, normal climate cycles and global climate change... A key component of natural disaster resilience is to understand the risk from natural hazard events and the ability to react rapidly.”²⁵³

Arup has built their own risk information system to monitor risk (such as that posed by extreme weather) using open data. Its Hazard Owl²⁵⁴ system uses real-time information from public data feeds to constantly assess the risk of damage to businesses’ portfolios of assets. If a pre-defined level of risk is exceeded, those with assets under protection are sent an alert so they can quickly initiate risk reduction, mitigation and business continuity actions.

As well as integrated risk information and monitoring systems, some organisations are developing tools that can help a city’s citizens understand the risks of specific extreme weather types.

An example of this is flooding. According to Swiss Re²⁵⁵, a global reinsurer, across more than 600 of the world’s largest metropolitan areas, around 400m urban dwellers are in danger of coastal or river flooding.²⁵⁶

In the UK, Shoothill²⁵⁷, a software development company that specialises in data visualisation and online mapping, has created a number of products using Environment Agency open flood data. The company’s FloodAlerts²⁵⁸ product is an online graphical representation of flood warnings, which provides localised updates to keep users informed about flooding in their area. Users can also visualise flood risks and calculate the risk of flooding to their property by river or sea, using its Check My Flood Risk²⁵⁹ tool.

GaugeMap²⁶⁰, a live map of river levels based on data from river level monitoring gauges, can significantly improve a city’s collective understanding of the risks floods pose. It is updated every 15 minutes and all gauges - over 2,400 - have been assigned a Twitter account for local citizens to follow and be alerted of their own flood risk.

Despite the development of significant risk monitoring and information systems such as these, a UNISDR’s Sendai Framework key strategic point describes the need “to substantially increase the availability of and access to multi-hazard early warning systems and disaster risk information and assessments to the people by 2030.”²⁶¹ A component of this increase is the availability of the data itself - multi-hazard early warning and risk information and monitoring systems rely on access to an abundance of different datasets.

4.2.2. Improving response, recovery and reconstruction

In the event of an extreme weather event, a city’s ability to respond, recover and reconstruct effectively is vital.

A lack of access to critical data in the aftermath of a disaster can significantly impede informed decision making, reduce the effectiveness of resource planning and limit the coordination of services.

²⁵³ <http://digital.arup.com/services/hazard-owl/> (last accessed on 2015-10-16)

²⁵⁴ Ibid

²⁵⁵ <http://www.swissre.com/> (last accessed on 2015-10-16)

²⁵⁶ <http://digital.edition-on.net/wardour/ArupDesignBook2015> (last accessed on 2015-10-16)

²⁵⁷ <http://www.shoothill.com/> (last accessed on 2015-10-16)

²⁵⁸ <http://www.shoothill.com/floodmap/> (last accessed on 2015-10-16)

²⁵⁹ <http://www.checkmyfloodrisk.co.uk/> (last accessed on 2015-10-16)

²⁶⁰ <http://www.gaugemap.co.uk/> (last accessed on 2015-10-16)

²⁶¹ http://www.unisdr.org/files/43291_sendaiframeworkfordrren.pdf (last accessed on 2015-10-20)

The power of mapping

In her work responding to natural disasters in Nepal, Pavitra Rana [found that](#) “the full potential of information to aid disaster response largely rests on how much information is available and accessible, in what format, through which channel and how well it is shared... open data could be an effective tool to improve access to a timely flow of information, helping to effectively and efficiently deliver relief assistance.”

In particular, community mapping, whereby a community is engaged to gather data on the built environment through crowdsourcing, has had a significant impact on the way cities respond to natural disasters. According to The World Bank “advances in technology — including rising communications access, falling hardware costs, and a growing movement toward open data, open source, and open innovation — have created a new opportunity to engage local communities.”²⁶²

HumanitarianResponse.info²⁶³, a platform developed by the UN Office for the Coordination of Humanitarian Affairs²⁶⁴, was used in the aftermath of Cyclone Pam, the most intense tropical cyclone in the southern hemisphere in 2015. It helped responders to coordinate their work in Vanuatu - government data on roads, accessibility and population numbers was enhanced with data gathered as part of NGOs’ development work on the ground to assess population distributions and plan targeted responses.²⁶⁵

The Humanitarian OpenStreetMap Team (HOT)²⁶⁶ was incorporated in the immediate aftermath of the Haiti earthquake 2010 and became a registered charitable organisation in 2013. HOT creates open data remotely to respond to humanitarian disasters around the world. Using crowdsourcing to very quickly build detailed maps of affected areas, published as open data, it enables response teams to deliver relief that they may have otherwise been unable to and reach those in critical need. As well as responding to Cyclone Pam in Vanuatu, HOT has also worked in the aftermath of a number of other extreme weather events, including the Tharparkar Drought in Pakistan, heavy flooding in Paraguay and Typhoon Haiyan in the Philippines in 2014 alone.

