Using Big Data to improve the UKWIR's Chemicals Investigation Programme

Juan Manuel Ponce Romero
Cranfield University
http://www.dream-cdt.ac.uk

Using Big Data to improve the UKWIR's Chemicals Investigation Programme

Juan Manuel Ponce Romero

Cranfield University
Newcastle University
University of Cambridge
University of Birmingham

Residential
Commercial
Industrial
Surface runoff

WWTPs are the receptors of these effluents

Large amounts of biosolids are generated

Some pollutants can pose an environmental hazard…

… and also a hazard to humans

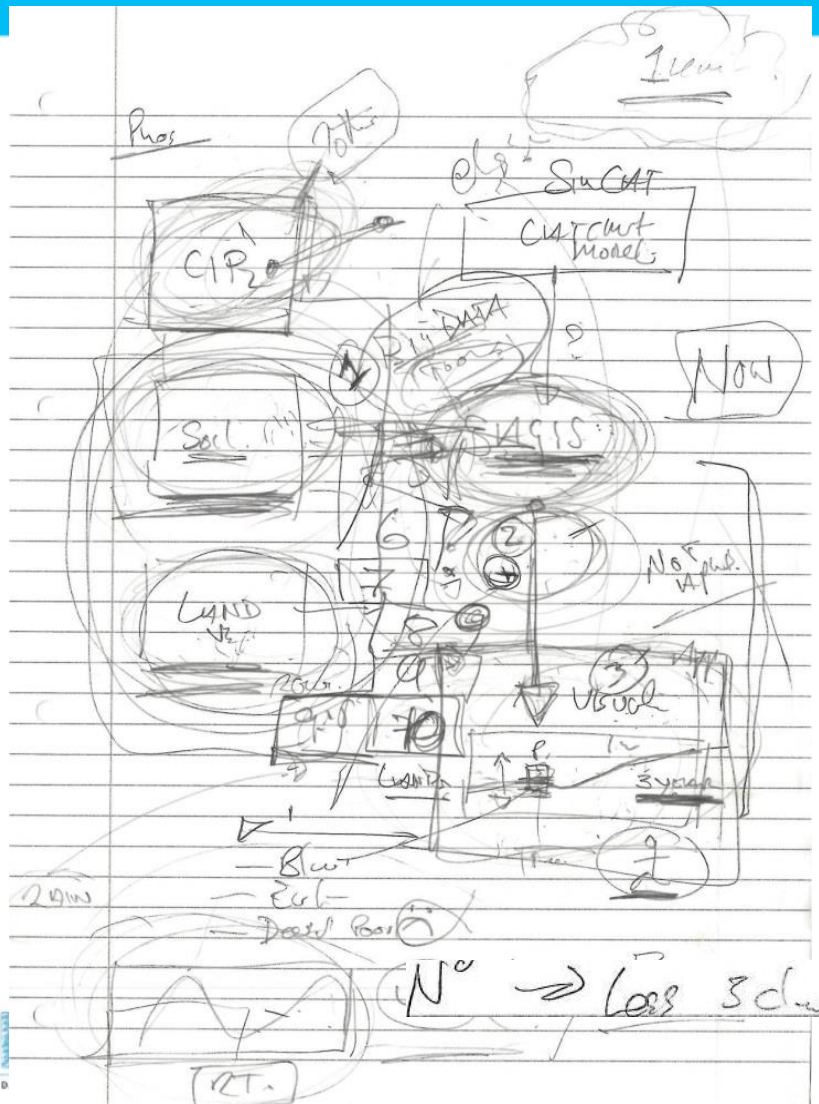
Some substances can be transferred into agricultural land

DREAM
DATA, RISK & ENVIRONMENTAL ANALYTICAL METHODS

Using Big Data to improve the UKWIR's Chemicals Investigation Programme

Juan Manuel Ponce Romero

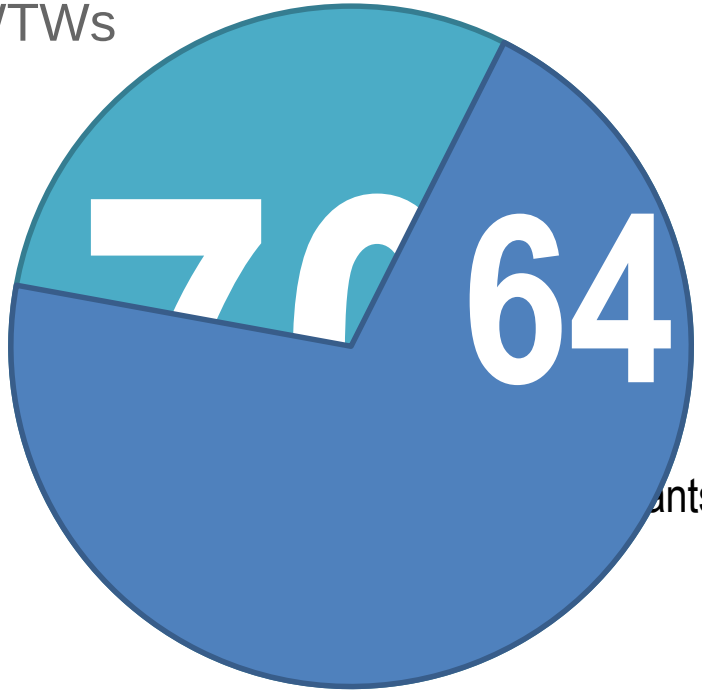
Cranfield University
Newcastle University
University of Cambridge
University of Birmingham