Opening up relief efforts

In addition to mapping, open data has a significant role to play in other areas of response, recovery and reconstruction.

For instance, if governments and NGO’s alike were to publish more data regarding their relief efforts in affected areas (such as types of assistance provided, quantities and types of materials distributed, numbers of assisted persons, etc) it would be much more readily available to other responders. It could then be used by the responders to help identify need and target their own relief operations - such as in areas initially deprived of assistance. It can also help to reduce the duplication of relief efforts.

The publication of detailed relief data increases the transparency and accountability of the assistance itself. Analysis of the response to a disaster, such as one caused by an extreme weather event, can

²⁶² <http://blogs.worldbank.org/ic4d/ideathon-code-resilience-mixes-tech-and-disaster-risk-experts-spark-innovation> (last accessed on 2015-10-16)

²⁶³ <https://www.humanitarianresponse.info/home> (last accessed on 2015-10-16)

²⁶⁴ <http://www.unocha.org/> (last accessed on 2015-10-16)

²⁶⁵ <https://www.devex.com/news/data-in-action-the-role-of-data-in-humanitarian-disasters-86565> (last accessed on 2015-10-16)

²⁶⁶ <http://hotosm.org/> (last accessed on 2015-10-16)

enable citizens to determine whether it was effective and may help to plan and improve future responses.

Open data related to a city's demographics can also prove extremely valuable in the immediate aftermath of an extreme weather event. It is used by relief responders to determine the varying requirements of different areas, prioritise assistance and prepare tailored relief.

When it comes to the reconstruction of a city hit by an extreme weather event, citizens can often be left in the dark concerning what, how and when they'll receive the allocated funds for rebuilding their homes and businesses. Pavitra Rana argues that "open data can play a role in improving [citizens access to reconstruction information]. Extracting all the information on financial relief packages from different government documents, sorting it to ensure cohesiveness and easy understanding, and publishing it in an open format could help people to get complete information from a single source."²⁶⁷

Open data does not only have a role to play in improving response, recovery and reconstruction to extreme weather events in the developing world. In December 2014, the US government launched a new open data portal to meet the specific needs of responders to disaster situations²⁶⁸. The portal provides public access to a number of datasets related to earthquakes, floods, hurricanes, severe winter weather, tornadoes and wildfires, as well as tools and applications that can be used in the event of these extreme weather disasters.

4.3 How can the DataGraft platform help?

The DataGraft platform could be of clear benefit in addressing the challenge of improving the resilience of cities to extreme weather events in both the understanding of risk and response, recovery and reconstruction settings.

4.3.1 Making it easier to publish data with limited resources

Critically, with regard to developing an improved understanding of the risks extreme weather may pose, DataGraft may enable a wider range of stakeholders to publish their data openly. As discussed, the data required to develop an effective risk information and monitoring system is often held by a number of different parties - not only the government or city itself. Equipping these parties, which may not otherwise possess the resources of government or city departments to publish data, with the DataGraft tooling that helps them to release high quality, usable open data with minimal resources, could lead to the improved efficacy of these systems.

DataGraft can also be used as a centralised, low-cost hosting platform for the open data itself. This further increases the prospect of more stakeholders publishing their data openly, as it removes another significant barrier to doing so.

The fluctuating nature of the risks posed by extreme weather makes the availability of real time data (where possible) related to them extremely useful. The platform's ability to publish and handle real time data may prove a significant feature in the development of risk information and monitoring capabilities. It also allows data publishers to save and repeat functions carried out on one dataset to others - enabling them to near automate the process of preparing real time data.

²⁶⁷ <http://opennepal.net/blog/challenges-using-information-disaster-response-how-can-open-data-help> (last accessed on 2015-10-16)

²⁶⁸ <http://www.data.gov/disasters/> (last accessed on 2015-10-16)

4.3.2 Enabling the publishing of interoperable linked data

In the response, recovery and reconstruction setting, the DataGraft platform's focus on producing interoperable open data to common standards is particularly useful.

As Pavitra Rana explains "a wide range of information is necessary for the delivery of effective relief operations aimed at reducing suffering in the wake of disasters. The availability of this information needs to be improved by publishing it in easier to access and reusable formats."²⁶⁹

Publishing relief assistance data using the platform's standard workflow methodology and supportive technical tools would help to ensure that open data published by different stakeholders - such as NGOs and cities themselves - can be aggregated with other datasets. Again, for stakeholders with critical data but limited resources, DataGraft may be particularly useful, as it reduces the burden of preparing high quality, interoperable open data.