Using Big Data to improve the UKWIR's Chemicals Investigation Programme

Juan Manuel Ponce Romero

- UKWIR
- National Chemicals Investigation Programme (CIP)
- 160 WWTWs
- £25 M



trace contaminants

- Trace metals
- Pharmaceuticals
- Polycyclic aromatic hydrocarbons (PAHs)
- Emerging organic pollutants
- Regulated organic pollutants

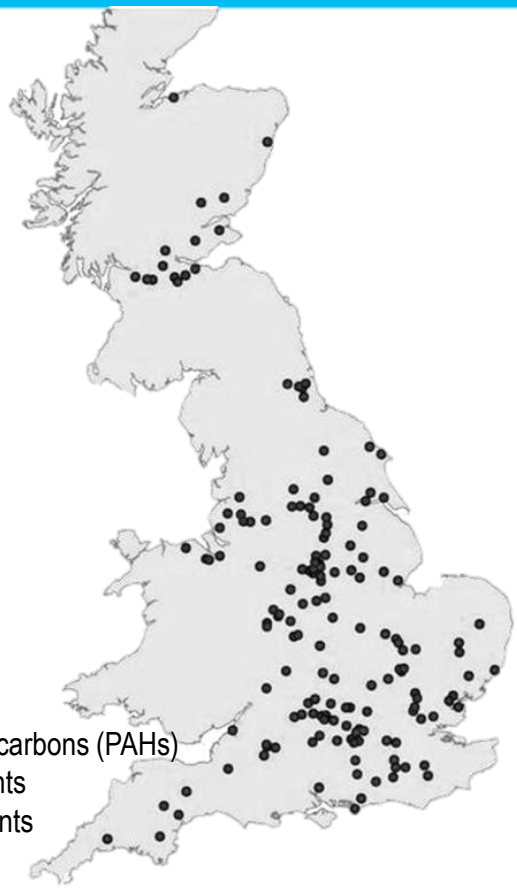


Image: Gardner et al. 2012

DREAM
DATA, RISK & ENVIRONMENTAL ANALYTICAL METHODS

Using Big Data to improve the UKWIR's Chemicals Investigation Programme

Juan Manuel Ponce Romero

Phases of the CIP

Phase I	Phase II	Phase III
• 2010 – 2013 • Identification of relative risk • Prioritisation • Sources	• In progress • Assess WWTPs performance	• Sources investigation

Gardner et al., 2012; UKWIR, 2012

Using Big Data to improve the UKWIR's Chemicals Investigation Programme

Juan Manuel Ponce Romero

Objectives of the CIP:

- Facilitate the **design of** effective, proportionate, and justifiable **control measures**
- Valuable platform for **policy development**
- Quality **technical basis** that companies can use **for approaching future environmental quality requirements**
- Differentiate between measures that can justifiably involve treatment and those that require **other strategies**
- Provide Regulatory Agencies with a demonstrable technical rationale by which **justify** its **proposed actions**

Using Big Data to improve the UKWIR's Chemicals Investigation Programme

Juan Manuel Ponce Romero

What the CIP has demonstrated so far:

- Trace contaminants concentrations in WWTWs' effluents can exceed existing or proposed EQS values
- For the majority of the hazardous chemicals, tenfold dilution in the receiving water will ensure compliance with EQS
- In some cases, the dilution is not enough to comply with EQS. Additional management options to be considered (source control, substance substitution, tertiary treatment, optimisation of processes…)

Gardner et al., 2012

Using Big Data to improve the UKWIR's Chemicals Investigation Programme

Juan Manuel Ponce Romero

Gaps of the CIP (I):