An in-depth study²⁷⁰ into the data behind disaster management found that although crowdsourcing - resulting in community generated datasets - has had crucial impacts around the world (such as in Nepal and Haiti), its unstructured nature can often limit its utility through problems with processing and integration. In particular, relief organisations working in disaster-struck areas are unable to cope with unstructured data as they simply lack the time to integrate it into existing information systems. It suggests that linked open data practice could significantly increase the effectiveness of these efforts.

It prescribes "linked open data to alleviate the integration problems of crowdsourced data and to improve the exploitation of crowdsourced data in disaster management. We suggest to engage people in processing unstructured observations into structured RDF to linked open data principles... This will increase the impact of crowdsourced data in disaster management and help humanitarian agencies to make informed decisions."²⁷¹

Whilst acknowledging the lack of resource available to integrate unstructured data, especially in highly time sensitive operations, what is perhaps foregone is an appreciation of the resources required to prepare RDF linked data until now. DataGraft's ability to make it relatively easy for users to publish to this standard will enable their data to be powerfully connected with other datasets, and as the study describes, unlock further value in the response to extreme weather events.

²⁶⁹ <http://opennepal.net/blog/challenges-using-information-disaster-response-how-can-open-data-help> (last accessed on 2015-10-16)

²⁷⁰ <http://ceur-ws.org/Vol-798/paper2.pdf> (last accessed on 2015-10-26).

²⁷¹ Ibid.

5 Bringing new drugs to a global market

5.1 What is the challenge?

The average cost of bringing a new drug to market is now estimated to be \$2.6bn.²⁷² The ever increasing complexity of pharmaceutical Research and Development (R&D), combined with an inability of new drugs to capture the huge reimbursement value of their 'blockbuster' counterparts of previous years, makes bringing new drugs to market an arduous challenge.

Between 2011 and 2016 more than \$125bn worth of brand drug sales will have lost patent protection - a process known as the 'patent cliff'. These are generally blockbuster drugs that account for a large percentage of big pharmaceutical company sales. In 2015 alone potential branded drug losses will reach an estimated \$33.5 billion²⁷³, as generic drug companies become free to create and sell their own replicas. This will increase pressure on many pharmaceutical companies to develop new and clinically differentiated products, which require the strong focus on R&D and innovation they have often been accused of lacking in recent years.

According to Rizwan Ahmed, "the major changes [in the pharmaceutical sector] will come in the R&D [approach] of companies. Innovation in drug discovery over the last decade has been considerably slower than before, and the necessity for a differentiable drug is imperative for sustained success among large pharmaceutical companies."²⁷⁴

The limited duration of patent protection may also put pressure on pharmaceutical companies to accelerate their R&D processes (and thus get their drugs to market more quickly) in order to have a longer period in which to market them.

Once a new drug is brought to market, the development of health technology assessment (HTA) practices across various stakeholders can make it even harder for pharmaceutical companies to achieve high reimbursement value for newly launched products. This is driven by the trend towards healthcare reimbursement being based on the demonstration of real world effectiveness, in addition to clinical trial efficacy.

This pressure on reimbursement compounds the high R&D costs already associated with bringing a new drug to market. As PWC have reported, "the sober reality [for the pharmaceutical sector] could be a continuously declining return on each dollar invested in R&D. To escape this vicious circle, a fundamental overhaul of the prevailing pharmaceutical R&D model is required... pharmaceutical companies need to put into practice an integrated evidence development model that seamlessly brings together randomized controlled trials and Real World Evidence (RWE)-based approaches."²⁷⁵

Open data will play an increasingly important role in both the clinical trials process and real world evidence approaches to bringing a new drug to market - helping the pharmaceutical sector to shift from the inward-facing, partially shared R&D environment to a more collaborative, distributed and open approach.

²⁷² <http://www.scientificamerican.com/article/cost-to-develop-new-pharmaceutical-drug-now-exceeds-2-5b/> (last accessed on 2015-10-16)

²⁷³ <http://www.michaelbaileyassociates.com/news/pharmaceutical/what-do-patent-expirations-mean-for-the-pharma-industry> (last accessed on 2015-10-16)

²⁷⁴ <http://triplehelixblog.com/2014/07/the-patent-cliff-implications-for-the-pharmaceutical-industry/> (last accessed on 2015-10-16)

²⁷⁵ <http://www.strategyand.pwc.com/global/home/what-we-think/reports-white-papers/article-display/revitalizing-pharmaceutical-> (last accessed on 2015-10-16)

5.2 How can we address the challenge using data?

5.2.1 Clinical trials data

A clinical trial is designed to collect data on the safety and efficacy of a healthcare intervention, such as a new drug. Clinical trials, and their results, form a significant part of the authorisation process a pharmaceutical company must undertake to bring a newly developed product to market.