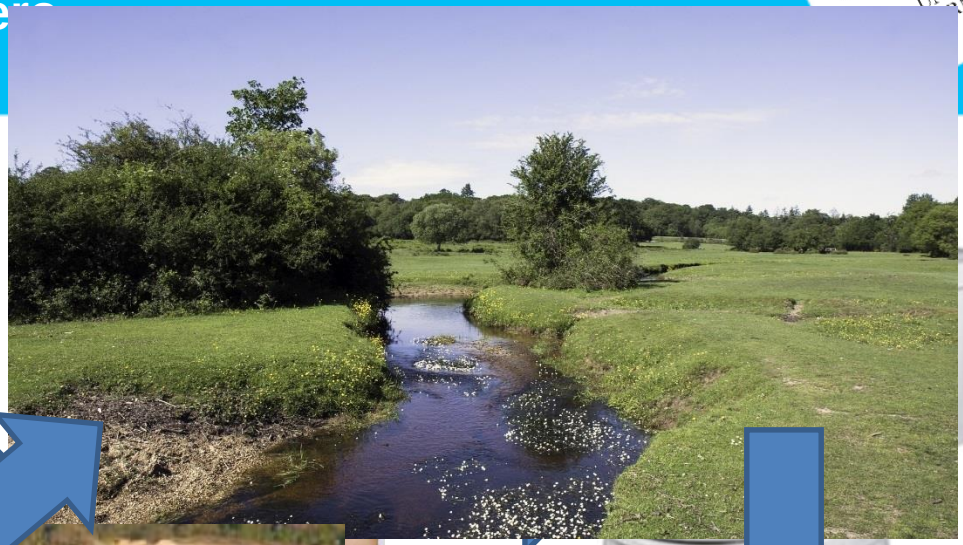
- Large and diverse measurements
- Lack of context for the measures
- Unknown repercussion of the presence of some of the chemicals
- Influence of a chemical or a group of chemicals on others
- Data generated in 2011
- *Can we extract more valuable information from it?*