Over the past decade there has been a growing demand for the clinical trials process to become more open. In part, this has been fuelled by the The Declaration of Helsinki²⁷⁶ - the World Medical Association's (WMA)²⁷⁷ statement of principles for medical research involving people, which states that every investigator running a clinical trial should register it and report its results.

In some regions, public registration of a clinical trial has become law when bringing a new drug to market and numerous registries have been developed to collect key information related to them. The World Health Organisation (WHO)²⁷⁸ International Clinical Trials Registry Platform (ICTRP)²⁷⁹ brings together trials registered on a number of national and multinational registries, including the EU Clinical Trials Register (EU-CTR)²⁸⁰. It aims to provide a single public access point to trial registry information to enable the unambiguous identification of clinical trials.

The call for clinical trials to become more open extends beyond just their registration, however. In April 2015 the WHO published a new statement²⁸¹ on the public disclosure of clinical trial results, calling for results-reporting of older but still unpublished trials, and outlining steps to improve linkages between clinical trial registry entries and their published results. The statement added to the widespread and increasing pressure on pharmaceutical companies to disclose more information related to the results of their clinical trials.

Clinical trial results

According to AllTrials²⁸², there are generally three levels of information involved in clinical trial results reporting: a summary of the trial's results, full clinical study reports about the trial, and patient and study-level data from the trial.²⁸³

Summary results from a clinical study include data on the primary and any secondary outcomes of taking a drug and their statistical analysis. Although regulations vary by region, some current (and most proposed) regulatory frameworks suggest that summary results for every registered trial must be made public within one year of the trial's completion, although there is debate surrounding the extent to which many pharmaceutical companies comply.²⁸⁴ When summary results are produced, they are often published in a variety of formats, and submitted and published in different ways; across journal papers and reports to grant giving bodies, for example.

Full clinical study reports contain a large amount of data on the methods, analysis, results and conclusions of a clinical trial. Pharmaceutical companies bringing a new drug to market must produce a clinical study report in order to receive marketing authorisation, following a standard structure set out by

²⁷⁶ <http://www.wma.net/en/30publications/10policies/b3/> (last accessed on 2015-10-16)

²⁷⁷ <http://www.wma.net/> (last accessed on 2015-10-16)

²⁷⁸ <http://www.who.int/> (last accessed on 2015-10-16)

²⁷⁹ <http://www.who.int/ictip/> (last accessed on 2015-10-16)

²⁸⁰ <https://www.clinicaltrialsregister.eu/> (last accessed on 2015-10-16)

²⁸¹ http://www.who.int/ictip/results/WHO_Statement_results_reporting_clinical_trials.pdf?ua=1 (last accessed on 2015-10-16)

²⁸² <http://www.alltrials.net/> (last accessed on 2015-10-16)

²⁸³ <http://www.alltrials.net/find-out-more/all-trials/> (last accessed on 2015-10-16)

²⁸⁴ <http://www.bmj.com/content/344/bmj.d7373> (last accessed on 2015-10-16)

International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH)²⁸⁵ Good Clinical Practice guidelines²⁸⁶. There are different approaches to the release of clinical study reports, with some regulators/agencies adopting a policy of releasing reports it holds on request and many pharmaceutical companies themselves choosing to share reports with selected researchers.

Suitably anonymized patient and study-level data, which does not allow for the identification of trial participants, is sometimes made available when the clinical study report is shared. A number of pharmaceutical companies have agreed to share this type of trial data with selected researchers via the Clinical Study Data Request platform²⁸⁷.

McKinsey have reported that "some pharmaceutical companies start to improve collaboration by identifying data elements to share with specific sets of trusted partners. Such steps are only the beginning, however, as they are essentially just a way to expand the 'circle of trust' to select partners."²⁸⁸

Rather than sharing only with select partners, pharmaceutical company GlaxoSmithKline (GSK)²⁸⁹ has adopted a more open approach to publishing their clinical study reports. The detailed reports are redacted for any patient data (that could identify trial participants) and are made available to the public. The company's approach to clinical study reports - shifting them along the data spectrum²⁹⁰ from closed to shared, and now towards open - is leading the way in the sector.

The benefits of access to clinical trial information

The benefits of making more clinical trial information publicly available could be wide reaching.

Secondary analyses can change clinical practice by demonstrating ineffective, unsafe or other uses of a drug. The pharmaceutical company Lilly²⁹¹ recently announced an apparent breakthrough in the treatment of Alzheimer's disease²⁹². The new evidence of the effectiveness of the drug solanezumab is based on an extension of a previous phase III trial that had been declared a failure in 2012. Fresh analytics of the data revealed a 34% reduction in the rate of mental decline among a subset of mild Alzheimer's patients, suggesting solanezumab could have an effect if prescribed before the disease became too advanced. Although the secondary analysis was internal, this demonstrates the potential benefit of making clinical trial data more widely available.

Sir Tim Berners-Lee, co-founder of the Open Data Institute²⁹³, has recently called for clinical trials data to be open by default.²⁹⁴ He said that "if you have access to, for example, a history of all the clinical trials people have done with different compounds on different patients, you can find yourself in a situation where you want to use a particular drug in conjunction [with a disease] and you can understand a whole lot about it just by running back over old clinical trials [data]."

Another key advantage of making clinical trial data more widely available is the elimination of unnecessary duplicative efforts among pharmaceutical companies, common when the data exists in sealed

²⁸⁵ <http://www.ich.org/home.html> (last accessed on 2015-10-16)

²⁸⁶ <https://www.gov.uk/guidance/good-clinical-practice-for-clinical-trials> (last accessed on 2015-10-16)

²⁸⁷ <https://www.clinicalstudydatarequest.com/> (last accessed on 2015-10-16)

²⁸⁸ http://www.mckinsey.com/insights/health_systems_and_services/how_big_data_can_revolutionize_pharmaceutical_r_and_d (last accessed on 2015-10-16)

²⁸⁹ <http://www.gsk.com/en-gb/home/> (last accessed on 2015-10-16)

²⁹⁰ <http://theodi.org/data-spectrum> (last accessed on 2015-10-16)

²⁹¹ <http://www.lilly.co.uk/en/index.aspx> (last accessed on 2015-10-16)

²⁹² <http://www.pri.org/stories/2015-07-27/new-drug-being-tested-raises-hopes-people-alzheimers> (last accessed on 2015-10-16)

²⁹³ <http://theodi.org/> (last accessed on 2015-10-16)

²⁹⁴ <http://www.v3.co.uk/v3-uk/news/2417313/tim-berners-lee-calls-for-healthcare-data-to-be-open-by-default> (last accessed on 2015-10-16)

corporate silos. Breaking down these silos could therefore lead to substantial cost savings, as the data generated by one study can be ingested and used to inform another. Brian L. Strom, chancellor of Rutgers Biomedical and Health Sciences²⁹⁵, has said that “data collection is the most expensive element in the development of new drugs, and the most risky to subjects. Data sharing allows this valuable resource to be used maximally.”²⁹⁶ Rather than shared data, it could, in fact, be that an open data approach ensures that clinical trial data is used to its full potential - users of the data are not bound, in number or by the terms of use, of a data sharing agreement.

Other potential cost-saving benefits include the accurate, data driven identification of opportunities in the market, and reduced exposure to patient safety risks.

This shift to an increasingly open approach - and a growing regard for clinical trial data as a pre-competitive element in bringing a new drug to market - is consistent with PWC's observation that “many leading pharmaceutical companies have started initiatives to incrementally improve R&D productivity... experimenting with pre-competitive consortia and partnerships.”²⁹⁷ Proposed EU Clinical Trials Regulation will likely seek to push this boundary further. It is expected to contain guidance that no information in a clinical study report should be considered commercially confidential once a marketing authorisation decision has been made.

The first phase of Open Trials²⁹⁸, an open database of information about the world's clinical research trials, is set to be completed in March 2017. A partnership between Open Knowledge²⁹⁹ and the Center for Open Science³⁰⁰, it will match trial registry data with documents containing trial results to identify results that have not been disclosed.

In July 2015 the non-profit foundation Medicines for Malaria Venture (MMV)³⁰¹ won an Open Data Innovation award for their Malaria Box³⁰². The Malaria Box is a collection of 400 compounds, which are provided free of charge to researchers who wish to develop new medicines for malaria - on the basis that their results are published in the public domain (as open data).

These are significant examples of this shift in the way clinical trials are regarded and to a more open approach to pharmaceutical R&D: the data from clinical trials is becoming a foundation for the industry to innovate on top of, rather than locked away in silo.