Gardner et al., 2012

Using Big Data to improve the UKWIR's Chemicals Investigation Programme

Juan Manuel Ponce Romero

Just an example…

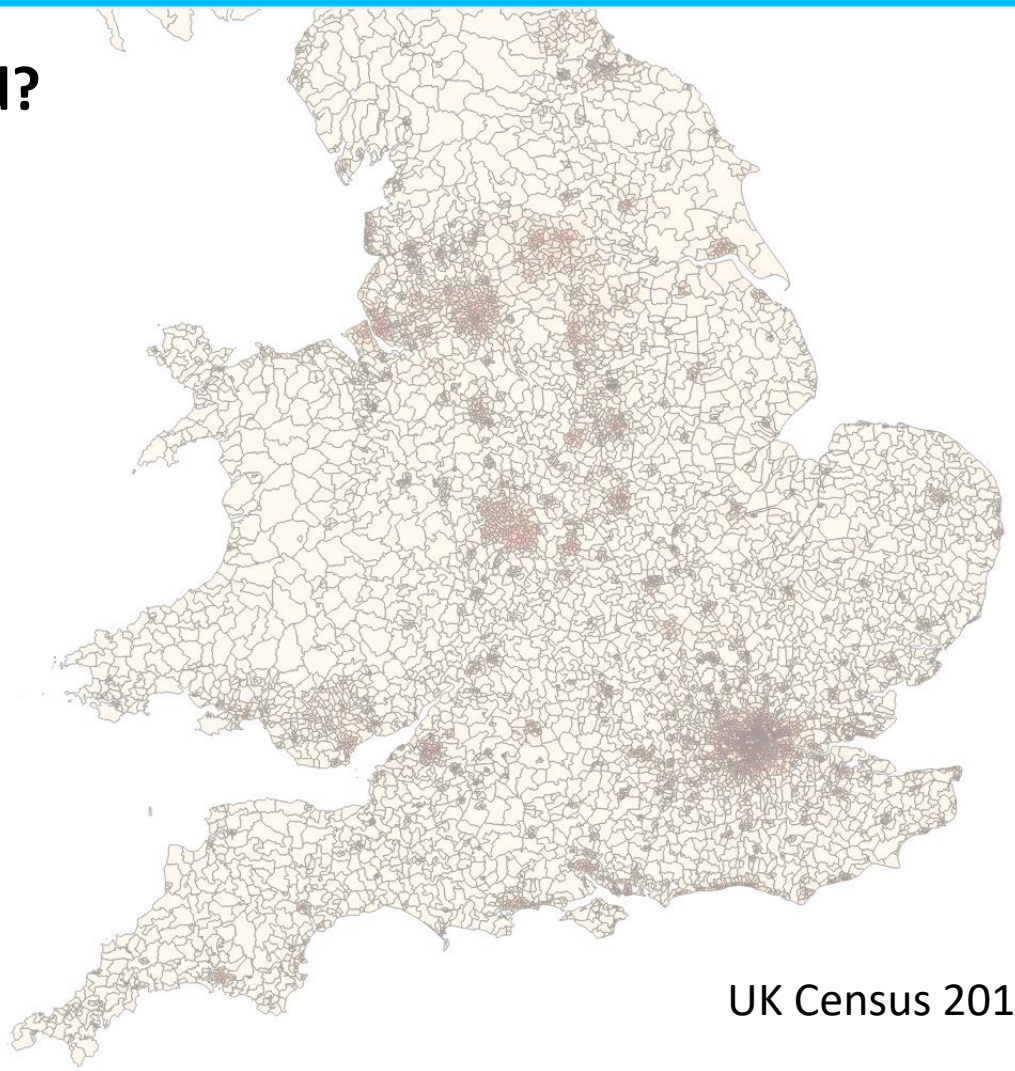


Using Big Data to improve the UKWIR's Chemicals Investigation Programme

Juan Manuel Ponce Romero

What information do we need?

- Female population proportion?
- Age of the female population?
- Other factors?



UK Census 2011

DREAM
DATA, RISK & ENVIRONMENTAL ANALYTICAL METHODS

Using Big Data to improve the UKWIR's Chemicals Investigation Programme

Juan Manuel Ponce Romero

Higher velocity in the transfer or the analytics allows a better response and decision making, in real time

Improving data veracity provides a more accurate and consistent information

Big Data gives the ability to extract more value from data analysis at higher volumes, velocities, varieties, or veracities.

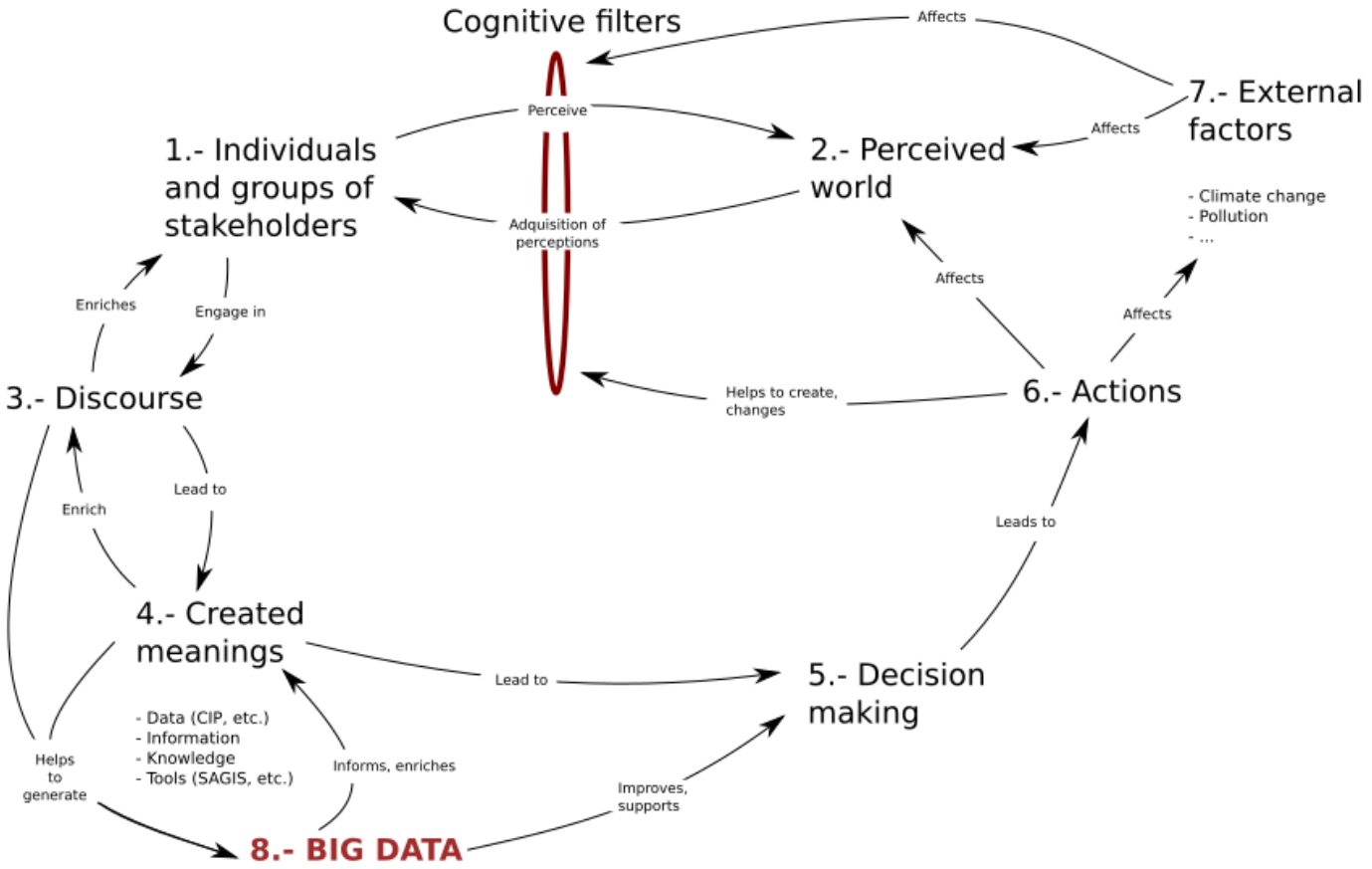
Higher amounts of information allows a more holistic understanding