5.2.2 Real world data

Real world data is the term used for data collected outside of traditional controlled clinical trials, under real-life (practical) circumstances. In the process of bringing a new drug to market, this type of data is used to demonstrate real world evidence of its effectiveness.

According to the Association of the British Pharmaceutical Industry (ABPI)³⁰³ “extrapolating results from clinical trials, for which patients are often highly selected for age, comorbidity (the existence of

²⁹⁵ <http://rbhs.rutgers.edu/> (last accessed on 2015-10-16)

²⁹⁶ <http://news.rutgers.edu/qa/data-sharing-pharmaceutical-industry-shows-progress/20141015#.ViZxPhCrTdS> (last accessed on 2015-10-16)

²⁹⁷ <http://www.strategyand.pwc.com/global/home/what-we-think/reports-white-papers/article-display/revitalizing-pharmaceutical-rd> (last accessed on 2015-10-16)

²⁹⁸ <http://blog.okfn.org/2015/04/21/open-trials-open-knowledge-announce-plans-for-open-online-database-of-clinical-trials/#sthash.e4tgU9XR.dpuf> (last accessed on 2015-10-16)

²⁹⁹ <https://okfn.org/> (last accessed on 2015-10-16)

³⁰⁰ <http://centerforopenscience.org/> (last accessed on 2015-10-16)

³⁰¹ <http://www.mmv.org/> (last accessed on 2015-10-16)

³⁰² <http://www.mmv.org/malariabox> (last accessed on 2015-10-16)

³⁰³ <http://www.abpi.org.uk/> (last accessed on 2015-10-16)

other diseases or disorders) and performance status, to clinical practice can be very challenging and restrictive... real world data can reduce the uncertainty exhibited by new medicines at the time of launch by improving the information on the benefits and risks and developing evidence on real-life effectiveness.”³⁰⁴

Real world data includes data from electronic medical records and observational studies, and also extends to data that isn't directly related to drugs, such as population surveys, national statistics and social media data. It is now used in both the pre- and post-approval settings for new drugs.

In the pre-approval setting, real world data can be used to understand market characteristics such as unmet clinical need, pathways of care, current treatment patterns and patient profiles.

After being brought to market - and in some cases before - the use of real world data to evaluate a new drug's effectiveness and safety can provide practical insight into numerous indicators; such as the impact and clinical outcomes of treatment, prescribing patterns, the impact of service delivery, patient experience and patient safety.

The availability of real world data

Some real world data used by pharmaceutical companies to bring a new drug to market is already published as open data, but this is generally limited to populations surveys and national statistics. Often, access to the most valuable real world data used in bringing drugs to market is restricted.

Using data from electronic medical records and patient registries to generate real world evidence of a drug's effectiveness is one of the fastest growing areas of pharmaceutical R&D. Modern healthcare practice generates huge amounts of data related to patient outcomes and trends, and this is only set to increase. In the UK, for example, “it is likely that the NHS will be collecting and using more real world data with which to make local decisions about healthcare delivery in the future” according to the ABPI.

How much of this data is, and should be, made available to the pharmaceutical industry is a topic of wide debate. A key component of the discussion is a patient's right to privacy and the very personal and sensitive nature of their electronic medical records.

Deloitte³⁰⁵ have stated that “due to a great deal of recent development in the usage of real world evidence, historic regulatory paradigms are becoming obsolete. Without a clear framework, market participants must carefully coordinate their approach... as all parties seek to achieve both public confidence in the privacy of their data and to realise the full potential of healthcare data to transform patient outcomes.”³⁰⁶

Currently, aggregated statistics from electronic medical records are sometimes shared with pharmaceutical companies as a result of specific partnerships or agreements. Nadia Foskett, Principal Real World Data Scientist at Roche³⁰⁷, recently delivered a presentation on the value of using real world data in pharma R&D at an Open Data Institute business breakfast. She described how Roche was understood to be the only pharmaceutical company to have access to aggregated statistics from the UK's Systemic Anti-Cancer Therapy (SACT) chemotherapy dataset. The dataset is a real world drug

³⁰⁴ <http://www.abpi.org.uk/our-work/library/industry/Documents/Vision-for-Real-World-Data.pdf> (last accessed on 2015-10-16)

³⁰⁵ <http://www2.deloitte.com/uk/en.html> (last accessed on 2015-10-16)

³⁰⁶ <http://www2.deloitte.com/uk/en/pages/life-sciences-and-healthcare/articles/real-world-evidence.html> (last accessed on 2015-10-16)

³⁰⁷ <https://www.roche.co.uk/> (last accessed on 2015-10-16)

registry that maintains data on around 200,000 patients in the UK who have received chemotherapy - which can be extremely valuable in evaluating the effectiveness of the drugs used in the treatment.

In order to collect, analyse and share real world data, many pharmaceutical companies have developed proprietary data networks based on exclusive agreements or partnerships, such as the one involving Roche and SACT. The varied way in which healthcare is delivered - and data collected - by nation and healthcare system means that there is no single model for this type of agreement. One characteristic common across nations and systems, however, is a large number of different stakeholders involved; pharmaceutical companies often have to work with a variety of them - from healthcare providers to insurance companies - in order to access this type of data.

Some are now asking whether more of this aggregated or anonymised real world data can be released as open data, rather than shared between select stakeholders. However, the issues surrounding privacy and control of data in this field are significant. It should be noted that there are risks to anonymisation and some argue that the security of personal records cannot be guaranteed through anonymisation procedures.³⁰⁸ The wider healthcare industry will need to work together to ensure that patients maintain their rights to privacy.

Using real world data

The availability of more real world data could have significant benefits to the pharmaceutical industry.

For those pharmaceutical companies currently engaged, and seeking, agreements or partnerships to access real world data, it would reduce the substantial financial burden of doing this. It would also open doors for smaller, or less connected, firms; particularly in the pre-approval setting. "Open data initiatives provide local researchers with more avenues to get involved in what previously may have been considered high-level, niche or expensive drug discovery research,"³⁰⁹ says Michelle Willmers, project manager of Open Data Africa Initiative.³¹⁰

The significant benefit to pharmaceutical companies of an increased availability of real word data is the sustained ability to monetise the drugs they bring to market. Using real world data to demonstrate the effectiveness of a new drug is now critical to successful, long term reimbursement. It is needed to satisfy the changing reimbursement environment, in which payments are increasingly linked to the demonstration of the real world impact of the medication. It is not only regulatory bodies that are demanding more long-term data demonstrating the effectiveness, safety and quality of a new drug. ABPI have said that "realistically, key drivers for this breadth of information are the demands of health technology assessors and payers, and their need for evidence-based health economic data, based on local conditions."³¹¹ Better access to the data needed to demonstrate the effectiveness of their drugs will enable pharmaceutical companies to meet these demands and help to ensure value capture.

According to PwC, "if pharmaceutical companies fail to build an effective RWE-based capabilities system, they are at risk of quickly losing control over the value communication around their own drugs, as other stakeholders such as payors, data analytics companies, and academia are currently enhancing their own capabilities. In consequence, this might potentially even lead to a significant decline in use and reimbursement."³¹²

³⁰⁸ <http://www.infosecurity-magazine.com/news/cambridge-professor-questions-the-viability-of/> (last accessed on 2015-10-16)

³⁰⁹ <http://www.scidev.net/sub-saharan-africa/data/news/open-data-key-to-tackling-neglected-tropical-diseases.html#sthash.Zp7WslKW.dpuf> (last accessed on 2015-10-16)

³¹⁰ <http://opendataforafrica.org/> (last accessed on 2015-10-16)

³¹¹ <http://www.abpi.org.uk/our-work/library/industry/Documents/Vision-for-Real-World-Data.pdf> (last accessed on 2015-10-16)

³¹² <http://www.strategyand.pwc.com/global/home/what-we-think/reports-white-papers/article-display/revitalizing-pharmaceutical-rd> (last accessed on 2015-10-16)

More widespread access to valuable real world data could also have a huge impact on the cost of bringing a new drug to market. It would fuel the industry's shift in R&D approach to an integrated clinical trials/real world evidence model - in which a randomised registry (or other pragmatic trial format) replaces the controlled trial approach at the end of phase IIa of bringing a new drug to market. This is provided of course that a real world data project can be carried out outside of the clinical trial setting within the current regulatory environment.

PwC have found that “[the shift], if handled with robust and reliable guidance, would lead to earlier generation of clinical effectiveness data... It would enable the sponsor to collect precious clinical effectiveness data right away. It would also result, five years later, in meaningful evidence of a drug's risk/benefit and cost benefit profile that is in line with payors' expectations. Finally, it would protect the sponsor from big safety surprises in phase III programs. In short, the fully integrated RCT/RWE model would lead to results with higher external validity and generalized applicability earlier.”³¹³

This earlier discovery of a drug's effectiveness (or lack thereof) can have a critical effect on pharmaceutical R&D costs; by enabling an effective drug to be brought to market more quickly, and by providing researchers with the evidence they need to halt or reduce investment earlier in a failing drug. It is estimated that this approach could reduce cycle time by around 5 years and decrease R&D investment per product by up to 60%.³¹⁴

In summary, widespread access to real word data could have a dramatic impact on a pharmaceutical company's ability to both capture the value of a new drug brought to market, and reduce the huge costs of bringing it there.