The use of different types of information allows a broader understanding of the subject

http://www.ibmbigdatahub.com/infographic/extracting-business-value-4-vs-big-data

Using Big Data to improve the UKWIR's Chemicals Investigation Programme

Juan Manuel Ponce Romero



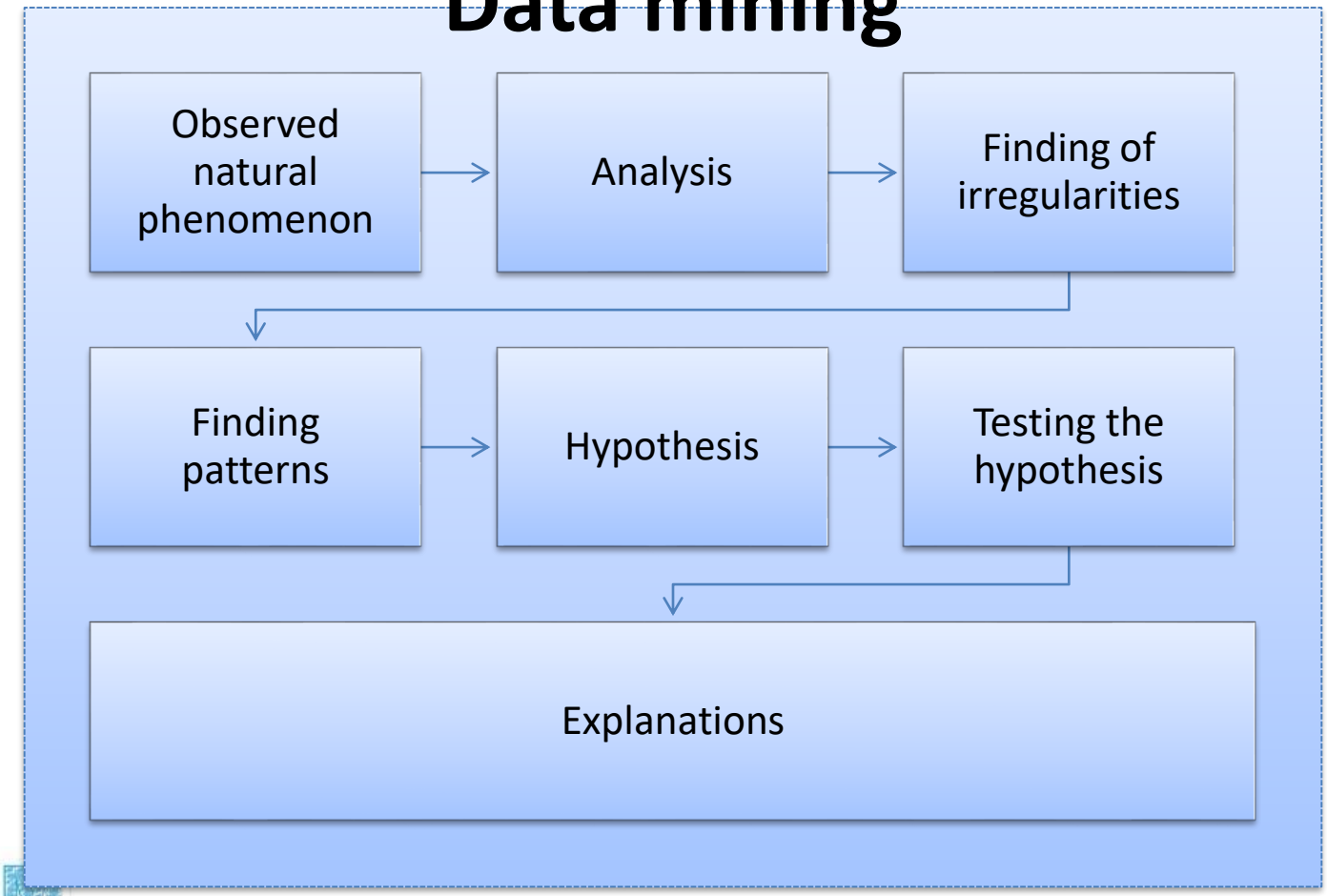
- Improvement of available datasets (CIP, etc.)
- Improvement of current tools (SAGIS, etc.)
- Generation of new datasets
- Integration of feedback and perceptions

After Checkland and Holwell (1998, p.106)

Using Big Data to improve the UKWIR's Chemicals Investigation Programme

Juan Manuel Ponce Romero

Data mining



DREAM
DATA, RISK & ENVIRONMENTAL ANALYTICAL METHODS

Using Big Data to improve the UKWIR's Chemicals Investigation Programme

Juan Manuel Ponce Romero

Machine learning is "the process by which a computer can work more accurately as it collects and learns from the data it is given" (Cielen et al., 2016)

Supervised Machine Learning

- Some training data is provided beforehand
- Classification, regression

Unsupervised Machine Learning

- No training provided, it search for a hidden structure in the data
- Clustering, association mining

Using Big Data to improve the UKWIR's Chemicals Investigation Programme

Juan Manuel Ponce Romero

Cranfield University · Newcastle University · University of Cambridge · University of Birmingham

Retrieve → Store → Analyse → Communicate

CIP (1)
- Semi-structured data
- Large volume

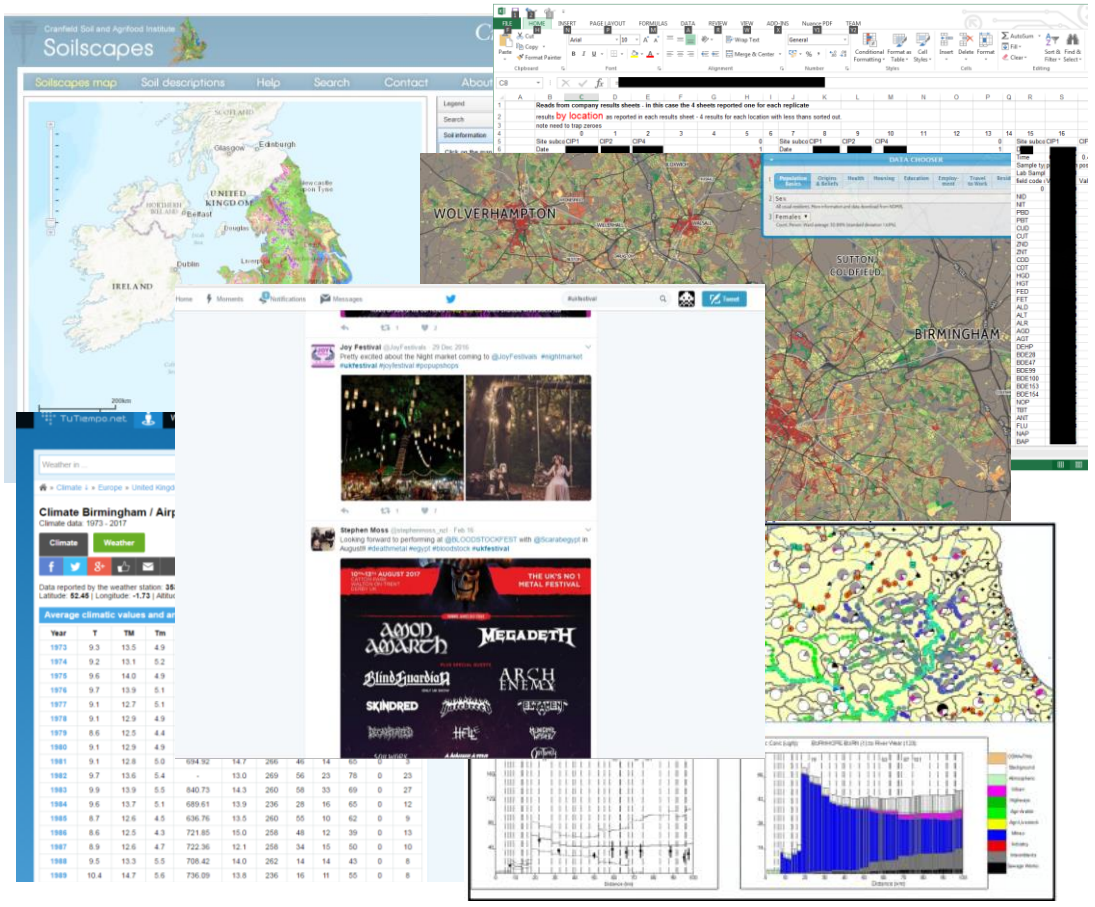
SAGIS
- Geo-referenced data

Climate data
- Different sources
- High volume
- Different formats

Soils classification
- Structured data
- Large volume

Census data
- Structured data
- Large volume

Other
- Social networks, health information…
- Diverse, large volume, different availability, velocity, and veracity