5.3 How can the DataGraft platform help?

There are two key ways in which the DataGraft platform can help address the challenge of bringing a new drug to market.

5.3.1 Publishing clinical trial data effectively

Even as the shift to more clinical trial data being made openly available gathers momentum, there are concerns that the benefits derived from it will be hampered by a lack of publishing standards. Although the platform cannot alleviate overnight the pharmaceutical industry's ongoing issues with ontology - different stakeholders describe the same, or similar, data very differently - it can help to ensure that data is published on clinical trial registers in a way that makes it most useful to others.

According to AllTrials, “registers currently have different formats for reporting results. Ideally every major register could require results to be uploaded in a format that allows the main reported items to be searchable and enables sharing of information between registries... Global registries would certainly have to do this to be useful and manageable.”³¹⁵

One of the DataGraft platform's primary tasks is to facilitate the publication of interoperable data. In this case, publishing clinical trial data using its standard workflow methodology and supportive technical tools would help to ensure that the data made available by different publishers is workable when

³¹³ <http://www.strategyand.pwc.com/global/home/what-we-think/reports-white-papers/article-display/revitalizing-pharmaceutical-rd> (last accessed on 2015-10-16)

³¹⁴ Ibid

³¹⁵ <http://www.alltrials.net/find-out-more/all-trials/> (last accessed on 2015-10-16)

mixing datasets together. It provides a set of tools which help publishers to clean, transform and link the data. For smaller pharmaceutical firms with very limited resources, DataGraft may be particularly useful - it can help to reduce the time and resource burden for preparing high quality, usable open data.

DataGraft's focus on helping users to publish in linked data (RDF) formats may also be significant in helping open clinical trial data to be published and used effectively. A commonly held opinion among stakeholders is that having data that is consistent, reliable, and well linked is one of the biggest challenges facing pharmaceutical R&D. Linked data is a method of data publishing that enables it to be powerfully connected with other datasets, to derive detailed insight or value. Encouraging clinical trial data to be published in linked data formats where possible could go a long way in enabling the discussed benefits of more widely available data to be unlocked.

5.3.2 Exploiting the value of real world data

Secondly, the DataGraft platform's technical features could be used to ingest and derive value from the rising quantities of available real world data. This is a core challenge to pharmaceuticals companies, as it can often be extremely challenging to integrate external data from a wide range of sources and exploit it. This is often referred to as 'out-side in' innovation - the incorporation of external knowledge into a business' own commercial processes.

Regardless of whether the data is released openly or subject to a data sharing agreement - which are common in the industry - pharma companies must be able to quickly consume and exploit these data sources, ideally quicker than other indirect competitors such as healthcare providers and data analytics companies, who are rapidly developing advanced HTA practices.

In meeting this challenge, the project's PaaS-layer could be very useful. It offers business users efficient datastore access, data import and transformation services, data enrichment and linking, search and indexing, and caching tools. These are some of the essential services required to tweak and process external data in a way that it fits with the internal demands of an organisation.

The tools DaPaaS provides in reporting the data may also help pharmaceutical companies take control of the value perception of its products. By reporting on the efficacy and effectiveness of their products quickly and simply, it will ensure that they can impart key messaging and be heard among the growing number of voices assessing them.

Ultimately, the effect clinical trial and real world data can have in developing new drugs will depend on the amount of data that is made available. Deloitte argue the need for "a shift in mind-set, moving away from withholding all data and toward identifying which data can be shared and with whom. In addition, pharmaceutical enterprises must understand and mitigate the legal, regulatory, and intellectual-property risks associated with a more collaborative approach."³¹⁶ As this shift occurs, the DaPaaS could help pharmaceutical companies, old and new, to exploit the data and use it to meet the challenge of bringing new drugs to market.

³¹⁶ <http://www2.deloitte.com/uk/en/pages/life-sciences-and-healthcare/articles/real-world-evidence.html> (last accessed on 2015-10-16)

6 Conclusions

This report has explored the potential of the DataGraft platform and tools with respect to specific real world challenges. These real world challenges from a range of domains around health, the environment and smart cities, all major concerns in Europe and beyond. Given the role that data, particularly open and linked data, plays in solving these challenges there are clear potential implementations for the DataGraft platform. The potential is recognised in its ability to maximise the application of data to solutions through high-quality, accessible linked open data. As such each of the use cases demonstrates a tangible potential impact the DataGraft platform could have in tackling real world problems.