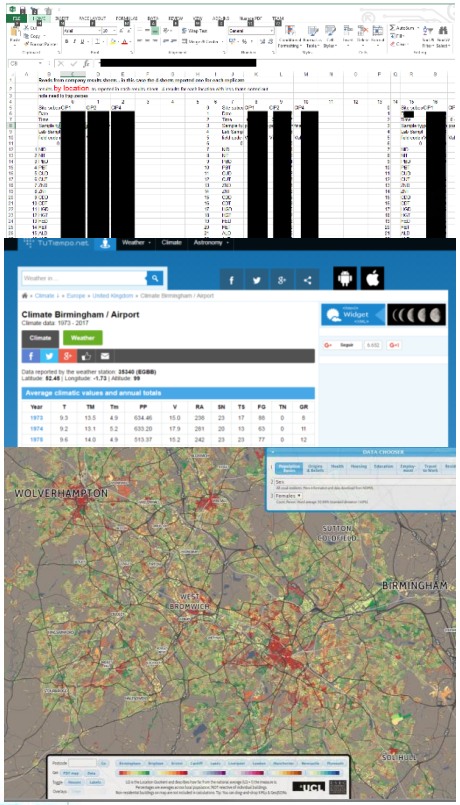


DREAM
DATA, RISK & ENVIRONMENTAL ANALYTICAL METHODS

Using Big Data to improve the UKWIR's Chemicals Investigation Programme

Juan Manuel Ponce Romero

Retrieve | **Store** | Analyse | Communicate




```
{
  _id: "6545646",
  name: "Severn",
  locations: [
        {
          east: "473839",
          north: "5559361",
          name: "Bristol",
          water_quality: "2"
        },
        {
          east: "549769",
          north: "5745397",
          name: "Gloucester",
          water_quality: "3"
        }
      ]
}
```

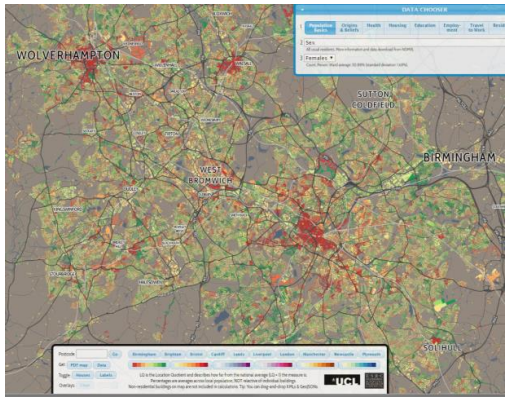
DREAM
DATA, RISK & ENVIRONMENTAL ANALYTICAL METHODS

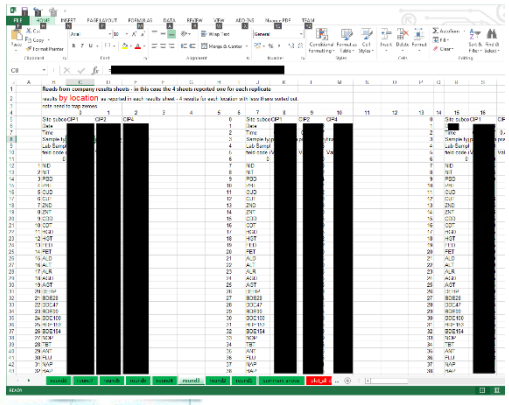
Using Big Data to improve the UKWIR's Chemicals Investigation Programme

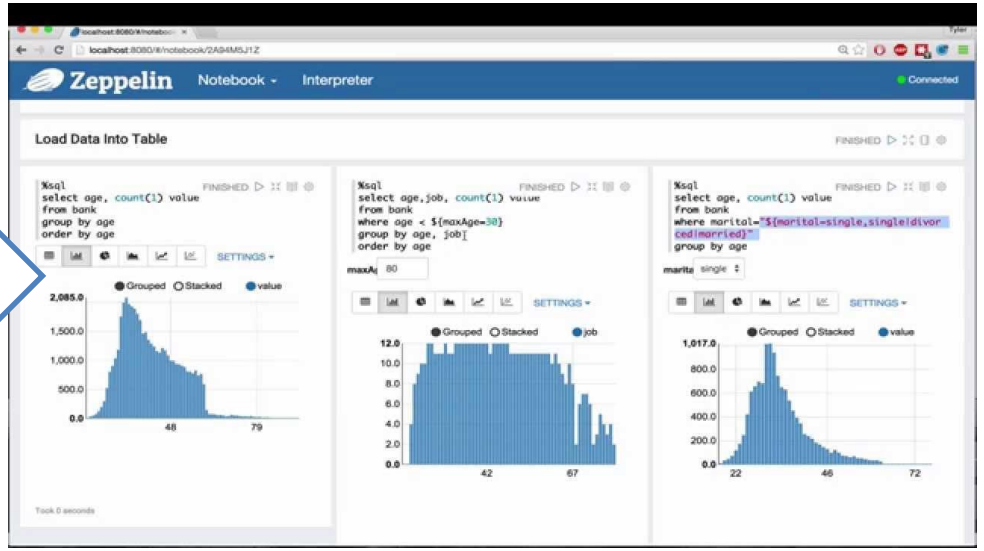
Juan Manuel Ponce Romero

Cranfield University · Newcastle University · University of Cambridge · University of Birmingham

Retrieve → Store → **Analyse** → Communicate







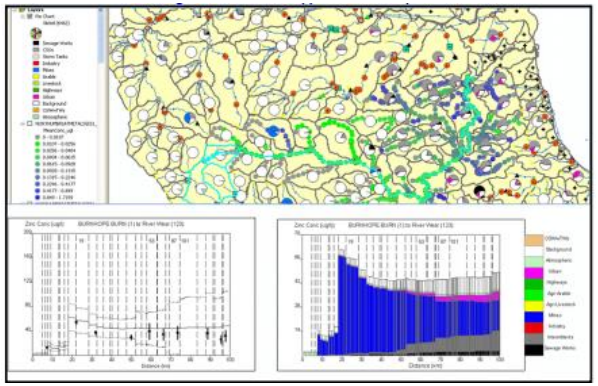
DREAM
DATA, RISK & ENVIRONMENTAL ANALYTICAL METHODS

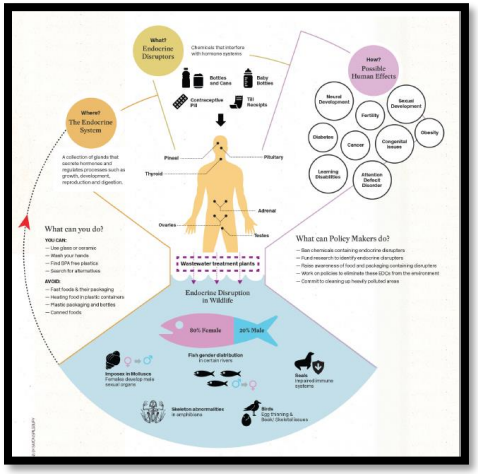
Using Big Data to improve the UKWIR's Chemicals Investigation Programme

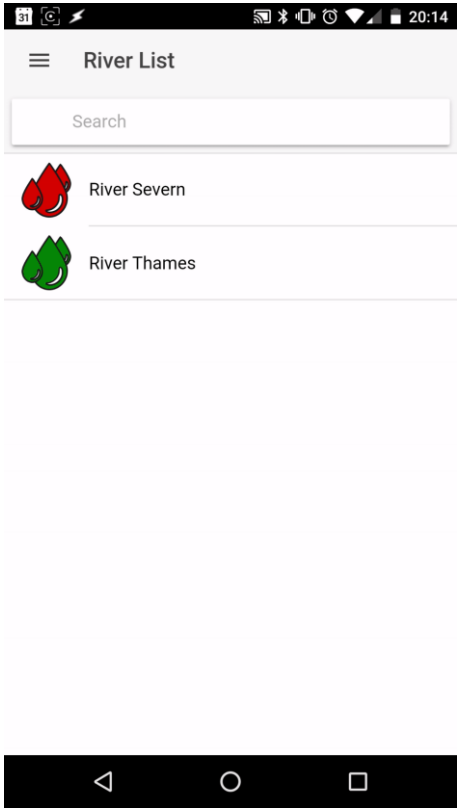
Juan Manuel Ponce Romero

Cranfield University · Newcastle University · University of Cambridge · University of Birmingham

Retrieve → Store → Analyse → **Communicate**







DREAM
DATA, RISK & ENVIRONMENTAL ANALYTICAL METHODS

References

- Gardner, M. et al. (2012) ‘The significance of hazardous chemicals in wastewater treatment works effluents’, *Science of the Total Environment*, 437, pp. 363–372.
- UKWIR (2012) *The UKWIR Chemicals Investigation Programme – A Mid-programme Update*.
- Checkland, P., & Holwell, S. (1998). *Information, systems, and information systems : making sense of the field*. Wiley
- Cielen, D.; Meysman, A.; Ali, M. *Introducing data science : big data, machine learning, and more, using Python tools*; 2016.